UNIVERSITÄT ZU LÜBECK

Aus dem Institut für Mathematik
der Universität zu Lübeck

Direktor: Prof. Dr. rer. nat. Jürgen Prestin

L^p and Pathwise Convergence of the Milstein Scheme for Stochastic Delay Differential Equations

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

aus der Sektion Informatik/Technik

vorgelegt von

Jan Pleis, M. Sc.

aus Oldenburg (Oldb)

Lübeck, August 2019

1. Berichterstatter: Prof. Dr. rer. nat. Andreas Rößler

2. Berichterstatter: Prof. Dr. rer. nat. Andreas Neuenkirch

Tag der mündlichen Prüfung: 3. März 2020

Zum Druck genehmigt. Lübeck, den 4. März 2020

Danksagung

Ich danke *allen*, die mich auf dem Weg zur Promotion unterstützt haben – insbesondere meiner Schwester Wiebke und meiner Freundin Julia. Genauso danke ich den lieben Kolleginnen und Kollegen am Institut für Mathematik; vielen Dank für die schöne Zeit mit Euch!

Ich möchte an dieser Stelle auch Andreas Neuenkirch für die Begutachtung meiner – um es mit Andreas Worten zu sagen – *enzyklopädischen Arbeit* danken. Vielen Dank!

Ein ganz besonders großer Dank gilt Andreas Rößler.
Lieber Andreas, ich danke Dir herzlich für Dein Vertrauen und die Möglichkeit, selbstständig und frei forschen zu können. Du hattest immer Zeit, meine Ideen zu diskutieren und hast, gerade auch durch Deine kritischen und interessierten Rückfragen, maßgeblich zum Gelingen dieser Arbeit beigetragen. Ich danke Dir für die guten und lustigen Gespräche, auch abseits der Mathematik, sowie die vielen gemeinsam genossenen Kaffees.
Herzlichen Dank für die wunderbare Betreuung und die schöne Zeit!

Zusammenfassung

Die Entwicklung des Milstein Verfahrens war ein großer Fortschritt in der Approximation von Lösungen stochastischer (gewöhnlicher) Differentialgleichungen. Dessen Konvergenz wurde umfassend untersucht und ist von starker Ordnung eins. Kaum analysiert wurde im Gegensatz dazu die Konvergenz des Milstein Verfahrens für stochastische *retardierte* Differentialgleichungen. Dessen numerische Analyse ist wesentlich schwieriger als im Falle von stochastischen gewöhnlichen Differentialgleichungen und ist der Schwerpunkt dieser Arbeit.

Bislang wurde die Konvergenz lediglich im quadratischen Mittel unter starken Voraussetzungen an die Differentialgleichungen betrachtet. In der vorliegenden Arbeit wird gezeigt, dass das Milstein Verfahren mit Ordnung eins in L^p für beliebige $p \in [1, \infty[$ sowie pfadweise mit Ordnung $1 - \varepsilon$ für beliebige $\varepsilon > 0$ konvergiert. Die Voraussetzungen an die Koeffizienten der Differentialgleichungen konnten dabei abgeschwächt und stochastische Prozesse als Anfangsbedingungen berücksichtigt werden.

Darüber hinaus liegt ein besonderer Fokus auf der effizienten Approximation iterierter stochastischer Integrale, die im Milstein Verfahren im Falle von nichtkommutativem Rauschen auftreten. In dieser Arbeit werden verschiedene Algorithmen vorgestellt und ihre Konvergenz in L^p für beliebige $p \in [2, \infty[$ analysiert. Bislang wurde in der Literatur lediglich die Konvergenz in L^2 betrachtet. Mit den hier präsentierten stärkeren Konvergenzaussagen ergibt sich die Konvergenz des Milstein Verfahrens, das auf Approximationen der iterierten stochastischen Integralen basiert, in L^p für beliebige $p \in [2, \infty[$ sowie pfadweise. Die Rechenkomplexität wird dabei im Vergleich zu dem Ergebnis von Hu, Mohammed und Yan (*Ann. Probab.*, 32(1A):265–314, 2004. DOI: `10.1214/aop/1078415836`) deutlich verbessert. Zwei der hier vorgestellten Algorithmen zur Approximation von iterierten stochastischen Integralen reduzieren außerdem den Rechenaufwand gegenüber dem von Wiktorsson vorgestellten Algorithmus (*Ann. Appl. Probab.*, 11(2):470–487, 2001. DOI: `10.1214/aoap/1015345301`) erheblich.

Abschließend werden einige numerische Simulationen präsentiert, um die zuvor beschriebenen theoretischen Ergebnisse zur Konvergenz des Milstein Verfahrens zu veranschaulichen. Dabei werden insbesondere nichtlineare stochastische retardierte Differentialgleichungen mit mehrdimensionalem und kommutativem Rauschen betrachtet. Deren analytischen Lösungen werden in dieser Arbeit erstmals exakt und fehlerfrei simuliert.

Abstract

The development of the Milstein scheme was a great advance in the approximation of solutions of stochastic (ordinary) differential equations. Its convergence has been extensively studied and is of strong order one. In contrast, the convergence of the Milstein scheme for stochastic *delay* differential equations has hardly been analyzed. Its numerical analysis is much more difficult than in the case of stochastic ordinary differential equations and is the focus of this thesis.

So far, the convergence has only been considered in the quadratic mean under strong assumptions regarding the differential equations. In this thesis, we prove that the Milstein scheme converges with order one in L^p for arbitrary $p \in [1, \infty[$ and with order $1 - \varepsilon$ in the pathwise sense for arbitrary $\varepsilon > 0$. Here, the assumptions on the coefficients of the differential equations are weakened and stochastic processes are considered as initial conditions.

Furthermore, a special focus is on the efficient approximation of iterated stochastic integrals that occur in the Milstein scheme in case of noncommutative noise. In this thesis, we present various algorithms and analyze their convergence in L^p for arbitrary $p \in [2, \infty[$. So far, in the literature, the convergence has been considered in L^2 only. The stronger findings on the convergence presented here result in the convergence of the Milstein scheme, which is based on approximations of the iterated stochastic integrals, in L^p for arbitrary $p \in [2, \infty[$ as well as in the pathwise sense. The computational complexity of the Milstein scheme with approximated iterated stochastic integrals is significantly improved in comparison to the result by Hu, Mohammed, and Yan (*Ann. Probab.*, 32(1A):265–314, 2004. DOI: `10.1214/aop/1078415836`). Moreover, two of the algorithms for the approximation of iterated stochastic integrals presented here reduce the computational effort substantially compared to the algorithm derived by Wiktorsson (*Ann. Appl. Probab.*, 11(2):470–487, 2001. DOI: `10.1214/aoap/1015345301`).

Finally, numerical simulations are presented in order to illustrate and confirm the theoretical results on the convergence of the Milstein scheme. Here, we especially consider nonlinear stochastic delay differential equations with multidimensional and commutative noise. In this thesis, their analytical solutions are simulated exactly and error-free for the first time.

Contents

I
Introduction

The Milstein scheme was a major advancement in the approximation of solutions of stochastic ordinary differential equations (SODEs). This method was introduced by Milstein in [106], and it was the first that converges with order $\mathcal{O}(h)$ in L^2 to the SODEs' solutions as maximum step size $h \to 0$.

Priorly, Maruyama showed that an Euler-type scheme converges to the solution of SODEs in L^2, see [101, Theorem 1]. According to [105, p. 4], Gihman and Skorohod were the first who proved that the convergence of the so-called Euler-Maruyama method is of order $\mathcal{O}(\sqrt{h})$ in L^2, see [46, pp. 237–241]. Thus, this order of convergence is lower than the one of the Euler method, introduced by Euler [39, pp. 424–425], in the case of deterministic ordinary differential equations. There, the Euler method converges with order $\mathcal{O}(h)$ as $h \to 0$, cf. [54, Section I.7]. This already indicates that randomness in stochastic differential equations (SDEs) has a large influence on the order of convergence and makes the analysis of numerical methods more difficult.

Since Milstein's article [106] in 1975, approximations of solutions of SODEs are extensively studied. Their convergence is analyzed, among others, in L^p for $p \in [1, \infty[$, and pathwise. We refer to [41, 77, 78, 105, 125] to name only a selection of references. Recently, numerical solutions of stochastic partial differential equations (SPDEs) have also been studied relative comprehensively, see, among others, [11, 13, 25, 69, 71, 89].

In contrast to this, numerical solutions, especially those that converge with a higher order than $\mathcal{O}(\sqrt{h})$, of stochastic *delay* differential equations (SDDEs) have not been analyzed to the same extent. SDDEs are SDEs whose evolution in time depends on its past history, that is, coefficients of an SDDE incorporate, in addition to the current state, discrete information about prior states of the equation's solution, cf. [65, 98, 107]. Whereas the convergence of the Euler-Maruyama scheme for SDDEs is broadly studied, see e. g. [2, 52, 77, 83, 99, 100], there are only few results on the Milstein scheme or other methods of higher order. Contrary to what one would expect, the convergence analysis of numerical solutions, that converge with a higher order than $\mathcal{O}(\sqrt{h})$, turned out to be substantially more difficult in case of SDDEs in comparison to SODEs. In this regard, we refer to [60, 80, 137], where the convergence in L^2 of the Milstein scheme to the SDDE's solution is shown. In order to prove the convergence of order $\mathcal{O}(h)$ as $h \to 0$, their numerical analyses are on the one hand based on the Malliavin calculus, see [60, 137], and on the other, on the differentiation of the SDDE's solution with respect to its initial condition, see [80]. Thus, their proofs involve more sophisticated techniques from stochastic analysis in comparison to convergence analyses of the Euler-Maruyama scheme for SDDEs, approximations of SODEs' solutions, or numerical solutions of SPDEs. In the latter cases, the proofs of convergence are mainly based on Itô's calculus, cf. [11, 71, 77, 78, 105]. We see in [60, 80, 137] that the difficulties occurring in the convergence analyses are caused by the delay in the drift coefficient of the SDDE. In [104, 124], higher order approximations of solutions of SDDEs are considered. However, the presented numerical schemes are not optimal in the sense that a scheme of order $\mathcal{O}(h)$ contains itself terms of order $\mathcal{O}(h)$, for example.

In this thesis, the focus is on the Milstein scheme for SDDEs. As a main result, we prove in Chapter IV that the Milstein scheme converges to the SDDE's solution with order $\mathcal{O}(h)$ in L^p for $p \in [1, \infty[$ as its maximum step size $h \to 0$. Here, we allow the SDDE to have a stochastic process as initial condition, consider the supremum over time inside the expectation of the L^p-norm, and postulate less restrictive assumptions on the SDDE's coefficients than in [60, 80, 137]. The pathwise convergence subsequently follows as a corollary from the main result, cf. [2, 41, 77]. Thus, we improve the results obtained in [60, 80, 137] eminently as they only considered the convergence in L^2 for SDDEs with deterministic initial conditions.

The Milstein scheme contains iterated stochastic integrals in order to achieve the higher order of convergence $\mathcal{O}(h)$. These stochastic integrals can be simulated by normally distributed increments of the Wiener process only if the diffusion coefficients do not depend on the prior development of the SDDE's solution and satisfy a so-called commutativity condition, cf. [23]. Thus, in order to make the Milstein scheme applicable in general, the iterated stochastic integrals in the Milstein scheme have to be substituted by appropriate approximations. In Chapter V, we consider various approximations. At first, we prove the convergence in L^p of a so-called Fourier method for nondelayed- and delayed-iterated stochastic integrals. This method was first developed by Milstein in case of nondelayed-iterated stochastic integrals occurring in the Milstein scheme for SODEs, see [105], and afterwards extended by Yan to the case of delayed-iterated stochastic integrals, see [60, 137]. Both only proved the convergence in L^2. Moreover, we improve the computational complexity of the Milstein method for SDDEs with approximated iterated stochastic integrals compared to the result in [60]. In Chapter V, we further focus on nondelayed-iterated stochastic integrals in particular. Here, we improve the algorithm that Wiktorsson developed in [136] and significantly reduce the computational effort. Moreover, we prove the convergence of our algorithm in L^p for arbitrary $p \in [2, \infty[$, and not just that it is convergent in L^2. Thus, we obtain that the Milstein scheme with approximated iterated stochastic integrals converges in L^p for arbitrary $p \in [1, \infty[$ and pathwise as well.

In Chapter VI, we present some numerical simulations in order to illustrate and confirm our theoretical results. In order to compare the approximations obtained by the Euler-Maruyama scheme and the Milstein scheme, we first derive analytical solutions of SDDEs that can be simulated error-free. Here, we consider linear SDDEs with additive noise on the one hand, but also more general SDDEs with multidimensional and commutative noise on the other. These results make the difficulty and complexity in the exact simulation of analytical solutions clear and emphasize the demand for efficient numerical methods. To the best of our knowledge, the presented numerical simulations are the first that compare the Milstein approximations with the correctly simulated analytical solutions of SDDEs.

Chapter II and Chapter III provide fundamentals for the numerical analysis of SDDEs. We present the existence and uniqueness of strong solutions for SDDEs and state important properties of their strong solutions in Chapter II. Further, we derive inequalities for time-discrete and time-continuous martingales that are similar to the well-known Burkholder inequalities.

However, the here presented inequalities have smaller constants and are therefore highly valuable in the analysis of numerical methods for SDEs. Chapter III serves as an introduction to the Malliavin calculus. Here, we prove a more general chain rule for the Malliavin derivative and show that solutions of SDDEs with deterministic initial conditions are differentiable in the sense of Malliavin. These results are used in the numerical analysis of the Milstein scheme in Chapter IV.

Finally, we conclude this thesis with Chapter VII, where we summarize our several new results and mention some open problems.

Throughout this thesis, we consider SDEs whose stochastic integrals are defined in the sense of Itô, cf. [63]. We refer to [29] for a result by which SDEs with Stratonovich-stochastic integrals, cf. [129], can be converted to Itô-SDEs.

II
Stochastic Delay Differential Equations

The SDDEs considered in this thesis as well as the involved spaces and stochastic processes are introduced in this chapter. Further, in Section II.2, we depict the well-known Burkholder inequalities and derive more accurate inequalities of a similar type that are highly valuable for the analysis of solutions of SDEs and their numerical approximations. The concept of strong solutions of SDDEs is introduced below. In Section II.3, the existence and uniqueness of such solutions are shown.

Throughout this thesis, the points in time $t_0, T \in \mathbb{R}$ with $t_0 < T$ denote the starting point and the finite time horizon of the evolution of the SDDEs under consideration. Let $(\Omega, \mathscr{F}, \mathrm{P})$ be a complete probability space with a filtration $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$ that satisfies the usual condition, see [75, Definition 1.2.25], and where $\tau \geq 0$ is a constant, which is specified later. Further, we consider an m-dimensional Wiener process W on $(\Omega, \mathscr{F}, \mathrm{P})$ with respect to filtration $(\mathscr{F}_t)_{t\in[t_0,T]}$ that is defined similarly to [12, Definition 40.1] and [75, Definition 2.5.1].

Definition II.1 (Wiener Process)
Let $m \in \mathbb{N}$ and Q be a probability measure on $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$. A measurable stochastic process $W\colon [t_0,T] \times \Omega \to \mathbb{R}^m$ on $(\Omega, \mathscr{F}, \mathrm{P})$ that is adapted to $(\mathscr{F}_t)_{t\in[t_0,T]}$ is called (m-dimensional) Wiener process *with respect to filtration $(\mathscr{F}_t)_{t\in[t_0,T]}$ and with initial distribution Q if*

i) $\mathrm{P}[W_{t_0} \in B] = Q[B]$ *for all* $B \in \mathscr{B}(\mathbb{R}^m)$,

ii) the realizations $t \mapsto W_t$ are P*-almost surely continuous,*

iii) for all $s, t \in [t_0, T]$ with $s < t$, the increments $W_t - W_s$ are independent of $\mathscr{F}_s$, and

iv) for all $s, t \in [t_0, T]$ with $s < t$, the increments $W_t - W_s$ are $\mathrm{N}(0, (t-s)I_m)$*-distributed, that is, they are normally distributed with expectation $0 \in \mathbb{R}^m$ and covariance $(t-s)I_m$, where I_m is the identity matrix in $\mathbb{R}^{m\times m}$.*

Evidently, the Wiener process W is adapted to its augmented natural filtration $(\mathscr{F}_t^W)_{t\in[t_0,T]}$, which is defined by $\mathscr{F}_t^W := \sigma(\{W_s : s \in [t_0,t]\} \cup \mathscr{N})$, where $\mathscr{N} = \{A \in \mathscr{F} : \mathrm{P}[A] = 0\}$ is the collection of all P-null sets, see [75, Section 2.7]. Moreover, the process $(B_t)_{t\in[0,T-t_0]}$ with $B_t = W_{t_0+t} - W_{t_0}$ is clearly the standard Wiener process or Brownian motion with $B_0 = 0$ P-almost surely. The advantage of the generality of the Wiener process W with initial distribution Q from Definition II.1 is described later.

As the coefficients of SDDEs incorporate discrete information about the prior development of their solutions, we have to introduce the time lags that specify these retardations. Define $\tau_0 \coloneqq 0$, and let $\tau_l \in \,]0,\infty[$ be constants for $l \in \{1,\dots,D\}$, where $D \in \mathbb{N}_0$ is the number of different positive delays of the SDDE under consideration.

Now, we describe the d-dimensional SDDEs considered in this thesis, where $d \in \mathbb{N}$. Let $a, b^j \colon \mathbb{R}^{1\times(D+1)} \times \mathbb{R}^{d\times(D+1)}$ be Borel-measurable functions for $j \in \{1,\dots,m\}$. Further, consider a measurable stochastic process $\xi \colon [t_0-\tau, t_0] \times \Omega \to \mathbb{R}^d$ that is adapted to the filtration $(\mathscr{F}_t)_{t\in[t_0-\tau,t_0]}$ and serves as the initial condition, where $\tau \geq \max_{l\in\{0,1,\dots,D\}} \tau_k$ is arbitrary but fixed. Then, the d-dimensional SDDE with D delays and initial condition ξ is given by

$$X_t = \begin{cases} \xi_t \quad \text{if } t \in [t_0-\tau, t_0], \\ \xi_{t_0} + \displaystyle\int_{t_0}^{t} a(s, s-\tau_1,\dots,s-\tau_D, X_s, X_{s-\tau_1},\dots,X_{s-\tau_D})\,\mathrm{d}s \\ \qquad + \displaystyle\sum_{j=1}^{m}\int_{t_0}^{t} b^j(s, s-\tau_1,\dots,s-\tau_D, X_s, X_{s-\tau_1},\dots,X_{s-\tau_D})\,\mathrm{d}W_s^j \quad \text{if } t \in \,]t_0, T], \end{cases} \tag{II.1}$$

where the stochastic integral is defined in the sense of Itô. We refer to e. g. [21], [75], and [84] for monographs addressing the stochastic integration. Note that the SDDE (II.1) above simply reduces to an SODE in case of $D = 0$. Further, the generality of the Wiener process with initial distribution Q allows us to consider a Wiener process $\widetilde{W}$ that actually starts in $t_0 - \tau$ by choosing $Q[B] = \mathrm{P}^{\widetilde{W}_{t_0}}[B] = \mathrm{P}[\widetilde{W}_{t_0} \in B]$ for all $B \in \mathscr{B}(\mathbb{R}^m)$ in Definition II.1. Then, the initial condition ξ can also depend on the Wiener process $\widetilde{W}$ and can thus be semimartingale for example.

Throughout this thesis, we consider strong solutions of the SDDEs (II.1) that are defined as follows, cf. [75, Definition 5.2.1] in case of SODEs and [98, Definition 5.2.1].

Definition II.2
A measurable stochastic process $X \colon [t_0-\tau, T] \times \Omega \to \mathbb{R}^d$ is called strong solution *of SDDE* (II.1) *with respect to the fixed Wiener process W and to initial condition ξ if*

i) $(X_t)_{t\in[t_0-\tau,T]}$ *is adapted to* $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$,

ii) $(X_t)_{t\in[t_0,T]}$ *has* P*-almost surely continuous realizations,*

iii) $$\int_{t_0}^{t} \|a(s, s-\tau_1,\dots,s-\tau_D, X_s, X_{s-\tau_1},\dots,X_{s-\tau_D})\| + \sum_{j=1}^{m}\|b^j(s, s-\tau_1,\dots,s-\tau_D, X_s, X_{s-\tau_1},\dots,X_{s-\tau_D})\|^2\,\mathrm{d}s < \infty$$
holds P*-almost surely for all* $t \in [t_0, T]$,

iv) and if equation (II.1) *holds for all* $t \in [t_0-\tau, T]$ P*-almost surely.*

Note that in contrast to definitions of strong solutions in [98, Definition 5.2.1], [107, page 35], and [109, page 10], our definition does not impose the continuity of the solution on $[t_0-\tau, t_0]$. This will be important later on in Chapter III. As we only consider strong solutions in this thesis, we may omit the adjective *strong* from time to time, and a solution of SDDE (II.1) always refers to a strong solution. If we want to emphasize the initial condition ξ of a solution X, we use the notation X^{ξ}.

In the following, we give some remarks on the filtration $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$. Instead of considering filtration $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$, one can also use the augmented natural filtration $(\mathscr{H}_t)_{t\in[t_0-\tau,T]}$ defined by

$$\mathscr{H}_t := \begin{cases} \sigma\big(\{\xi_s : s \in [t_0-\tau, t]\} \cup \mathcal{N}\big) & \text{if } t \in [t_0-\tau, t_0[\text{ and} \\ \sigma\big(\mathscr{F}_t^W \cup \{\xi_s : s \in [t_0-\tau, t_0]\}\big) & \text{if } t \in [t_0, T], \end{cases}$$

where $\mathcal{N} = \{A \in \mathscr{F} : \mathrm{P}[A] = 0\}$. Initial condition ξ is clearly adapted to $(\mathscr{H}_t)_{t\in[t_0-\tau,t_0]}$, therefore, this does not have to be imposed a priori. However, we then have to suppose that ξ_t is independent of σ-algebra $\mathscr{G} := \mathscr{G}_T$ for all $t \in [t_0-\tau, t_0]$, where

$$\mathscr{G}_t := \sigma\big(\{W_s - W_{t_0} : s \in [t_0, t]\} \cup \mathcal{N}\big) \tag{II.2}$$

for $t \in [t_0, T]$. As filtration $(\mathscr{H}_t)_{t\in[t_0-\tau,T]}$ is more restrictive, we stick to filtration $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$. The assumption of the independence of σ-algebra $\mathscr{G}$ does not have to be imposed because we presume the existence of the Wiener process, and thus it is fulfilled anyway. Nevertheless, σ-algebra $\mathscr{G}$ generated by the normally distributed random variables $W_s - W_{t_0}$, $s \in [t_0, T]$, plays an important rule in Malliavin's calculus, see Chapter III.

II.1. On Measurability of Stochastic Processes, Notations, and Spaces

In order to keep formulas and terms concise, we introduce the following abbreviations. Considering SDDE (II.1), we define

$$\mathcal{T}(t, X_t) := (t, t-\tau_1, \ldots, t-\tau_D, X_t, X_{t-\tau_1}, \ldots, X_{t-\tau_D}) \tag{II.3}$$

for $t \in [t_0, T]$. Then, SDDE (II.1) can be rewritten to

$$X_t = \begin{cases} \xi_t & \text{if } t \in [t_0-\tau, t_0], \\ \xi_{t_0} + \displaystyle\int_{t_0}^t a(\mathcal{T}(s, X_s))\,\mathrm{d}s + \sum_{j=1}^m \int_{t_0}^t b^j(\mathcal{T}(s, X_s))\,\mathrm{d}W_s^j & \text{if } t \in \,]t_0, T] \end{cases}$$

in short notation. Further, for the sake of brevity, we write $x \vee y := \max\{x, y\}$ and $x \wedge y := \min\{x, y\}$ for $x, y \in \mathbb{R}$ throughout this thesis.

In the following, consider a real separable Hilbert space E with inner product $\langle\cdot,\cdot\rangle_E$ and norm $\|\cdot\|_E$. If $E = \mathbb{R}^d$ for some $d \in \mathbb{N}$ with $d > 1$, we neglect the subscript on the norm and simply write $\|\cdot\| := \|\cdot\|_{\mathbb{R}^d}$ for the Euclidean norm in $\mathbb{R}^d$. In the case of $E = \mathbb{R}$, the notation $|\cdot| := \|\cdot\|_{\mathbb{R}}$ is used for the absolute value.

The Banach space of all (equivalence classes of) E-valued random variables on $(\Omega, \mathscr{F}, \mathrm{P})$ with finite p-mean, $p \in [1, \infty[$, is denoted by $L^p(\Omega; E) := L^p((\Omega, \mathscr{F}, \mathrm{P}); (E, \mathscr{B}(E)))$ and endowed with the norm

$$\|\cdot\|_{L^p(\Omega;E)} = \big(\mathrm{E}[\|\cdot\|^p]\big)^{\frac{1}{p}},$$

where random variables that equal P-almost surely are identified, and $\mathrm{E}[\cdot]$ denotes the expectation on $(\Omega, \mathscr{F}, \mathrm{P})$. In this thesis, equivalence classes and their representatives are however not distinguished. Thus, $Z \in L^p(\Omega; E)$ is referred to as a fixed $\mathscr{F}/\mathscr{B}(E)$-measurable function instead of an equivalence class.

Let λ denote the Lebesgue-measure on $\mathbb{R}$, and consider an interval $A \subseteq [t_0 - \tau, T]$. The space $H^p(A \times \Omega; E) := H^p((A \times \Omega, \mathscr{B}(A) \otimes \mathscr{F}, \lambda|_A \otimes \mathrm{P}); (E, \mathscr{B}(E)))$ with $p \in [1, \infty[$ denotes the space of all (equivalence classes of) E-valued and $(\mathscr{F}_t)_{t\in A}$-progressively measurable processes $Z\colon A \times \Omega \to E$ with the finite norm

$$\|Z\|_{H^p(A\times\Omega;E)} := \left(\mathrm{E}\left[\left(\int_A \|Z_t\|_E^2 \,\mathrm{d}t\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}}$$

where indistinguishable processes are identified, cf. [45], [44], [103], [33, p. 274], and [119, p. 195]. Further, the space $S^p(A \times \Omega; E) := S^p((A \times \Omega, \mathscr{B}(A) \otimes \mathscr{F}, \lambda|_A \otimes \mathrm{P}); (E, \mathscr{B}(E)))$ with $p \in [1, \infty[$ denotes the space of all (equivalence classes of) E-valued and $(\mathscr{F}_t)_{t\in A}$-progressively measurable processes $Z\colon A \times \Omega \to E$ whose realizations are P-almost surely càdlàg and the norm

$$\|Z\|_{S^p(A\times\Omega;E)} := \left(\mathrm{E}\left[\sup_{t\in A} \|Z_t\|_E^p\right]\right)^{\frac{1}{p}}$$

is finite, where indistinguishable processes are identified again, cf. [103], [33, p. 253], [118, p. 339], and [119, p. 250]. A realization is called càdlàg – continue à droite limites à gauche – if it is right continuous with left-hand limits [32, p. 90]. The letter S of the space $S^p(A \times \Omega; E)$ stands for *s*upremum.

In Section II.3, we show that under certain conditions there exists a strong solution X of SDDE (II.1) that belongs to $S^p([t_0 - \tau, T] \times \Omega; \mathbb{R}^d)$, where $p \in [2, \infty[$.

For $Z \in H^p(A \times \Omega; E)$ or $Z \in S^p(A \times \Omega; E)$, the process Z again is a fixed representative. Moreover, the term $\|Z\|_E$ is referred to as the $(\mathscr{F}_t)_{t\in A}$-progressively measurable process $\|Z\|_E\colon A \times \Omega \to \mathbb{R}$ with $(t, \omega) \mapsto \|Z_t(\omega)\|_E$, where $\|Z\| \in H^p(A \times \Omega; \mathbb{R})$ and $S^p(A \times \Omega; \mathbb{R})$, respectively.

Consider an integrable and $(\mathscr{F}_t)_{t\in[t_0,T]}$-progressively measurable process Z, e. g. $Z \in S^2([t_0, T] \times \Omega; \mathbb{R})$. Then, Fubini's theorem implies that the process $t \mapsto \int_{t_0}^t Z_s \,\mathrm{d}s$, $t \in [t_0, T]$, is adapted to the filtration $(\mathscr{F}_t)_{t\in[t_0,T]}$. In general this is not the case, when we only assume the process Z to be measurable and adapted, cf. [21, Section 3.2 and Section 3.4] and [72]. Especially in [21, Example, p. 62], an example of a measurable and adapted process that is not progressively measurable is provided.

Although not every measurable and $(\mathscr{F}_t)_{t\in A}$-adapted stochastic process is $(\mathscr{F}_t)_{t\in A}$-progressively measurable, any processes of this kind have an $(\mathscr{F}_t)_{t\in A}$-progressively measurable modification [32, Theorem IV.30 on p. 99], see also [72, Theorem 1]. However, a modification does not preserve continuity properties of that process in general. But we have the following. If every realization of an $(\mathscr{F}_t)_{t\in A}$-adapted and measurable process is right continuous, the process is $(\mathscr{F}_t)_{t\in A}$-progressively measurable, see [32, Theorem IV.15 on p. 89] or [75, Proposition 1.1.13]. However, this is not the case, when only P-almost all realizations are right continuous. Consider a measurable and $(\mathscr{F}_t)_{t\in[t_0,T]}$-adapted stochastic process $Z\colon [t_0, T] \times \Omega \to E$ that has P-almost surely right continuous realizations. The process Z is $(\mathscr{F}_t)_{t\in[t_0,T]}$-progressively measurable if and only if $Z|_{[t_0,t]} := (Z_s)_{s\in[t_0,t]}$ is $\mathscr{B}([t_0,t]) \otimes \mathscr{F}_t/\mathscr{B}(E)$-measurable for all $t \in [t_0, T]$. That is, for all $B \in \mathscr{B}(E)$ the preimage $Z|_{[t_0,t]}^{-1}(B)$ belongs to $\mathscr{B}([t_0,t]) \otimes \mathscr{F}_t$ for all $t \in [t_0,T]$. Let $N \in \mathscr{F}$ be a P-null set such that $Z(\omega)$ is right continuous for all $\omega \in \Omega \setminus N$. Recall that probability space $(\Omega, \mathscr{F}, \mathrm{P})$ is complete and P-null set $N \in \mathscr{F}_t$ for all $t \in [t_0,T]$. Thus, we have $[t_0,t] \times N \in \mathscr{B}([t_0,t]) \otimes \mathscr{F}_t$, and it holds $(\lambda \otimes \mathrm{P})\big[[t_0,t] \times N\big] = \lambda\big[[t_0,t]\big]\mathrm{P}[N] = 0$ for all $t \in [t_0,T]$. Consider the preimage $Z|_{[t_0,t]}^{-1}(B) \in \mathscr{B}([t_0,t]) \otimes \mathscr{F}_t$ for arbitrary $B \in \mathscr{B}(E)$, where

$$Z|_{[t_0,t]}^{-1}(B) = \big(Z|_{[t_0,t]}^{-1}(B) \cap ([t_0,t] \times \Omega \setminus N)\big) \cup \big(Z|_{[t_0,t]}^{-1}(B) \cap ([t_0,t] \times N)\big)$$

for all $t \in [t_0,T]$. The set $Z|_{[t_0,t]}^{-1}(B) \cap ([t_0,t] \times N)$ can be a subset of a null set in general. But product-σ-algebra $\mathscr{B}([t_0,t]) \otimes \mathscr{F}_t$ is not complete, thus, this set may not belong to $\mathscr{B}([t_0,t]) \otimes \mathscr{F}_t$. In general, $Z|_{[t_0,t]}$ is not $\mathscr{B}([t_0,t]) \otimes \mathscr{F}_t/\mathscr{B}(E)$-measurable, and thus, process Z is not $(\mathscr{F}_t)_{t\in[t_0,T]}$-progressively measurable.

A measurable and $(\mathscr{F}_t)_{t\in A}$-adapted stochastic process $Z : \Omega \times A \to E$ with P-almost surely right continuous realizations is however indistinguishable from an $(\mathscr{F}_t)_{t\in A}$-progressively measurable process $\tilde{Z} : \Omega \times A \to E$. Here, $A \subseteq [t_0 - \tau, T]$ is still an interval, and E is a real separable Hilbert space. In order to verify that assertion, let $N \in \mathscr{F}$ with $\mathrm{P}[N] = 0$ be the null set so that $Z(\omega)$ is right continuous for all $\omega \in \Omega \setminus N$. Then, let $\tilde{Z}(\omega) = Z(\omega)$ for all $\omega \in \Omega \setminus N$, and set $\tilde{Z}_t(\omega) = 0$ for all $(t,\omega) \in A \times N$ for example. Such a process $\tilde{Z}$ is clearly indistinguishable from Z. Since all realizations of $\tilde{Z}$ are right continuous, the process is further $(\mathscr{F}_t)_{t\in A}$-progressively measurable, see [32, Theorem IV.15 on p. 89] or [75, Proposition 1.1.13]. The same holds true when the process Z has P-almost surely left continuous, continuous, or càdlàg realizations.

Due to this, $(\mathscr{F}_t)_{t\in A}$-adapted measurable stochastic processes that have P-almost surely right continuous realizations, e. g. càdlàg processes, can be modified on a P-null set, by preserving the regularity property of P-almost all realizations, such that they are $(\mathscr{F}_t)_{t\in A}$-progressively measurable and indistinguishable from the originated processes. Throughout this thesis, the $(\mathscr{F}_t)_{t\in A}$-adapted measurable processes and its indistinguishable, $(\mathscr{F}_t)_{t\in A}$-progressively measurable variant are not distinguished because the processes below will only be unique up to indistinguishability. Hence, without the loss of generality, the $(\mathscr{F}_t)_{t\in A}$-progressive measurability property of such processes can be assumed.

We continue introducing some further notations and spaces. Let $C(A;\mathbb{R}^d)$ denote the space of continuous functions $f\colon A \to \mathbb{R}^d$ where e. g. $A \subset \mathbb{R}$ or $A = \mathbb{R}^{d\times(D+1)}$. The latter case is important in consideration of the coefficients of SDDE (II.1). Having drift coefficient $a = (a^1, \ldots, a^d)^{\mathrm{T}}$ and diffusion coefficient $b^j = (b^{1,j}, \ldots, b^{d,j})^{\mathrm{T}}$ for $j \in \{1, \ldots, m\}$ in mind, we consider functions $f = (f^1, \ldots, f^d)^{\mathrm{T}}\colon \mathbb{R}^{d\times(D+1)} \to \mathbb{R}^d$ in the following. We denote the partial derivatives of f by $\partial_{x_l^i} f = (\partial_{x_l^i} f^1, \ldots, \partial_{x_l^i} f^d)^{\mathrm{T}}$ for $i \in \{1, \ldots, d\}$ and $l \in \{0, 1, \ldots, D\}$ where $x_l = (x_l^1, \ldots, x_l^d)^{\mathrm{T}} \in \mathbb{R}^d$. If the function f and all these partial derivatives exist and are continuous, we write $f \in C^1(\mathbb{R}^{d\times(D+1)};\mathbb{R}^d)$. If, in addition, the partial derivatives of second order $\partial_{x_k^j}\partial_{x_l^i} f$ exist and are continuous for all $i,j \in \{1, \ldots, d\}$ and $k,l \in \{0,1,\ldots,D\}$, we write $f \in C^2(\mathbb{R}^{d\times(D+1)};\mathbb{R}^d)$. In the assumptions regarding the Milstein scheme, see Section IV.2, we suppose for example that the spatial partial derivatives up the second order of the coefficients of the SDDE (II.1) exist and are continuous. That is $a(t, t-\tau_1, \ldots, t-\tau_D, \cdot, \ldots, \cdot), b^j(t, t-\tau_1, \ldots, t-\tau_D, \cdot, \ldots, \cdot) \in C^2(\mathbb{R}^{d\times(D+1)};\mathbb{R}^d)$ for all $t \in [t_0,T]$ and $j \in \{1, \ldots, m\}$. We emphasize that the term $\partial_{x_l^i} f$ denotes the partial derivative of f, and the symbol $\partial_{x_l^i}$ should not be understood as the derivative operator. That is $\partial_{x_l^i} f(x_0, x_1, \ldots, x_D) g(x)$ refers to as $\big(\partial_{x_l^i} f(x_0, x_1, \ldots, x_D)\big) g(x)$, where $g\colon A \to \mathbb{R}$ and $x \in A$.

II.2. On Inequalities for Martingales

In order to estimate martingales in $L^p(\Omega;\mathbb{R}^d)$ or $S^p([t_0,T]\times\Omega;\mathbb{R}^d)$ for $p\in[2,\infty[$, the time-discrete and time-continuous Burkholder inequalities provide powerful estimates.

Theorem II.3 (Discrete Burkholder Inequalities, [19])
Let $p\in[2,\infty[$ and $N\in\mathbb{N}$. Consider a discrete martingale $(M_n)_{n\in\{0,1,\dots,N\}}$ in $L^p(\Omega;\mathbb{R}^d)$ with respect to the filtration $(\mathscr{F}_{t_n})_{n\in\{0,1,\dots,N\}}$, where $M_n=\sum_{k=0}^n d_k$ with $d_0=M_0$ and $d_k=M_k-M_{k-1}$ for $k\in\{1,\dots,N\}$. Then, it holds

$$\|M_n\|^2_{L^p(\Omega;\mathbb{R}^d)}\le(p-1)^2\bigg\|\sum_{k=0}^n\|d_k\|^2\bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})}$$

and

$$\bigg\|\sup_{\nu\in\{0,1,\dots,n\}}\|M_\nu\|\bigg\|^2_{L^p(\Omega;\mathbb{R})}\le p^2\bigg\|\sum_{k=0}^n\|d_k\|^2\bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})}$$

for all $n\in\{0,1,\dots,N\}$. The constants are best possible.

Proof. The first inequality is stated in [19, Theorem 3.1]. The second follows from Doob's maximal inequality [35, Theorem 3.4 on p. 317] and is stated in [19, Inequality (3.4)], cf. [20]. □

The inequalities in Theorem II.3 carry over to time-continuous martingales [19, 34].

Theorem II.4 (Burkholder Inequalities, [19])
Let $p\in[2,\infty[$, and let $f^j\in H^p([t_0,T]\times\Omega;\mathbb{R}^d)$ for $j\in\{1,\dots,m\}$. Then, it holds

$$\bigg\|\sum_{j=1}^m\int_{t_0}^t f_u^j\,\mathrm{d}W_u^j\bigg\|^2_{L^p(\Omega;\mathbb{R}^d)}\le(p-1)^2\bigg\|\int_{t_0}^t\sum_{j=1}^m\big\|f_u^j\big\|^2\,\mathrm{d}u\bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})}$$

and

$$\bigg\|\sum_{j=1}^m\int_{t_0}^{\cdot} f_u^j\,\mathrm{d}W_u^j\bigg\|^2_{S^p([t_0,t]\times\Omega;\mathbb{R}^d)}\le p^2\bigg\|\int_{t_0}^t\sum_{j=1}^m\big\|f_u^j\big\|^2\,\mathrm{d}u\bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})}$$

for all $t\in[t_0,T]$. The constants are best possible.

Proof. See [19, Inequality (4.1)] for the first inequality. The second inequality then follows from Doob's maximal inequality, see e. g. [122, Theorem II.1.7]. □

Usually, the expressions on the right-hand sides of the inequalities in Theorem II.3 and Theorem II.4 are not needed explicitly. The triangle inequality is often applied to the $L^{\frac{p}{2}}(\Omega;\mathbb{R})$-norms in order to obtain suitable upper bounds. In the following, we show that for such inequalities the constants can be reduced in case of $p>2$ compared to the constants in Theorem II.3 and Theorem II.4.

Theorem II.5 (Discrete Burkholder-type Inequalities)
Let $p \in [2,\infty[$ and $N \in \mathbb{N}$. Consider a discrete martingale $(M_n)_{n\in\{0,1,\dots,N\}}$ in $L^p(\Omega;\mathbb{R}^d)$ with respect to the filtration $(\mathscr{F}_{t_n})_{n\in\{0,1,\dots,N\}}$, where $M_n = \sum_{k=0}^n d_k$ with $d_0 = M_0$ and $d_k = M_k - M_{k-1}$ for $k \in \{1,\dots,N\}$. Then, it holds

$$\|M_n\|^2_{L^p(\Omega;\mathbb{R}^d)} \leq (p-1)\sum_{k=0}^n \|d_k\|^2_{L^p(\Omega;\mathbb{R}^d)} \tag{II.4}$$

and

$$\Big\| \sup_{\nu\in\{0,1,\dots,n\}} \|M_\nu\| \Big\|^2_{L^p(\Omega;\mathbb{R})} \leq \frac{p^2}{p-1}\sum_{k=0}^n \|d_k\|^2_{L^p(\Omega;\mathbb{R}^d)}$$

for all $n \in \{0,1,\dots,N\}$. The constants are best possible.

Proof. The proof is stated in Section II.4, see p. 16. □

In 1967, Zakai proved inequality (II.5) of the following theorem in the case of $d = m = 1$, see [139, Theorem 1]. Since his paper is older than Burkholder's, we call, in honor of Zakai, the inequalities of the following theorem Zakai inequalities.

Theorem II.6 (Zakai Inequalities)
Let $p \in [2,\infty[$, and let $f^j \in H^p([t_0,T]\times\Omega;\mathbb{R}^d)$ for $j \in \{1,\dots,m\}$. Then, it holds

$$\Big\| \sum_{j=1}^m \int_{t_0}^t f^j_u \,\mathrm{d}W^j_u \Big\|^2_{L^p(\Omega;\mathbb{R}^d)} \leq (p-1)\int_{t_0}^t \Big\| \sum_{j=1}^m \|f^j_u\|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \mathrm{d}u \tag{II.5}$$

and

$$\Big\| \sum_{j=1}^m \int_{t_0}^{\cdot} f^j_u \,\mathrm{d}W^j_u \Big\|^2_{S^p([t_0,t]\times\Omega;\mathbb{R}^d)} \leq \frac{p^2}{p-1}\int_{t_0}^t \Big\| \sum_{j=1}^m \|f^j_u\|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \mathrm{d}u$$

for all $t \in [t_0,T]$.

Proof. The proof is stated in Section II.4, see p. 17. □

The smaller constants in the inequalities of Theorem II.5 and Theorem II.6 make the estimates highly valuable in stochastic analysis and stochastic numerics for example.

II.3. Strong Solutions

In this section, the existence and uniqueness of a strong solution X of SDDE (II.1) is shown. We further show that the solution is bounded in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$, Hölder continuous in time with order $\frac{1}{2}$ in $L^p(\Omega;\mathbb{R}^d)$, and Lipschitz continuous with respect to its initial condition in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$. But first, we state some inequalities that are used throughout this theses.

Hölder's inequality implies for $p \in [1, \infty[$ that

$$\left| \sum_{i=1}^{N} c_i \right|^p \leq N^{p-1} \sum_{i=1}^{N} |c_i|^p, \tag{II.6}$$

where $c_i \in \mathbb{R}$ and $N \in \mathbb{N}$. Moreover, using $(c_1 - c_2)^2 = c_1^2 - 2c_1c_2 + c_2^2$, we obtain the inequality

$$c_1 c_2 \leq \frac{1}{2} c_1^2 + \frac{1}{2} c_2^2 \tag{II.7}$$

for all $c_1, c_2 \in \mathbb{R}$. Further, the following lemma of Gronwall is frequently used.

Lemma II.7 (Gronwall's Lemma, [46, Lemma 2.6.1])
Let $f\colon [t_0, T] \to \mathbb{R}$ be Borel-measurable bounded function. Then, given a Borel-measurable bounded function $g\colon [t_0, T] \to \mathbb{R}$ and a constant $C > 0$ satisfying

$$f(t) \leq g(t) + C \int_{t_0}^{t} f(s) \, \mathrm{d}s,$$

it holds

$$f(t) \leq g(t) + C \int_{t_0}^{t} \mathrm{e}^{C(t-s)} g(s) \, \mathrm{d}s$$

for all $t \in [t_0, T]$.

Throughout this thesis, we impose a global Lipschitz and a linear growth condition on the Borel-measurable drift coefficient $a = (a^1, \ldots, a^d)^\mathrm{T}$ and diffusion coefficients $b^j = (b^{1,j}, \ldots, b^{d,j})^\mathrm{T}$, $j \in \{1, \ldots, m\}$, of SDDE (II.1). The SDDE's coefficients are said to satisfy the global *Lipschitz condition* if there exist constants $L_a, L_b > 0$ such that

$$\begin{aligned} &\sup_{t \in [t_0, T]} \| a(t, t - \tau_1, \ldots, t - \tau_D, x_0, x_1, \ldots, x_D) - a(t, t - \tau_1, \ldots, t - \tau_D, y_0, y_1, \ldots, y_D) \| \\ &\quad \leq L_a \max_{l \in \{0,1,\ldots,D\}} \| x_l - y_l \| \end{aligned} \tag{II.8}$$

and

$$\begin{aligned} &\sup_{t \in [t_0, T]} \max_{j \in \{1,\ldots,m\}} \| b^j(t, t - \tau_1, \ldots, t - \tau_D, x_0, x_1, \ldots, x_D) \\ &\qquad\qquad - b^j(t, t - \tau_1, \ldots, t - \tau_D, y_0, y_1, \ldots, y_D) \| \\ &\quad \leq L_b \max_{l \in \{0,1,\ldots,D\}} \| x_l - y_l \| \end{aligned} \tag{II.9}$$

for all $x_l, y_l \in \mathbb{R}^d$, $l \in \{0, 1, \ldots, D\}$, as well as the *linear growth condition* if there exist constants $K_a, K_b > 0$ such that

$$\sup_{t \in [t_0, T]} \| a(t, t - \tau_1, \ldots, t - \tau_D, x_0, x_1, \ldots, x_D) \| \leq K_a \max_{l \in \{0,1,\ldots,D\}} \left(1 + \|x_l\|^2\right)^{\frac{1}{2}} \tag{II.10}$$

and

$$\sup_{t\in[t_0,T]} \max_{j\in\{1,\dots,m\}} \|b^j(t,t-\tau_1,\dots,t-\tau_D,x_0,x_1,\dots,x_D)\| \le K_b \max_{l\in\{0,1,\dots,D\}} \left(1+\|x_l\|^2\right)^{\frac{1}{2}} \tag{II.11}$$

for all $x_l \in \mathbb{R}^d$, $l \in \{0,1,\dots,D\}$.

In [98, Theorem 5.2.2] and [107, Theorem II.2.1] for example, the existence and uniqueness of strong solutions of stochastic functional differential equations (SFDEs) are proven, also cf. [65], [119, Theorem V.3.7], and [122, Theorem IX.2.1]. Since SDDEs are a subclass of SFDEs, the existence and uniqueness of a strong solution of SDDEs follows immediately, see e. g. [98, Section 5.3]. Thus, the result of Theorem II.8 below is not entirely new, cf. [98, Theorem 5.2.2, p. 156 and Theorem 5.4.1] and [119, Theorem V.3.7].

Nevertheless, we present a proof of Theorem II.8 that is similar to the one of [98, Theorem 5.2.2] but takes the specific SDDE (II.1) as well as the global Lipschitz and linear growth conditions from above into account. Moreover, the realizations of the initial condition $\xi \in S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)$ may only be P-almost surely càdlàg and not continuous as in [98, Theorem 5.2.2] and [107, Theorem II.2.1], also cf. [107, Section VII.3]. In addition, stochastic processes are defined on a product space whereas the processes in [98, 107] are considered to be random variables with values in the space of continuous functions.

Theorem II.8 (Existence and Uniqueness of Strong Solutions)
Let the Borel-measurable drift a and diffusion b^j, $j \in \{1,\dots,m\}$, of SDDE (II.1) *satisfy the global Lipschitz and linear growth conditions* (II.8), (II.9), (II.10), *and* (II.11). *Moreover, let $\xi \in S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)$ for some $p \in [2,\infty[$.*

Then, there exists a unique (up to indistinguishability) stochastic process $X \in S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$, which is the strong solution of Itô SDDE (II.1) *with respect to the Wiener process W and initial condition ξ.*

Moreover, it holds

$$1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \le \left(1+2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\right)\mathrm{e}^{2\left(K_a\sqrt{T-t_0}+\frac{pK_b\sqrt{m}}{\sqrt{p-1}}\right)^2(T-t_0)}. \tag{II.12}$$

Proof. The proof is stated in Section II.4, see p. 18. □

Similarly to the result in [98, Theorem 5.4.3], we obtain the Hölder continuity with exponent $\frac{1}{2}$ of the solution in $L^p(\Omega;\mathbb{R}^d)$.

Lemma II.9
Let X be the strong solution of SDDE (II.1) *with initial condition ξ, and let the Borel-measurable coefficients a, b^j, $j \in \{1,\dots,m\}$, satisfy the linear growth conditions* (II.10) *and* (II.11). *Further, let $\xi \in S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)$ for some $p \in [2,\infty[$. Then, it holds*

$$\|X_t - X_s\|_{L^p(\Omega;\mathbb{R}^d)} \le \left(K_a\sqrt{T-t_0}+\sqrt{p-1}K_b\sqrt{m}\right)\left(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\right)^{\frac{1}{2}}\sqrt{|t-s|}$$

for all $s,t \in [t_0,T]$.

Proof. The proof is stated in Section II.4, see p. 24. □

Moreover, the solution X is, with respect to its initial condition, Lipschitz continuous in $S^p([t_0-\tau, T] \times \Omega; \mathbb{R}^d)$, cf. [107, Theorem II.3.1].

Lemma II.10
Let the Borel-measurable drift a and diffusion b^j, $j \in \{1, \ldots, m\}$, of SDDE (II.1) *satisfy the global Lipschitz and linear growth conditions* (II.8), (II.9), (II.10), *and* (II.11). *Moreover, let $\xi, \zeta \in S^p([t_0-\tau, t_0] \times \Omega; \mathbb{R}^d)$ for some $p \in [2, \infty[$, and let X^ξ and $X^\zeta \in S^p([t_0-\tau, T] \times \Omega; \mathbb{R}^d)$ be strong solutions of SDDE* (II.1) *with respect to the initial conditions ξ and ζ, respectively. Then, it holds*

$$\|X^\xi - X^\zeta\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq \sqrt{2}\|\xi-\zeta\|_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} \mathrm{e}^{\left(\sqrt{T-t_0}L_a+\frac{pL_b\sqrt{m}}{\sqrt{p-1}}\right)^2(T-t_0)}.$$

Proof. The proof is stated in Section II.4, see p. 25. □

Now, we extend Theorem II.8 to slightly more general SDDEs. This result will be needed in Chapter III for Theorem III.26. Let $Z\colon [t_0-\tau, T] \times \Omega \to \mathbb{R}^{\tilde{d}}$ be an $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$-progressively measurable process that realizations are P-almost surely càdlàg, where $\tilde{d} \in \mathbb{N}$. Further, let $A, B^j\colon \mathbb{R}^{1\times(D+1)} \times \mathbb{R}^{\tilde{d}\times(D+1)} \times \mathbb{R}^{d\times(D+1)}$ be Borel-measurable functions for $j \in \{1, \ldots, m\}$ and the initial condition ξ as in SDDE (II.1). Consider the SDDE

$$X_t = \begin{cases} \xi_t & \text{if } t \in [t_0-\tau, t_0], \\ \xi_{t_0} + \displaystyle\int_{t_0}^t A(\mathcal{T}(s, Z_s, X_s))\,\mathrm{d}s + \sum_{j=1}^m \int_{t_0}^t B^j(\mathcal{T}(s, Z_s, X_s))\,\mathrm{d}W_s^j & \text{if } t \in \,]t_0, T] \end{cases} \tag{II.13}$$

where

$$\mathcal{T}(t, Z_t, X_t) := (t, t-\tau_1, \ldots, t-\tau_D, Z_t, Z_{t-\tau_1}, \ldots, Z_{t-\tau_D}, X_t, X_{t-\tau_1}, \ldots, X_{t-\tau_D})$$

for all $t \in [t_0, T]$, cf. formula (II.3). Note that we recover SDDE (II.1) if the coefficients A and B^j, $j \in \{1, \ldots, d\}$ do not depend on the process Z. We consider the following definition of a strong solution for SDDE (II.13), which is a generalization of Definition II.2.

Definition II.11
A measurable stochastic process $X\colon [t_0-\tau, T]\times\Omega \to \mathbb{R}^d$ is called strong solution *of SDDE* (II.13) *with respect to the fixed Wiener process W, initial condition ξ, and process Z if*

i) $(X_t)_{t\in[t_0-\tau,T]}$ *is adapted to* $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$,

ii) $(X_t)_{t\in[t_0,T]}$ *has* P*-almost surely continuous realizations,*

iii) $\displaystyle\int_{t_0}^t \|A(\mathcal{T}(s, Z_s, X_s))\| + \sum_{j=1}^m \|B^j(\mathcal{T}(s, Z_s, X_s))\|^2\,\mathrm{d}s < \infty$
holds P*-almost surely for all* $t \in [t_0, T]$,

iv) and if the equation (II.13) *holds* P*-almost surely for all* $t \in [t_0-\tau, T]$.

Similarly to the global Lipschitz and linear growth conditions (II.8), (II.9), (II.10), and (II.11), the following conditions on the coefficients of SDDE (II.13) are considered. The coefficients of SDDE (II.13) are said to satisfy the global *Lipschitz condition* if there exist constants $L_A, L_B > 0$ such that

$$\begin{aligned}\sup_{\substack{t\in[t_0,T]\\ z_l\in\mathbb{R}^d:\ l\in\{0,1,\ldots,D\}}} &\|A(t,t-\tau_1,\ldots,t-\tau_D,z_0,z_1,\ldots,z_D,x_0,x_1,\ldots,x_D)\\ &\quad - A(t,t-\tau_1,\ldots,t-\tau_D,z_0,z_1,\ldots,z_D,y_0,y_1,\ldots,y_D)\|\\ \leq L_A &\max_{l\in\{0,1,\ldots,D\}}\|x_l-y_l\|\end{aligned} \tag{II.14}$$

and

$$\begin{aligned}\sup_{\substack{t\in[t_0,T]\\ z_l\in\mathbb{R}^d:\ l\in\{0,1,\ldots,D\}}} &\max_{j\in\{1,\ldots,m\}}\|B^j(t,t-\tau_1,\ldots,t-\tau_D,z_0,z_1,\ldots,z_D,x_0,x_1,\ldots,x_D)\\ &\quad - B^j(t,t-\tau_1,\ldots,t-\tau_D,z_0,z_1,\ldots,z_D,y_0,y_1,\ldots,y_D)\|\\ \leq L_B &\max_{l\in\{0,1,\ldots,D\}}\|x_l-y_l\|\end{aligned} \tag{II.15}$$

for all $x_l, y_l \in \mathbb{R}^d$, $l \in \{0,1,\ldots,D\}$, as well as the *linear growth condition* if there exist constants $K_A, K_B > 0$ such that

$$\begin{aligned}\sup_{\substack{t\in[t_0,T]\\ z_l\in\mathbb{R}^d:\ l\in\{0,1,\ldots,D\}}} &\|A(t,t-\tau_1,\ldots,t-\tau_D,z_0,z_1,\ldots,z_D,x_0,x_1,\ldots,x_D)\|\\ \leq K_A &\max_{l\in\{0,1,\ldots,D\}}\left(1+\|x_l\|^2\right)^{\frac{1}{2}}\end{aligned} \tag{II.16}$$

and

$$\begin{aligned}\sup_{\substack{t\in[t_0,T]\\ z_l\in\mathbb{R}^d:\ l\in\{0,1,\ldots,D\}}} &\max_{j\in\{1,\ldots,m\}}\|B^j(t,t-\tau_1,\ldots,t-\tau_D,z_0,z_1,\ldots,z_D,x_0,x_1,\ldots,x_D)\|\\ \leq K_B &\max_{l\in\{0,1,\ldots,D\}}\left(1+\|x_l\|^2\right)^{\frac{1}{2}}\end{aligned} \tag{II.17}$$

for all $x_l \in \mathbb{R}^d$, $l \in \{0,1,\ldots,D\}$. We obtain the following existence and uniqueness theorem for SDDE (II.13).

Theorem II.12 (Existence and Uniqueness of Strong Solutions)
Let the Borel-measurable drift A and diffusion B^j, $j \in \{1,\ldots,m\}$, of SDDE (II.13) *satisfy the global Lipschitz and linear growth conditions* (II.14), (II.15), (II.16), *and* (II.17). *Moreover, let $Z\colon [t_0-\tau,T]\times\Omega \to \mathbb{R}^{\tilde{d}}$ be an $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$-progressively measurable process that realizations are* P*-almost surely càdlàg, where $\tilde{d} \in \mathbb{N}$, and let $\xi \in S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)$ for some $p \in [2,\infty[$.*

Then, there exists a unique (up to indistinguishability) stochastic process $X \in S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$, which is the strong solution of Itô SDDE (II.13) *with respect to the Wiener process W, initial condition ξ, and process Z.*

Moreover, it holds

$$1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq \left(1+2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\right)\mathrm{e}^{2\left(K_A\sqrt{T-t_0}+\frac{pK_B\sqrt{m}}{\sqrt{p-1}}\right)^2(T-t_0)}. \tag{II.18}$$

Proof. Due to the linear growth conditions (II.16) and (II.17), no assumptions on the boundedness of process Z have to be made, and the proof is completely analogous to the proof of Theorem II.8. $\square$

In case of SODEs, a similar theorem is stated in [113, Lemma 2.2.1]. There, the SODEs' coefficients are allowed to have a polynomial growth regarding the argument of the process Z. In return, some boundedness assumption on the process Z is supposed. Mohammed considered in [107, Theorem V.4.3] the existence and uniqueness of SFDEs whose coefficients are allowed to be random. However, he supposed that the coefficients are $\mathscr{F}_{t_0}$-measurable, and our theorem above is thus more general in case of SDDEs.

II.4. Proofs

Proof of Theorem II.5

Proof of Theorem II.5. In the case of $p = 2$, the statement follows from the discrete Burkholder inequalities in Theorem II.3, and therefore we assume $p \in \,]2, \infty[$ in the following. Inequality (II.4) and its sharpness are proven in [123, Section 2] by Rio in the case of $d = n = 1$.

We amend his proof to general $d \in \mathbb{N}$ and $n \in \{1, \dots, N\}$ with $N \in \mathbb{N}$. For this, we generalize [123, Proposition 2.1]. Let $X, Y \in L^p(\Omega; \mathbb{R}^d)$, and let $\mathcal{F} \in \mathscr{F}$ be some sub-σ-algebra such that X is $\mathcal{F}/\mathscr{B}(\mathbb{R}^d)$-measurable and $\mathrm{E}[Y|\mathcal{F}] = 0$ P-almost surely. Then, we first prove that

$$\|X + Y\|^2_{L^p(\Omega;\mathbb{R}^d)} \leq \|X\|^2_{L^p(\Omega;\mathbb{R}^d)} + (p-1)\|Y\|^2_{L^p(\Omega;\mathbb{R}^d)}. \tag{II.19}$$

If $X = 0$ or $Y = 0$, this inequality is clearly true, so we assume $\|X\|_{L^p(\Omega;\mathbb{R}^d)} > 0$ and $\|Y\|_{L^p(\Omega;\mathbb{R}^d)} > 0$. Define the function $\varphi\colon [0,1] \to \mathbb{R}$ by $\varphi(t) = \|x + ty\|^p$, where $x, y \in \mathbb{R}^d$. Using Taylor's formula [57, p. 284], it holds

$$\varphi(1) = \varphi(0) + \varphi'(0) + \int_0^1 \varphi''(t)(1-t)\,\mathrm{d}t$$

and thus

$$\begin{aligned}\|X + Y\|^p = \|X\|^p + p\|X\|^{p-2}\sum_{i=1}^d X^i Y^i + p\int_0^1 \|X + tY\|^{p-2}\|Y\|^2(1-t)\,\mathrm{d}t \\ + p(p-2)\int_0^1 \|X + tY\|^{p-4}\left(\sum_{i=1}^d (X^i + tY^i)Y^i\right)^2 (1-t)\,\mathrm{d}t.\end{aligned}$$

Considering the integrand of the last integral in the Taylor formula above, it holds

$$\|X + tY\|^{p-4}\left(\sum_{i=1}^d (X^i + tY^i)Y^i\right)^2 \leq \|X + tY\|^{p-2}\|Y\|^2$$

by Cauchy-Schwarz inequality, and due to this, it follows

$$\|X + Y\|^p \leq \|X\|^p + p\|X\|^{p-2}\sum_{i=1}^d X^i Y^i + p(p-1)\int_0^1 \|X + tY\|^{p-2}\|Y\|^2(1-t)\,\mathrm{d}t. \tag{II.20}$$

According to the assumptions, it holds

$$\mathrm{E}\big[\mathrm{E}[\|X\|^{p-2}X^iY^i|\mathcal{F}]\big] = \mathrm{E}\big[\|X\|^{p-2}X^i\mathrm{E}[Y^i|\mathcal{F}]\big] = 0.$$

Then, since

$$\mathrm{E}\big[\|X+tY\|^{p-2}\|Y\|^2\big] \le \big(\mathrm{E}[\|X+tY\|^p]\big)^{\frac{p-2}{p}}\big(\mathrm{E}[\|Y\|^p]\big)^{\frac{2}{p}}$$

by Hölder's inequality with $\frac{p-2}{p}+\frac{2}{p}=1$, we obtain by taking the expectation on both sides of inequality (II.20) and using Fubini's theorem that

$$\mathrm{E}[\|X+Y\|^p] \le \mathrm{E}[\|X\|^p] + p(p-1)\int_0^1 \big(\mathrm{E}[\|X+tY\|^p]\big)^{\frac{p-2}{p}}\big(\mathrm{E}[\|Y\|^p]\big)^{\frac{2}{p}}(1-t)\,\mathrm{d}t.$$

This is a multidimensional version of [123, Inequality (2.1)]. Next, we use a Gronwall-type inequality, that is, we apply [139, Lemma on p. 171] with $\alpha = \frac{2}{p}$. It follows

$$\mathrm{E}[\|X+Y\|^p] \le \left(\mathrm{E}[\|X\|^p]^{\frac{2}{p}} + \frac{2}{p}p(p-1)\int_0^1 (1-t)\,\mathrm{d}t\big(\mathrm{E}[\|Y\|^p]\big)^{\frac{2}{p}}\right)^{\frac{p}{2}},$$

and since $\int_0^1(1-t)\,\mathrm{d}t = \frac{1}{2}$, inequality (II.19) holds by raising both sides of the inequality above to the power of $\frac{2}{p}$. Due to [123, Remark 2.1], the constant $p-1$ in inequality (II.19) is best possible.

We remark that the considerations after [123, Inequality (2.1)] on [123, p. 150] prove essentially Zakai's Gronwall-type inequality in [139, Lemma on p. 171].

Now, we consider inequality (II.4). Since $(M_n)_{n\in\{0,1,\dots,N\}}$ is a martingale, it holds $\mathrm{E}[M_n|\mathscr{F}_{t_{n-1}}] = M_{n-1}$ P-almost surely for $n \in \{1,\dots,N\}$, that is, $\mathrm{E}[d_n|\mathscr{F}_{t_{n-1}}] = 0$ P-almost surely. Thus, inequality (II.4) follows from applying inequality (II.19) to $M_n = M_{n-1} + d_n$ and by induction over $n \in \{1,\dots,N\}$, cf. [123, Theorem 2.1].

Finally, Doob's maximal inequality [35, Theorem 3.4 on p. 317] implies the second inequality of this theorem. Since Doob's inequality is sharp, the constant is best possible, cf. [19, p. 87] and [36, Theorem 2]. □

Proof of Theorem II.6

Proof of Theorem II.6. The first inequality is proven by Zakai [139, Theorem 1] in case of $d = m = 1$. We extend his proof to the case of m-dimensional Wiener processes and $\mathbb{R}^d$-valued integrands. Due to Burkholder's inequality in Theorem II.4 and the assumption $f^j \in H^p([t_0,T]\times\Omega;\mathbb{R}^d)$ for $j \in \{1,\dots,m\}$, we have

$$\Big\|\sum_{j=1}^m \int_{t_0}^t f_u^j\,\mathrm{d}W_u^j\Big\|_{L^p(\Omega;\mathbb{R}^d)} < \infty.$$

Let $\delta > 0$ and $\varphi \in C^2(\mathbb{R}^d;\mathbb{R})$ with $\varphi(x) = (\delta + \|x\|^2)^{\frac{p}{2}}$ for $x \in \mathbb{R}^d$. Using Itô's formula, see e. g. [64] or [75, p. 153], with function φ, [139, Equation (6)] reads in the multidimensional case as

$$
\begin{aligned}
&\left(\delta + \left\| \sum_{j=1}^{m} \int_{t_0}^{t} f_u^j \,\mathrm{d}W_u^j \right\|^2 \right)^{\frac{p}{2}} - \delta^{\frac{p}{2}} \\
&= \frac{p}{2} \int_{t_0}^{t} \left(\delta + \left\| \sum_{j=1}^{m} \int_{t_0}^{s} f_u^j \,\mathrm{d}W_u^j \right\|^2 \right)^{\frac{p}{2}-1} \sum_{j=1}^{m} \sum_{i=1}^{d} (f_s^{i,j})^2 \,\mathrm{d}s \\
&\quad + \frac{p(p-2)}{2} \int_{t_0}^{t} \left(\delta + \left\| \sum_{j=1}^{m} \int_{t_0}^{s} f_u^j \,\mathrm{d}W_u^j \right\|^2 \right)^{\frac{p}{2}-2} \\
&\qquad \times \sum_{l=1}^{m} \sum_{i,k=1}^{d} \left(\sum_{j=1}^{m} \int_{t_0}^{s} f_u^{i,j} \,\mathrm{d}W_u^j \right) \left(\sum_{j=1}^{m} \int_{t_0}^{s} f_u^{k,j} \,\mathrm{d}W_u^j \right) f_s^{i,l} f_s^{k,l} \,\mathrm{d}s \\
&\quad + p \sum_{l=1}^{m} \int_{t_0}^{t} \left(\delta + \left\| \sum_{j=1}^{m} \int_{t_0}^{s} f_u^j \,\mathrm{d}W_u^j \right\|^2 \right)^{\frac{p}{2}-1} \sum_{i=1}^{d} \left(\sum_{j=1}^{m} \int_{t_0}^{s} f_u^{i,j} \,\mathrm{d}W_u^j \right) f_s^{i,l} \,\mathrm{d}W_s^l
\end{aligned}
$$

P-almost surely. Taking the expectation and using the Cauchy-Schwarz inequality, we obtain

$$
\begin{aligned}
&\mathrm{E}\left[\left(\delta + \left\| \sum_{j=1}^{m} \int_{t_0}^{t} f_u^j \,\mathrm{d}W_u^j \right\|^2 \right)^{\frac{p}{2}} \right] - \delta^{\frac{p}{2}} \\
&= \frac{p}{2} \int_{t_0}^{t} \mathrm{E}\left[\left(\delta + \left\| \sum_{j=1}^{m} \int_{t_0}^{s} f_u^j \,\mathrm{d}W_u^j \right\|^2 \right)^{\frac{p}{2}-1} \sum_{j=1}^{m} \|f_s^j\|^2 \right] \mathrm{d}s \\
&\quad + \frac{p(p-2)}{2} \int_{t_0}^{t} \mathrm{E}\left[\left(\delta + \left\| \sum_{j=1}^{m} \int_{t_0}^{s} f_u^j \,\mathrm{d}W_u^j \right\|^2 \right)^{\frac{p}{2}-2} \sum_{l=1}^{m} \sum_{i=1}^{d} \left(\sum_{j=1}^{m} \int_{t_0}^{s} f_u^{i,j} \,\mathrm{d}W_u^j \, f_s^{i,l} \right)^2 \right] \mathrm{d}s \\
&\leq \frac{p(p-1)}{2} \int_{t_0}^{t} \mathrm{E}\left[\left(\delta + \left\| \sum_{j=1}^{m} \int_{t_0}^{s} f_u^j \,\mathrm{d}W_u^j \right\|^2 \right)^{\frac{p}{2}-1} \sum_{j=1}^{m} \|f_s^j\|^2 \right] \mathrm{d}s,
\end{aligned}
$$

which corresponds to the multidimensional variant of [139, Inequality (8)]. Then, the inequality (II.5) follows from the same arguments as in [139, pp. 171–172] by applying a Gronwall-type inequality [139, Lemma on p. 171] and letting $\delta \to 0$.

Applying Doob's submartingale inequality, see e. g. [122, Theorem II.1.7], then yields the second inequality of this theorem. □

Proof of Theorem II.8

The proof of Theorem II.8 is similar to the proof of [98, Theorem 5.2.2].

Proof of Theorem II.8. We start with the proof of uniqueness. Assume that X and $\hat{X}$ are two strong solutions of SDDE (II.1) with respect to the same Wiener process W and the same initial condition $\xi \in S^p([t_0 - \tau, t_0] \times \Omega; \mathbb{R}^d)$.

Let $n \in \mathbb{N}$, and define the $(\mathscr{F}_t)_{t\in[t_0,T]}$-stopping time $\sigma_n \colon \Omega \to [t_0,T]$ by

$$\sigma_n \coloneqq \inf\{t \in [t_0,T] : \|X_t\| \geq n\} \wedge \inf\{t \in [t_0,T] : \|\hat{X}_t\| \geq n\}$$

where $\inf\{\emptyset\} \coloneqq T$. The stopped processes $(X_{t\wedge\sigma_n})_{t\in[t_0-\tau,T]}$ and $(\hat{X}_{t\wedge\sigma_n})_{t\in[t_0-\tau,T]}$ are bounded, and thus they belong to $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$, cf. Section II.1. As P-almost all realizations of $(X_t)_{t\in[t_0,T]}$ and $(\hat{X}_t)_{t\in[t_0,T]}$ are continuous, it holds $\lim_{n\to\infty}\sigma_n = T$ P-almost surely.

At first, we show that the stopped processes $X_{\cdot\wedge\sigma_n}$ and $\hat{X}_{\cdot\wedge\sigma_n}$ are modifications of each other for all $n \in \mathbb{N}$. Since $X_t = \hat{X}_t$ for all $t \in [t_0-\tau,t_0]$ P-almost surely, it holds

$$\|X_{\cdot\wedge\sigma_n} - \hat{X}_{\cdot\wedge\sigma_n}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = \|X_{\cdot\wedge\sigma_n} - \hat{X}_{\cdot\wedge\sigma_n}\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}.$$

Using that the solutions X and $\hat{X}$ satisfy the equation (II.1) and applying the triangle inequality, we obtain by rewriting

$$\begin{aligned}
&\|X_{\cdot\wedge\sigma_n} - \hat{X}_{\cdot\wedge\sigma_n}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq \Big\| \int_{t_0}^{\cdot\wedge\sigma_n} a(\mathcal{T}(s,X_s)) - a(\mathcal{T}(s,\hat{X}_s))\,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad + \Big\| \sum_{j=1}^m \int_{t_0}^{\cdot\wedge\sigma_n} b^j(\mathcal{T}(s,X_s)) - b^j(\mathcal{T}(s,\hat{X}_s))\,\mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad= \Big\| \int_{t_0}^{\cdot} \mathbb{1}_{[t_0,\sigma_n]}(s)\big(a(\mathcal{T}(s,X_{s\wedge\sigma_n})) - a(\mathcal{T}(s,\hat{X}_{s\wedge\sigma_n}))\big)\,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad + \Big\| \sum_{j=1}^m \int_{t_0}^{\cdot} \mathbb{1}_{[t_0,\sigma_n]}(s)\big(b^j(\mathcal{T}(s,X_{s\wedge\sigma_n})) - b^j(\mathcal{T}(s,\hat{X}_{s\wedge\sigma_n}))\big)\,\mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}.
\end{aligned} \tag{II.21}$$

Considering the first term on the right-hand side of inequality (II.21) above, the triangle inequality and Lipschitz condition (II.8) imply

$$\begin{aligned}
&\Big\| \int_{t_0}^{\cdot} \mathbb{1}_{[t_0,\sigma_n]}(s)\big(a(\mathcal{T}(s,X_{s\wedge\sigma_n})) - a(\mathcal{T}(s,\hat{X}_{s\wedge\sigma_n}))\big)\,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq \Big\| \sup_{t\in[t_0,T]} \int_{t_0}^{t} \big\|\mathbb{1}_{[t_0,\sigma_n]}(s)\big(a(\mathcal{T}(s,X_{s\wedge\sigma_n})) - a(\mathcal{T}(s,\hat{X}_{s\wedge\sigma_n}))\big)\big\|\,\mathrm{d}s \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \int_{t_0}^{T} \|a(\mathcal{T}(s,X_{s\wedge\sigma_n})) - a(\mathcal{T}(s,\hat{X}_{s\wedge\sigma_n}))\|_{L^p(\Omega;\mathbb{R}^d)}\,\mathrm{d}s \\
&\quad\leq L_a \int_{t_0}^{T} \Big\| \sup_{l\in\{0,1,\ldots,D\}} \|X_{s\wedge\sigma_n-\tau_l} - \hat{X}_{s\wedge\sigma_n-\tau_l}\| \Big\|_{L^p(\Omega;\mathbb{R})}\,\mathrm{d}s \\
&\quad\leq L_a \int_{t_0}^{T} \|X_{\cdot\wedge\sigma_n} - \hat{X}_{\cdot\wedge\sigma_n}\|_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s \\
&\quad\leq L_a\sqrt{T-t_0}\bigg(\int_{t_0}^{T} \|X_{\cdot\wedge\sigma_n} - \hat{X}_{\cdot\wedge\sigma_n}\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s \bigg)^{\frac{1}{2}},
\end{aligned} \tag{II.22}$$

where the Cauchy-Schwarz inequality is used in the last step. Similar considerations for the second term on the right-hand side of inequality (II.21) yield with the Zakai inequality from

Theorem II.6 and Lipschitz condition (II.9) that

$$\begin{aligned}
&\Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} \mathbb{1}_{[t_0,\sigma_n]}(s)\big(b^j(\mathcal{T}(s, X_{s\wedge\sigma_n})) - b^j(\mathcal{T}(s, \hat{X}_{s\wedge\sigma_n}))\big)\, \mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le \frac{p}{\sqrt{p-1}} \bigg(\int_{t_0}^{T} \Big\| \sum_{j=1}^{m} \big\| \mathbb{1}_{[t_0,\sigma_n]}(s)\big(b^j(\mathcal{T}(s, X_{s\wedge\sigma_n})) - b^j(\mathcal{T}(s, \hat{X}_{s\wedge\sigma_n}))\big)\big\|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \mathrm{d}s \bigg)^{\frac{1}{2}} \\
&\quad \le \frac{p}{\sqrt{p-1}} \bigg(\int_{t_0}^{T} \sum_{j=1}^{m} \| b^j(\mathcal{T}(s, X_{s\wedge\sigma_n})) - b^j(\mathcal{T}(s, \hat{X}_{s\wedge\sigma_n}))\|^2_{L^p(\Omega;\mathbb{R}^d)} \, \mathrm{d}s \bigg)^{\frac{1}{2}} \\
&\quad \le \frac{pL_b\sqrt{m}}{\sqrt{p-1}} \bigg(\int_{t_0}^{T} \| X_{\cdot\wedge\sigma_n} - \hat{X}_{\cdot\wedge\sigma_n} \|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \, \mathrm{d}s \bigg)^{\frac{1}{2}}. \qquad \text{(II.23)}
\end{aligned}$$

Inserting estimates (II.22) and (II.23) into inequality (II.21) and squaring the result, we obtain

$$\begin{aligned}
&\| X_{\cdot\wedge\sigma_n} - \hat{X}_{\cdot\wedge\sigma_n} \|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le \bigg(L_a\sqrt{T-t_0} + \frac{pL_b\sqrt{m}}{\sqrt{p-1}} \bigg)^2 \int_{t_0}^{T} \| X_{\cdot\wedge\sigma_n} - \hat{X}_{\cdot\wedge\sigma_n} \|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \, \mathrm{d}s. \qquad \text{(II.24)}
\end{aligned}$$

Then, Gronwall's Lemma II.7 implies $\|X_{\cdot\wedge\sigma_n} - \hat{X}_{\cdot\wedge\sigma_n}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = 0$. Hence, for all $n \in \mathbb{N}$ the stopped processes $X_{\cdot\wedge\sigma_n}$ and $\hat{X}_{\cdot\wedge\sigma_n}$ are particularly modifications of each other. That is

$$\mathrm{P}[X_{t\wedge\sigma_n} = \hat{X}_{t\wedge\sigma_n}] = 1$$

for all $t \in [t_0 - \tau, T]$ and $n \in \mathbb{N}$. Due to this, it holds for all $t \in [t_0 - \tau, T]$ and $n \in \mathbb{N}$ that

$$\begin{aligned}
\mathrm{P}[X_t \neq \hat{X}_t] &= \mathrm{P}\bigg[\Big\{ \sup_{t\in[t_0,T]} \|X_t\| > n \Big\} \cup \Big\{ \sup_{t\in[t_0,T]} \|\hat{X}_t\| > n \Big\} \bigg] \\
&\le \mathrm{P}\Big[\sup_{t\in[t_0,T]} \|X_t\| > n \Big] + \mathrm{P}\Big[\sup_{t\in[t_0,T]} \|\hat{X}_t\| > n \Big].
\end{aligned}$$

Since the solutions $(X_t)_{t\in[t_0,T]}$ and $(\hat{X})_{t\in[t_0,T]}$ have P-almost surely continuous realizations, $\sup_{t\in[t_0,T]}\|X_t\|$ and $\sup_{t\in[t_0,T]}\|\hat{X}_t\|$ are P-almost surely bounded. Hence, for all $\varepsilon > 0$, an $N \in \mathbb{N}$ exists such that for all $n \ge N$ we have

$$\mathrm{P}[X_t \neq \hat{X}_t] \le \mathrm{P}\Big[\sup_{t\in[t_0,T]} \|X_t\| > n \Big] + \mathrm{P}\Big[\sup_{t\in[t_0,T]} \|\hat{X}_t\| > n \Big] < \varepsilon$$

for all $t \in [t_0 - \tau, T]$, that is, the solutions X and $\hat{X}$ are modification of each others, cf. [7, p. 107] and [47, p. 394]. Using that P-almost all realizations of X and $\hat{X}$ are càdlàg, both solutions X and $\hat{X}$ are indistinguishable [119, Corollary of Theorem I.2]. That is, if there exists a solution of the SDDE (II.1), the solution is unique up to indistinguishability.

In the following, the existence of the strong solution X of SDDE (II.1) is proven using the Picard's iterations. Let

$$X_t^{(0)} := \begin{cases} \xi_t & \text{if } t \in [t_0 - \tau, t_0] \text{ and} \\ \xi_{t_0} & \text{if } t \in \,]t_0, T] \end{cases}$$

as well as

$$X_t^{(\ell+1)} := \begin{cases} \xi_t & \text{if } t \in [t_0-\tau, t_0] \text{ and} \\ \xi_{t_0} + \int_{t_0}^t a(\mathcal{T}(s, X_s^{(\ell)}))\,\mathrm{d}s + \sum_{j=1}^m \int_{t_0}^t b^j(\mathcal{T}(s, X_s^{(\ell)}))\,\mathrm{d}W_s^j & \text{if } t \in\,]t_0, T] \end{cases}$$

for $\ell \in \mathbb{N}_0$. Since $\xi \in S^p([t_0-\tau, t_0] \times \Omega; \mathbb{R}^d)$, and since coefficients a and b^j, $j \in \{1, \ldots, m\}$, satisfy the linear growth conditions (II.10) and (II.11), also cf. Section II.1, it is evident by induction over $\ell \in \mathbb{N}_0$ that $X^\ell \in S^p([t_0-\tau, T] \times \Omega; \mathbb{R}^d)$ for every $\ell \in \mathbb{N}_0$. In order to be more precise regarding the upper bound of $X^{(\ell)}$ in $S^p([t_0-\tau, T] \times \Omega; \mathbb{R}^d)$, we make the following considerations.

Using inequality (II.6), we have

$$\begin{aligned} 1 &+ \|X^{(\ell+1)}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\ &\leq 1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} + 2\Bigg(\Bigg\| \int_{t_0}^{\cdot} a(\mathcal{T}(s, X_s^{(\ell)}))\,\mathrm{d}s \Bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\ &\qquad + \Bigg\| \sum_{j=1}^m \int_{t_0}^{\cdot} b^j(\mathcal{T}(s, X_s^{(\ell)}))\,\mathrm{d}W_s^j \Bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}\Bigg)^2 \end{aligned}$$

for all $\ell \in \mathbb{N}_0$, where

$$\begin{aligned} \Bigg\| &\int_{t_0}^{\cdot} a(\mathcal{T}(s, X_s^{(\ell)}))\,\mathrm{d}s \Bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\ &\leq \Bigg\| \sup_{t\in[t_0,T]} \int_{t_0}^t \|a(\mathcal{T}(s, X_s^{(\ell)}))\|\,\mathrm{d}s \Bigg\|_{L^p(\Omega;\mathbb{R})} \\ &\leq K_a \int_{t_0}^T \Bigg\| \sup_{l\in\{0,1,\ldots,D\}} \big(1 + \|X_{s-\tau_l}^{(\ell)}\|^2\big)^{\frac{1}{2}} \Bigg\|_{L^p(\Omega;\mathbb{R})} \mathrm{d}s \\ &\leq K_a \int_{t_0}^T \big(1 + \|X^{(\ell)}\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\,\mathrm{d}s \\ &\leq K_a\sqrt{T-t_0}\Bigg(\int_{t_0}^T 1 + \|X^{(\ell)}\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s\Bigg)^{\frac{1}{2}} \end{aligned} \tag{II.25}$$

by triangle inequality, linear growth condition (II.10), and Cauchy-Schwarz inequality, as well as

$$\begin{aligned} \Bigg\| &\sum_{j=1}^m \int_{t_0}^{\cdot} b^j(\mathcal{T}(s, X_s^{(\ell)}))\,\mathrm{d}W_s^j \Bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\ &\leq \frac{p}{\sqrt{p-1}}\Bigg(\int_{t_0}^T \Bigg\| \sum_{j=1}^m \|b^j(\mathcal{T}(s, X_s^{(\ell)}))\|^2 \Bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \mathrm{d}s\Bigg)^{\frac{1}{2}} \\ &\leq \frac{pK_b\sqrt{m}}{\sqrt{p-1}} \int_{t_0}^T \Bigg\| \sup_{l\in\{0,1,\ldots,D\}} \big(1 + \|X_{s-\tau_l}^{(\ell)}\|^2\big) \Bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \mathrm{d}s \\ &\leq \frac{pK_b\sqrt{m}}{\sqrt{p-1}} \Bigg(\int_{t_0}^T 1 + \|X^{(\ell)}\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s\Bigg)^{\frac{1}{2}} \end{aligned} \tag{II.26}$$

by Zakai's inequality from Theorem II.6 and linear growth condition (II.11). Thus, we have

$$\begin{aligned}
&1 + \|X^{(\ell+1)}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \leq 1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} \\
&\qquad + 2\bigg(K_a\sqrt{T-t_0} + \frac{pK_b\sqrt{m}}{\sqrt{p-1}}\bigg)^2 \int_{t_0}^T 1 + \big\|X^{(\ell)}\big\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s
\end{aligned}$$

that inductively ensures $\|X^{(\ell+1)}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} < \infty$. Now let $n \in \mathbb{N}_0$ be arbitrary fixed. Since

$$\begin{aligned}
&1 + \max_{\ell\in\{0,1,\ldots,n\}} \|X^{(\ell+1)}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \leq 1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} \\
&\qquad + 2\bigg(\sqrt{T-t_0}K_a + \frac{pK_b\sqrt{m}}{\sqrt{p-1}}\bigg)^2 \int_{t_0}^T 1 + \max_{\ell\in\{0,1,\ldots,n\}} \|X^{(\ell)}\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s,
\end{aligned}$$

it also holds

$$\begin{aligned}
&1 + \max_{\ell\in\{0,1,\ldots,n\}} \|X^{(\ell)}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \leq 1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} \\
&\qquad + 2\bigg(\sqrt{T-t_0}K_a + \frac{pK_b\sqrt{m}}{\sqrt{p-1}}\bigg)^2 \int_{t_0}^T 1 + \max_{\ell\in\{0,1,\ldots,n\}} \|X^{(\ell)}\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s,
\end{aligned}$$

and Gronwall's Lemma II.7 implies

$$\begin{aligned}
&1 + \max_{\ell\in\{0,1,\ldots,n\}} \|X^{(\ell)}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \leq \big(1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\big)\mathrm{e}^{2\left(\sqrt{T-t_0}K_a + \frac{pK_b\sqrt{m}}{\sqrt{p-1}}\right)^2(T-t_0)}.
\end{aligned}$$

As the right-hand side of the inequality above does not depend on $n \in \mathbb{N}_0$, we obtain

$$1 + \|X^{(\ell)}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq \big(1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\big)\mathrm{e}^{2\left(\sqrt{T-t_0}K_a + \frac{pK_b\sqrt{m}}{\sqrt{p-1}}\right)^2(T-t_0)} \tag{II.27}$$

for all $\ell \in \mathbb{N}_0$. Using the triangle inequality, Zakai's inequality from Theorem II.6, linear growth conditions (II.10) and (II.11) as well as the Cauchy-Schwarz inequality, it then follows by inequalities (II.25) and (II.26) that

$$\begin{aligned}
&\|X^{(1)} - X^{(0)}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \leq \Bigg(\bigg\|\int_{t_0}^{\cdot} a(\mathcal{T}(s, X_s^{(0)}))\,\mathrm{d}s\bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} + \bigg\|\sum_{j=1}^m \int_{t_0}^{\cdot} b^j(\mathcal{T}(s, X_s^{(0)}))\,\mathrm{d}W_s^j\bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}\Bigg)^2 \\
&\quad \leq \bigg(K_a\sqrt{T-t_0} + \frac{pK_b\sqrt{m}}{\sqrt{p-1}}\bigg)^2\big(1 + \|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\big)(T-t_0).
\end{aligned}$$

Similarly to inequality (II.24), the triangle inequality, Zakai's inequality from Theorem II.6, Lipschitz conditions (II.8) and (II.9) as well as the Cauchy-Schwarz inequality yield

$$\begin{aligned}
&\|X^{(\ell+1)} - X^{(\ell)}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \leq \bigg(L_a\sqrt{T-t_0} + \frac{pL_b\sqrt{m}}{\sqrt{p-1}}\bigg)^2 \int_{t_0}^T \|X^{(\ell)} - X^{(\ell-1)}\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s
\end{aligned} \tag{II.28}$$

for $\ell \in \mathbb{N}$. Raising inequality (II.28) to the $\frac{p}{2}$th power and using Hölder's inequality, it inductively holds

$$\|X^{(\ell+1)} - X^{(\ell)}\|^p_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq \|X^{(1)} - X^{(0)}\|^p_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \frac{\big(C(T-t_0)\big)^\ell}{\ell!}$$

for all $\ell \in \mathbb{N}_0$ where

$$C := \left(L_a\sqrt{T-t_0} + \frac{pL_b\sqrt{m}}{\sqrt{p-1}}\right)^p (T-t_0)^{\frac{p-2}{2}}.$$

Then, Markov's inequality yields

$$\begin{aligned}
&\mathrm{P}\left[\sup_{t\in[t_0-\tau,T]} \|X_t^{(\ell+1)} - X_t^{(\ell)}\| \geq 2^{-(\ell+1)}\right] \\
&\quad\leq 2^p\|X^{(1)} - X^{(0)}\|^p_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \frac{\big(2^pC(T-t_0)\big)^\ell}{\ell!}.
\end{aligned}$$

Using Borel-Cantelli lemma, there exists an $\Omega_1 \subset \Omega$ with $\mathrm{P}[\Omega_1] = 1$ and $N_\omega \in \mathbb{N}_0$, $\omega \in \Omega_1$, such that for every $\omega \in \Omega_1$ and $\ell \geq N_\omega$, it holds

$$\sup_{t\in[t_0-\tau,T]} \|X_t^{(\ell+1)}(\omega) - X_t^{(\ell)}(\omega)\| < 2^{-(\ell+1)}.$$

Due to Weierstraß's convergence criterion [126, Theorem 7.10], the sequence $(X^{(\ell)}(\omega))_{\ell\in\mathbb{N}_0}$ of realizations $X^{(\ell)}(\omega) = X^{(0)}(\omega) + \sum_{l=0}^{\ell-1} X^{(l+1)}(\omega) - X^{(l)}(\omega)$ converges uniformly on $[t_0-\tau, T]$ for all $\omega \in \Omega_1$. Hence, there exists an $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$-adapted process X with P-almost surely continuous realizations on $[t_0, T]$ such that

$$X_t = \lim_{\ell\to\infty} X_t^{(\ell)}$$

uniformly for all $t \in [t_0-\tau, T]$ holds P-almost surely. Moreover, the process X can be chosen to be $(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$-progressively measurable, see Section II.1. Applying Fatou's Lemma to inequality (II.27), it holds that $X \in S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$, and inequality (II.12) follows.

Now, it is left to show that X is indeed a solution of SDDE (II.1). By inequality (II.28), we have

$$\|X^{(\ell+1)} - X^{(\ell)}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq \|X^{(1)} - X^{(0)}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \frac{\big(\hat{C}\sqrt{T-t_0}\big)^\ell}{\sqrt{\ell!}}$$

for all $\ell \in \mathbb{N}_0$ where

$$\hat{C} := L_a\sqrt{T-t_0} + \frac{pL_b\sqrt{m}}{\sqrt{p-1}}.$$

Due to this, the series $\sum_{l\in\mathbb{N}_0} X^{(l+1)} - X^{(l)}$ converges in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ because

$$\begin{aligned}
\Big\|\sum_{l\in\mathbb{N}_0} X^{(l+1)} - X^{(l)}\Big\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} &\leq \sum_{l\in\mathbb{N}_0} \|X^{(l+1)} - X^{(l)}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&= \|X^{(1)} - X^{(0)}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \sum_{l\in\mathbb{N}_0} \frac{\big(\hat{C}\sqrt{T-t_0}\big)^l}{\sqrt{l!}} \\
&< \infty
\end{aligned}$$

by the root test and e. g. Stirling's approximation. Hence, it holds

$$\lim_{\ell\to\infty} \|X^{(\ell)} - X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = 0. \tag{II.29}$$

Then, similarly to the inequalities (II.22) and (II.23), we obtain by triangle inequality, Theorem II.6 and Lipschitz conditions (II.8) and (II.9) that

$$\begin{aligned}
&\lim_{\ell\to\infty} \left\| \int_{t_0}^{\cdot} a(\mathcal{T}(s, X_s^{(\ell)}))\,\mathrm{d}s - \int_{t_0}^{\cdot} a(\mathcal{T}(s, X_s))\,\mathrm{d}s \right\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le L_a(T - t_0) \lim_{\ell\to\infty} \|X^{(\ell)} - X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad = 0
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{\ell\to\infty} \left\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(s, X_s^{(\ell)}))\,\mathrm{d}W_s^j - \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(s, X_s))\,\mathrm{d}W_s^j \right\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le \frac{pL_b\sqrt{m}\sqrt{T-t_0}}{\sqrt{p-1}} \lim_{\ell\to\infty} \|X^{(\ell)} - X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad = 0.
\end{aligned}$$

Due to this and since $X_t = \xi_t$ for all $t \in [t_0 - \tau, t_0]$ P-almost surely, the stochastic process X must thus be the unique strong solution of SDDE (II.1). □

Proof of Lemma II.9

Proof of Lemma II.9. Since X is the strong solution of SDDE (II.1), and the SDDE's coefficients satisfy the linear growth conditions (II.10) and (II.11), it holds $X \in S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ by the proof of Theorem II.8.

Similarly to the inequalities (II.25) and (II.26), the triangle inequality, Zakai's inequality from Theorem II.6, and linear growth conditions (II.10) and (II.11) then imply for $s, t \in [t_0, T]$ with $s < t$ that

$$\begin{aligned}
&\|X_t - X_s\|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad \le \left\| \int_s^t a(\mathcal{T}(s, X_s))\,\mathrm{d}s \right\|_{L^p(\Omega;\mathbb{R}^d)} + \left\| \sum_{j=1}^{m} \int_s^t b^j(\mathcal{T}(s, X_s))\,\mathrm{d}W_s^j \right\|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad \le \int_s^t \|a(\mathcal{T}(s, X_s))\|_{L^p(\Omega;\mathbb{R}^d)}\,\mathrm{d}s + \sqrt{p-1}\left(\int_s^t \left\| \sum_{j=1}^{m} \|b^j(\mathcal{T}(s, X_s))\|^2 \right\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \mathrm{d}s \right)^{\frac{1}{2}} \\
&\quad \le \left(K_a\sqrt{T-t_0} + \sqrt{p-1}K_b\sqrt{m}\right)\left(1 + \|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\right)^{\frac{1}{2}}\sqrt{t-s}.
\end{aligned}$$

The case $t < s$ is completely analogous to the one above, and the case $s = t$ is trivial. □

Proof of Lemma II.10

Proof of Lemma II.10. At first, inequality (II.6) implies

$$\|X^\xi - X^\zeta\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq 2\|X^\xi - X^\zeta\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} + 2\|X^\xi - X^\zeta\|^2_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},$$

where

$$\|X^\xi - X^\zeta\|_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} = \|\xi - \zeta\|_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}.$$

Consider the term $\|X^\xi - X^\zeta\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}$ on the right-hand side of the inequality above. Similarly to inequalities (II.22) and (II.23), we obtain by triangle inequality, Zakai's inequality from Theorem II.6, Lipschitz conditions (II.8) and (II.9) as well as the Cauchy-Schwarz inequality that

$$\begin{aligned}
&\|X^\xi - X^\zeta\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq \left\| \int_{t_0}^{\cdot} a(\mathcal{T}(s, X^\xi_u)) - a(\mathcal{T}(s, X^\zeta_u))\,\mathrm{d}s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad + \left\| \sum_{j=1}^m \int_{t_0}^{\cdot} b^j(\mathcal{T}(s, X^\xi_s)) - b^j(\mathcal{T}(s, X^\zeta_s))\,\mathrm{d}W^j_s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq \left(L_a\sqrt{T-t_0} + \frac{pL_b\sqrt{m}}{\sqrt{p-1}} \right) \left(\int_{t_0}^T \|X^\xi - X^\zeta\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s \right)^{\frac{1}{2}}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\|X^\xi - X^\zeta\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq 2\|\xi - \zeta\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} \\
&\qquad + 2\left(L_a\sqrt{T-t_0} + \frac{pL_b\sqrt{m}}{\sqrt{p-1}} \right)^2 \int_{t_0}^T \|X^\xi - X^\zeta\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s,
\end{aligned}$$

and Gronwall's Lemma II.7 implies

$$\|X^\xi - X^\zeta\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq 2\|\xi - \zeta\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} \mathrm{e}^{2\left(L_a\sqrt{T-t_0} + \frac{pL_b\sqrt{m}}{\sqrt{p-1}}\right)^2 (T-t_0)}$$

Then, the assertion follows by taking the square root. □

III
Some Results on the Malliavin Calculus

Our numerical analysis of the Milstein scheme in Chapter IV involves techniques from the Malliavin calculus although all occurring stochastic integrals are well-defined in the sense of Itô. We need, among others, a chain rule for the Malliavin derivative, the Malliavin derivative of the solution of SDDE (II.1), and the Skorohod integral in order to prove that the error of the Milstein scheme in $S^p([t_0 - \tau, T] \times \Omega; \mathbb{R}^d)$ is of order $\mathcal{O}(h)$ as $h \to 0$, where h is the maximum step size of the scheme. For more details on the estimates involving the Malliavin calculus, we refer to the analysis of term $\mathcal{R}_5$ in the proof of Theorem IV.9. See inequality (IV.146) for the final estimate.

The Malliavin derivative as well as its adjoint operator, namely the Skorohod integral, are introduced in Section III.1, and some important properties of them are stated. Here, we also develop a chain rule for the Malliavin derivative that applies to functions whose derivatives are not imposed to be bounded. In Section III.2, we focus on the Malliavin derivative of solutions of SDDE (II.1) and give details on the upper bound in the $S^p([t_0 - \tau, T] \times \Omega; \mathbb{R}^d)$-norm.

We refer to [97, 113, 114] for monographs on the Malliavin calculus. The contents of this chapter are mainly based on Nualart's books [112, 113].

III.1. Malliavin Derivative and Skorohod Integral

First, we introduce the Malliavin derivative for $\mathbb{R}$-valued random variables and state some important properties of it. Especially, a more general chain rule is presented. Thereafter, the Skorohod integral is defined as the adjoint operator of the Malliavin derivative. We then extend the definition of the Malliavin derivative to Hilbert space valued random variables in order to study some valuable properties of the Skorohod integral.

This section follows sections 1.2 and 1.3 in Nualart's books [112, 113] to a large extent. However, we choose a slightly different representation that is more suitable for our considerations on the Wiener process W in case of SDDE (II.1).

To begin with, we introduce some notations. Let $C_p^\infty(\mathbb{R}^K; \mathbb{R})$, $K \in \mathbb{N}$, be the space of continuous functions $f\colon \mathbb{R}^K \to \mathbb{R}$ that are infinitely often continuously differentiable, and such that f and all of its partial derivatives have polynomial growth. Here, a function $g\colon \mathbb{R}^K \to \mathbb{R}$ is said to be of polynomial growth if there exist a constant $C \in \mathbb{R}$ with $C > 0$ and an exponent $q \in [0, \infty[$ so that $|g(x)| \leq C\big(1 + \|x\|^2\big)^{\frac{q}{2}}$ holds for all $x \in \mathbb{R}^K$, where q is the order of the growth.

Consider two real separable Hilbert spaces E_1 and E_2 with inner products $\langle \cdot, \cdot \rangle_{E_1}$ and $\langle \cdot, \cdot \rangle_{E_2}$. The norms $\|\cdot\|_{E_1}$ and $\|\cdot\|_{E_2}$ are assumed to be induced by the inner products $\langle \cdot, \cdot \rangle_{E_1}$ and $\langle \cdot, \cdot \rangle_{E_2}$. Let $L_{HS}(E_1; E_2)$ denote the space of Hilbert-Schmidt operators from E_1 to E_2 with inner product

$$\langle \cdot, \cdot \rangle_{L_{HS}(E_1;E_2)} \coloneqq \sum_{k \in \mathbb{N}} \langle \cdot\, e_k^1, \cdot\, e_k^1 \rangle_{E_2}$$

and norm

$$\|\cdot\|_{L_{HS}(E_1;E_2)} \coloneqq \left(\sum_{k \in \mathbb{N}} \|\cdot\, e_k^1\|_{E_2}^2 \right)^{\frac{1}{2}},$$

where the definitions are independent of the particular orthonormal basis $(e_k^1)_{k \in \mathbb{N}}$ in E_1, see e. g. [10, Lemma 3.4.2]. For $x, y \in E_1$ and $z \in E_2$, define the linear operator $x \otimes z \colon E_1 \to E_2$ by $(x \otimes z)\, y \coloneqq \langle x, y \rangle_{E_1} z$, cf. [55, p. 44]. It holds $x \otimes z \in L_{HS}(E_1; E_2)$ and

$$\begin{aligned} \|x \otimes z\|_{L_{HS}(E_1;E_2)}^2 &= \sum_{k \in \mathbb{N}} \|(x \otimes z)\, e_k^1\|_{E_2}^2 = \sum_{k \in \mathbb{N}} \|\langle x, e_k^1 \rangle_{E_1} z\|_{E_2}^2 = \sum_{k \in \mathbb{N}} |\langle x, e_k^1 \rangle_{E_1}|^2 \|z\|_{E_2}^2 \\ &= \|x\|_{E_1}^2 \|z\|_{E_2}^2 \end{aligned} \tag{III.1}$$

by Parseval's identity [138, Theorem III.4.2]. Moreover, the space $L_{HS}(E_1; E_2)$ is again a real separable Hilbert space with the orthonormal basis $(e_k^1 \otimes e_l^2)_{k,l \in \mathbb{N}}$, where $(e_l^2)_{l \in \mathbb{N}}$ is an orthonormal basis of E_2, see e. g. [116, Proposition B.0.7].

We define the real separable Hilbert space

$$\begin{aligned} \mathrm{H}_E &\coloneqq L^2([t_0, T]; L_{HS}(\mathbb{R}^m; E)) \\ &\coloneqq L^2(([t_0, T], \mathscr{B}([t_0, T]), \lambda|_{[t_0,T]}); (L_{HS}(\mathbb{R}^m; E), \mathscr{B}(L_{HS}(\mathbb{R}^m; E)))), \end{aligned} \tag{III.2}$$

where E is a real separable Hilbert space. As in Chapter II, an element $Z \in \mathrm{H}_E$ is referred to as a $\mathscr{B}([t_0, T])/\mathscr{B}(L_{HS}(\mathbb{R}^m; E))$-measurable function instead of an equivalence class throughout this thesis. In case of $E = \mathbb{R}$, we write $\mathrm{H} \coloneqq \mathrm{H}_{\mathbb{R}}$ for sake of simplicity, which is used frequently especially at the beginning of this section.

In order to see the advantage of this notation, we need the following considerations. In view of equation (III.1), it holds

$$\begin{aligned} \|x\|_{\mathbb{R}^m} &= \left(\sum_{j=1}^m |\langle x, e_j \rangle_{\mathbb{R}^m}|^2 \right)^{\frac{1}{2}} = \left(\sum_{j=1}^m |\langle x, e_j \rangle_{\mathbb{R}^m}\, 1|^2 \right)^{\frac{1}{2}} = \left(\sum_{j=1}^m |(x \otimes 1)\, e_j|^2 \right)^{\frac{1}{2}} \\ &= \|x \otimes 1\|_{L_{HS}(\mathbb{R}^m;\mathbb{R})} \end{aligned}$$

for all $x \in \mathbb{R}^m$, where $(e_j)_{j \in \{1,\ldots,m\}}$ is the canonical orthonormal basis of $\mathbb{R}^m$, that is, e_j denotes the jth unit vector in $\mathbb{R}^m$. Thus, the map $\iota \colon x \mapsto x \otimes 1$ is an isometric isomorphism between the spaces $\mathbb{R}^m$ and $L_{HS}(\mathbb{R}^m; \mathbb{R})$. We define

$$x^j \coloneqq (x \otimes 1)\, e_j = \langle x, e_j \rangle_{\mathbb{R}^m} \tag{III.3}$$

for all $x \in \mathbb{R}^m$ and $j \in \{1, \dots, m\}$. Then, the inner product of $g, h \in \mathrm{H}$ can be represented as

$$\langle g, h \rangle_{\mathrm{H}} \coloneqq \int_{t_0}^{T} \langle g(t), h(t) \rangle_{L_{HS}(\mathbb{R}^m;\mathbb{R})} \,\mathrm{d}t = \sum_{j=1}^{m} \int_{t_0}^{T} g^j(t) h^j(t) \,\mathrm{d}t. \tag{III.4}$$

Using these definitions, we can consistently write

$$\int_{t_0}^{T} h(s) \,\mathrm{d}W_s \coloneqq \sum_{j=1}^{m} \int_{t_0}^{T} h^j(s) \,\mathrm{d}W_s^j \tag{III.5}$$

for all $h \in \mathrm{H}$, where $W_s = (W_s^1, \dots, W_s^m)^{\mathrm{T}}$. This notation is also used in a more general context of Hilbert space valued Wiener processes, see e. g. [28, Section I.4]. Thus, space H introduced above is more usual in the context of SDEs than the space in [112, 113], see especially [113, Example 1.1.2]. The Itô isometry for a deterministic function $h \in \mathrm{H}$ reads as

$$\Big\| \int_{t_0}^{T} h(s) \,\mathrm{d}W_s \Big\|_{L^2(\Omega;\mathbb{R})}^2 = \int_{t_0}^{T} \sum_{j=1}^{m} (h^j(s))^2 \,\mathrm{d}s = \int_{t_0}^{T} \|h(s)\|_{L_{HS}(\mathbb{R}^m;\mathbb{R})}^2 \,\mathrm{d}s = \|h\|_{\mathrm{H}}^2$$

for example, cf. [28, Equation (I.4.30)]. In particular, the Wiener integral in equation (III.5) is $\mathrm{N}(0, \|h\|_{\mathrm{H}}^2)$-distributed.

We continue with the following definition.

Definition III.1 ([113, p. 25])
The set of $\mathbb{R}$*-valued* smooth random variables *is denoted by*

$$\mathscr{S}(\Omega;\mathbb{R}) \coloneqq \Bigg\{ F\colon \Omega \to \mathbb{R} : F = f\bigg(\int_{t_0}^{T} h_1(s) \,\mathrm{d}W_s, \dots, \int_{t_0}^{T} h_K(s) \,\mathrm{d}W_s \bigg) \\ \text{where } f \in C_p^{\infty}(\mathbb{R}^K;\mathbb{R}), h_k \in \mathrm{H} \text{ for } k \in \{1, \dots, K\}, \text{ and } K \in \mathbb{N} \Bigg\}.$$

Because $f \in C_p^{\infty}(\mathbb{R}^K;\mathbb{R})$ is of polynomial growth and because the Wiener integral in equation (III.5) is $\mathrm{N}(0, \|h\|_{\mathrm{H}}^2)$-distributed, it holds $\mathscr{S}(\Omega;\mathbb{R}) \subset L^p(\Omega;\mathbb{R})$ for all $p \in [1, \infty[$. However, not every random variable in $L^p(\Omega;\mathbb{R})$ can be approximated by a sequence $(F_n)_{n\in\mathbb{N}}$ of smooth random variables $F_n \in \mathscr{S}(\Omega;\mathbb{R})$, $n \in \mathbb{N}$.

Indeed, recall the P-completed σ-algebra $\mathscr{G} \coloneqq \mathscr{G}_T$ from Chapter II, cf. formula (II.2). The smooth random variables $F_n \in \mathscr{S}(\Omega;\mathbb{R})$ are $\mathscr{G}/\mathscr{B}(\mathbb{R})$-measurable, and hence, the limit $\lim_{n\to\infty} F_n$ in $L^p(\Omega;\mathbb{R})$, as it exists, is $\mathscr{G}/\mathscr{B}(\mathbb{R})$-measurable, too, see e. g. [24, Corollary 2.2.3]. But as $\mathscr{G} \subset \mathscr{F}$ in general, not every random variable in $L^p(\Omega;\mathbb{R}) = L^p((\Omega, \mathscr{F}, \mathrm{P}); (\mathbb{R}, \mathscr{B}(\mathbb{R})))$ is $\mathscr{G}/\mathscr{B}(\mathbb{R})$-measurable.

Due to this, we introduce the Banach space

$$L_{\mathscr{G}}^p(\Omega; E) \coloneqq L^p((\Omega, \mathscr{G}, \mathrm{P}|_{\mathscr{G}}); (E, \mathscr{B}(E))) = \{\mathrm{E}[Z|\mathscr{G}] : Z \in L^p(\Omega; E)\}$$

with norm $\|\cdot\|_{L_{\mathscr{G}}^p(\Omega;E)} \coloneqq \|\cdot\|_{L^p(\Omega;E)}$, $\mathrm{P}|_{\mathscr{G}}$-almost surely equal random variables are identified, for $p \in [1, \infty[$, where E is a real separable Hilbert space, cf. [121, Chapter II].

Using this notation, it can more precisely be said that

$$\mathscr{S}(\Omega;\mathbb{R}) = \mathscr{S}((\Omega,\mathscr{G},\mathrm{P}|_{\mathscr{G}});(\mathbb{R},\mathscr{B}(\mathbb{R}))) \subset L^p_{\mathscr{G}}(\Omega;\mathbb{R})$$

for every $p \in [1,\infty[$. Even more, the following lemma holds.

Lemma III.2 ([111, Lemma 2.3.1])
The space $\mathscr{S}(\Omega;\mathbb{R})$ is dense in $L^p_{\mathscr{G}}(\Omega;\mathbb{R})$ for every $p \in [1,\infty[$.

We now define the Malliavin derivative for an arbitrary random variable $F \in \mathscr{S}(\Omega;\mathbb{R})$.

Definition III.3 ([113, Definition 1.2.1])
Let $F \in \mathscr{S}(\Omega;\mathbb{R})$ with the representation

$$F = f\biggl(\int_{t_0}^T h_1(s)\,\mathrm{d}W_s,\dots,\int_{t_0}^T h_K(s)\,\mathrm{d}W_s\biggr).$$

The Malliavin derivative $\mathrm{D}\colon \mathscr{S}(\Omega;\mathbb{R}) \to L^p_{\mathscr{G}}(\Omega;\mathrm{H})$, *$p \in [1,\infty[$, of F is defined by*

$$\mathrm{D}F := \sum_{k=1}^K \partial_{x_k} f\biggl(\int_{t_0}^T h_1(s)\,\mathrm{d}W_s,\dots,\int_{t_0}^T h_K(s)\,\mathrm{d}W_s\biggr)h_k.$$

In particular, it thus holds

$$\mathrm{D}\int_{t_0}^T h(s)\,\mathrm{d}W_s = h \tag{III.6}$$

for all $h \in \mathrm{H}$. Considering $F, G \in \mathscr{S}(\Omega;\mathbb{R})$, we clearly have $FG \in \mathscr{S}(\Omega;\mathbb{R})$, and the *product rule*

$$\mathrm{D}(FG) = (\mathrm{D}F)G + F(\mathrm{D}G). \tag{III.7}$$

directly follows from the definition of the Malliavin derivative. The Malliavin derivative further satisfies the following lemma.

Lemma III.4 ([113, Lemma 1.2.1])
Let $F \in \mathscr{S}(\Omega;\mathbb{R})$ and $h \in H$. Then, it holds

$$\mathrm{E}[\langle \mathrm{D}F, h\rangle_{\mathrm{H}}] = \mathrm{E}\biggl[F\int_{t_0}^T h(t)\,\mathrm{d}W_t\biggr].$$

The equation in the lemma above can be seen as an integration by parts formula and is useful to prove that the Malliavin derivative D is a closable operator, cf. [111, Lemma 1.1.1 and Proposition 2.3.4]. The Malliavin derivative D is closable if and only if for all sequences $(F_n)_{n\in\mathbb{N}} \subset \mathscr{S}(\Omega;\mathbb{R})$ with $\lim_{n\to\infty} F_n = 0$ and $\lim_{n\to\infty} \mathrm{D}F_n = G \in L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ it follows $G = 0$ $\mathrm{P}|_{\mathscr{G}}$-almost surely, see e. g. [138, Definition II.6.2 and Proposition II.6.2] or [14, Subsection 12.2.2].

Proposition III.5 ([113, Proposition 1.2.1])
Let $p \in [1,\infty[$. The operator $\mathrm{D}\colon L^p_{\mathscr{G}}(\Omega;\mathbb{R}) \supset \mathscr{S}(\Omega;\mathbb{R}) \to L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ is closable.

Definition III.6 ([113, p. 27])
Let $p \in [1, \infty[$. The closure of the set $\mathscr{S}(\Omega;\mathbb{R})$ with respect to the graph norm

$$\|\cdot\|_{\mathscr{D}^p(\Omega;\mathbb{R})} \coloneqq \left(\|\cdot\|^p_{L^p_{\mathscr{G}}(\Omega;\mathbb{R})} + \|\mathrm{D}\cdot\|^p_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})}\right)^{\frac{1}{p}}$$

is denoted by $\mathscr{D}^p(\Omega;\mathbb{R}) \subset L^p_{\mathscr{G}}(\Omega;\mathbb{R})$.

According to Proposition III.5, the operator $\mathrm{D}\colon \mathscr{S}(\Omega;\mathbb{R}) \to L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ can be extended to $\mathscr{D}^p(\Omega;\mathbb{R})$. This closure $\overline{\mathrm{D}}$ of the operator D will, by a slight abuse of notation, again be denoted by $\mathrm{D}\colon \mathscr{D}^p(\Omega;\mathbb{R}) \to L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ in the following, cf. [111, 113].

Let $q, r \in]1, \infty[$, and consider $F \in \mathscr{D}^q(\Omega;\mathbb{R})$ and $G \in \mathscr{D}^r(\Omega;\mathbb{R})$. Using Hölder's inequality and the product rule (III.7), we obtain $FG \in \mathscr{D}^p(\Omega;\mathbb{R})$ where $\frac{1}{p} = \frac{1}{q} + \frac{1}{r}$ and

$$\mathrm{D}(FG) = (\mathrm{D}F)G + F(\mathrm{D}G) \tag{III.8}$$

cf. [68, Corollary 15.80]. Before we continue with further valuable properties of the Malliavin derivative, we make a remark on the measurability of the Malliavin derivative $\mathrm{D}F$ for $F \in \mathscr{D}^p(\Omega;\mathbb{R})$ and introduce some notations.

Remark III.7
Considering $F \in \mathscr{D}^p(\Omega;\mathbb{R})$, a representative $\mathrm{D}F \in L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ is a random variable with values in Hilbert space H, that is, $\mathrm{D}F(\omega)$ is actually an equivalence class. Of course, for all $\omega \in \Omega$ one can also pick a representative in this equivalence class. These representatives $t \mapsto DF(\omega, t)$, $t \in [t_0, T]$, are $\mathscr{B}([t_0,T])/\mathscr{B}(L_{HS}(\mathbb{R}^m;\mathbb{R}))$-measurable. But for fixed $t \in [t_0, T]$, the map $\omega \mapsto \mathrm{D}F(\omega, t)$ is not necessarily $\mathscr{G}/\mathscr{B}(L_{HS}(\mathbb{R}^m;\mathbb{R}))$-measurable, that is, $\mathrm{D}F(t)$ is not a random variable.

However, according to [37, Theorem III.11.17] and [61, Proposition 1.2.25], respectively, there exists a $\mathscr{B}([t_0,T]) \otimes \mathscr{G}/\mathscr{B}(L_{HS}(\mathbb{R}^m;\mathbb{R}))$-measurable function $Z\colon [t_0,T] \times \Omega \to L_{HS}(\mathbb{R}^m;\mathbb{R})$ such that $Z(\omega) = \mathrm{D}F(\omega) \in \mathrm{H}$ for $\mathrm{P}|_{\mathscr{G}}$-almost all $\omega \in \Omega$. Moreover, Z is uniquely determined except for a set $A \in \mathscr{B}([t_0,T]) \otimes \mathscr{G}$ with $(\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}})[A] = 0$. That is, Z is uniquely determined up to indistinguishability. In the following, a representative $\mathrm{D}F \in L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ is always assumed to be this $\mathscr{B}([t_0,T]) \otimes \mathscr{G}/\mathscr{B}(L_{HS}(\mathbb{R}^m;\mathbb{R}))$-measurable stochastic process $Z\colon [t_0,T] \times \Omega \to L_{HS}(\mathbb{R}^m;\mathbb{R})$ and

$$\mathrm{D}_t F(\omega) \coloneqq Z(t, \omega)$$

for all $(t, \omega) \in [t_0, T] \times \Omega$. Moreover, let $\mathrm{D}^j F\colon [t_0,T] \times \Omega \to \mathbb{R}$ denote the $\mathbb{R}$-valued measurable process defined by

$$\mathrm{D}^j_t F(\omega) \coloneqq Z^j(t, \omega) \coloneqq Z(t, \omega) e_j \tag{III.9}$$

for $j \in \{1, \ldots, m\}$, cf. formula (III.3)*. In particular, it holds for $F \in \mathscr{S}(\Omega;\mathbb{R})$ that*

$$\mathrm{D}^j_t F(\omega) = \sum_{k=1}^{K} \partial_{x_k} f\left(\int_{t_0}^{T} h_1(s)\,\mathrm{d}W_s, \ldots, \int_{t_0}^{T} h_K(s)\,\mathrm{d}W_s\right)(\omega) h^j_k(t)$$

for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(t,\omega) \in [t_0,T] \times \Omega$ and all $j \in \{1, \ldots, m\}$. Note that $\mathrm{D}F$ for $F \in \mathscr{S}(\Omega;\mathbb{R})$ is a priori $\mathscr{B}([t_0,T]) \otimes \mathscr{G}/\mathscr{B}(L_{HS}(\mathbb{R}^m;\mathbb{R}))$-measurable. The measurability of representatives $\mathrm{D}F \in L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ with $F \in \mathscr{D}^p(\Omega;\mathbb{R})$ gets lost by passing to the equivalence classes.

Having adapted stochastic processes in mind, the following corollary will be important later on. Recall filtration $(\mathscr{G}_t)_{t\in[t_0,T]}$, see equation (II.2).

Corollary III.8 ([113, Corollary 1.2.1])
Let $p \in [1,\infty[$ and $F \in \mathscr{D}^p(\Omega;\mathbb{R})$ be $\mathscr{G}_t/\mathscr{B}(\mathbb{R})$-measurable for some $t \in [t_0,T]$. Then, it holds $\mathrm{D}_s^j F(\omega) = 0$ for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(s,\omega) \in]t,T] \times \Omega$ and all $j \in \{1,\dots,m\}$.

The next proposition states the chain rule for the Malliavin derivative, cf. [68, Theorem 15.78], [97, p. 36], and [113, Proposition 1.2.3 and Proposition 1.5.1]. The extended version in the theorem below is very useful in order to calculate the Malliavin derivative explicitly.

Theorem III.9
Let $F = (F_1,\dots,F_L)$ with $F_l \in \mathscr{D}^p(\Omega;\mathbb{R})$ for $l \in \{1,\dots,L\}$ and some $p \in [1,\infty[$. Further, let $\varphi \in C^1(\mathbb{R}^L;\mathbb{R})$ with $|\partial_{x_l}\varphi(x)| \leq C\big(1+\|x\|^2\big)^{\frac{\chi}{2}}$ for all $x \in \mathbb{R}^L$ and some $\chi \in [0,p-1]$, where $C>0$ is a constant. Then $\varphi(F) \in \mathscr{D}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})$, and it $\mathrm{P}|_{\mathscr{G}}$-almost surely holds

$$\mathrm{D}\varphi(F) = \sum_{l=1}^{L} \partial_{x_l}\varphi(F)\,\mathrm{D}F_l.$$

Proof. The proof is stated in Section III.3, see p. 41. □

In the case of $\varphi \in C(\mathbb{R}^L;\mathbb{R})$ having bounded partial derivatives $\partial_{x_l}\varphi$, $l \in \{1,\dots,L\}$, the proposition holds with $\chi = 0$, and thus, the statement is more general than the result in [113, Proposition 1.2.3].

In the following, the adjoint operator of the Malliavin derivative $\mathrm{D}\colon L^p_{\mathscr{G}}(\Omega;\mathbb{R}) \supset \mathscr{D}^p(\Omega;\mathbb{R}) \to L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ is considered for $p \in]1,\infty[$, cf. [97, Definition II.6.1] or [68, Definition 15.130 and Theorem 15.132], and in the case of $p=2$, cf. [113, Definition 1.3.1]. We refer to e. g. [62, pp. 521–522] for the definition of the adjoint operator.

Definition III.10
Let $p \in]1,\infty[$. The subspace $\mathrm{dom}\,\delta \subset L^p_{\mathscr{G}}(\Omega;\mathrm{H})$, $p \in]1,\infty[$, denotes the set of random variables $G \in L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ such that $F \mapsto \mathrm{E}[\langle \mathrm{D}F, G\rangle_{\mathrm{H}}]$ is continuous for all $F \in \mathscr{D}^q(\Omega;\mathbb{R})$ with $\frac{1}{p}+\frac{1}{q}=1$. That is $G \in \mathrm{dom}\,\delta$ if and only if there exists a constant $C_G > 0$ such that

$$|\mathrm{E}[\langle \mathrm{D}F, G\rangle_{\mathrm{H}}]| \leq C_G \|F\|_{L^q_{\mathscr{G}}(\Omega;\mathbb{R})}$$

for all $F \in \mathscr{D}^q(\Omega;\mathbb{R})$.

If $G \in \mathrm{dom}\,\delta$, then $\delta(G) \in L^p_{\mathscr{G}}(\Omega;\mathbb{R})$ is characterized by the duality formula

$$\mathrm{E}[F\delta(G)] = \mathrm{E}[\langle \mathrm{D}F, G\rangle_{\mathrm{H}}] \tag{III.10}$$

for any $F \in \mathscr{D}^q(\Omega;\mathbb{R})$ with $\frac{1}{p}+\frac{1}{q}=1$. The operator $\delta\colon L^p_{\mathscr{G}}(\Omega;\mathrm{H}) \supset \mathrm{dom}\,\delta \to L^p_{\mathscr{G}}(\Omega;\mathbb{R})$ is called divergence operator.

The random variable $\delta(G) \in L^p_{\mathscr{G}}(\Omega;\mathbb{R})$ is *unique* in the sense that if $G, \hat{G} \in \operatorname{dom}\delta \subset L^p_{\mathscr{G}}(\Omega;\mathrm{H})$ with $G = \hat{G}\ \lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost surely, then it $\mathrm{P}|_{\mathscr{G}}$-almost surely holds $\delta(G) = \delta(\hat{G})$.

Since $\mathscr{S}(\Omega;\mathbb{R})$ is a dense subset of $L^p_{\mathscr{G}}(\Omega;\mathbb{R})$ for every $p \in]1,\infty[$, see Lemma III.2, it is enough to verify the continuity of $F \mapsto \mathrm{E}[\langle \mathrm{D}F, G\rangle_{\mathrm{H}}]$ and the duality formula only for all $F \in \mathscr{S}(\Omega;\mathbb{R})$ in the previous definition, cf. [62, p. 522]. The divergence operator $\delta\colon \operatorname{dom}\delta \to L^p_{\mathscr{G}}(\Omega;\mathbb{R})$ is closed as it is the adjoint of the densely defined operator $\mathrm{D}\colon \mathscr{D}^p(\Omega;\mathbb{R}) \to L^p_{\mathscr{G}}(\Omega;\mathrm{H})$, cf. [113, p. 37] and [14, Theorem 12.3.1]. Equation (III.10) is also called *integration by parts formula*, cf. Lemma III.4. It follows immediately from equation (III.10) that the divergence operator δ is linear. Choosing further $F = c \in \mathbb{R}$ in equation (III.10), it holds $\mathrm{E}[\delta(G)] = 0$ if $G \in \operatorname{dom}\delta$. According to Lemma III.4 and equation (III.6), we have $\delta(h) = \int_{t_0}^T h(s)\,\mathrm{d}W_s$ and $\mathrm{D}\delta(h) = h$ for $h \in \mathrm{H}$. Similar relations of the divergence operator are derived in a more general setting in the following. In order to do this, the definition of the Malliavin derivative needs to be extended to stochastic processes.

In Nualart's book [113], the Malliavin derivative is extended to random variables that take values in an arbitrary separable Hilbert space. It should be mentioned that this can also be done for separable Banach spaces or even Banach spaces satisfying only a UMD property. The theory about stochastic integration and the Malliavin calculus in UMD-Banach spaces is developed by, among others, Maas, Pronk, van Nerven, Veraar and Weis, see e. g. [61, 93, 94, 117, 132, 133].

Using Definition III.1, we first define smooth random variables that take values in a real separable Hilbert space. Considering $F \in L^p(\Omega;\mathbb{R})$ and $x \in E$, let $F\cdot x \in L^p(\Omega;E)$ be defined by $(F\cdot x)(\omega) := F(\omega)x$ for all $\omega \in \Omega$.

Definition III.11 ([113, p. 31])
Let E be real separable Hilbert space. The set of E-valued smooth random variables *is denoted by*

$$\mathscr{S}(\Omega;E) := \Big\{ F\colon \Omega \to E : F = \sum_{k=1}^{n} F_k \cdot x_k \\ \text{where } F_k \in \mathscr{S}(\Omega;\mathbb{R}) \text{ and } x_k \in E \text{ for } k \in \{1,\dots,n\} \text{ and } n \in \mathbb{N} \Big\}.$$

Similar to Lemma III.2, the space $\mathscr{S}(\Omega;E)$ is dense in $L^p_{\mathscr{G}}(\Omega;E)$, cf. [94, Lemma 3.1].

Lemma III.12
Let E be real separable Hilbert space. The space of smooth E-valued random variables $\mathscr{S}(\Omega;E)$ is dense in $L^p_{\mathscr{G}}(\Omega;E)$ for every $p \in [1,\infty[$.

Proof. Since $\mathscr{S}(\Omega;\mathbb{R})$ is dense in $L^p_{\mathscr{G}}(\Omega;\mathbb{R})$ for every $p \in [1,\infty[$, see Lemma III.2, it follows that $\mathscr{S}(\Omega;E)$ is dense in $L^p_{\mathscr{G}}(\Omega;E)$ for every $p \in [1,\infty[$, see e. g. [31, Subsection I.7.2 on p. 78]. □

In order to define the Malliavin derivative for E-valued smooth random variables, cf. [94, p. 154] and [113, p. 31], recall the space $\mathrm{H}_E = L^2([t_0,T];L_{HS}(\mathbb{R}^m;E))$, where $\mathrm{H} = \mathrm{H}_{\mathbb{R}}$, see formula (III.2).

Definition III.13
Let $G \in \mathscr{S}(\Omega;\mathbb{R})$ and $x \in E$, and let $p \in [1,\infty[$. The Malliavin derivative $\mathrm{D}\colon \mathscr{S}(\Omega;E) \to L^p_{\mathscr{G}}(\Omega;\mathrm{H}_E)$ *of $G\cdot x \in \mathscr{S}(\Omega;E)$ is defined by*

$$\mathrm{D}^j_t(G\cdot x)(\omega) \coloneqq \mathrm{D}^j_t G(\omega)\, x,$$

for all $(t,\omega) \in [t_0,T] \times \Omega$ and all $j \in \{1,\dots,m\}$. This definition extends to general $F = \sum_{k=1}^n F_k\cdot x_k \in \mathscr{S}(\Omega;E)$ with $F_k \in \mathscr{S}(\Omega;\mathbb{R})$, $x_k \in E$, $k \in \{1,\dots,n\}$, and $n \in \mathbb{N}$ by linearity, that is

$$\Big(\mathrm{D}^j_t \sum_{k=1}^n F_k\cdot x_k\Big)(\omega) \coloneqq \sum_{k=1}^n \mathrm{D}^j_t(F_k\cdot x_k)(\omega) = \sum_{k=1}^n \mathrm{D}^j_t F_k(\omega)\, x_k$$

for all $(t,\omega) \in [t_0,T] \times \Omega$ and all $j \in \{1,\dots,m\}$.

Similar to Proposition III.5, the Malliavin derivative operator for E-valued smooth random variables is closable, too.

Proposition III.14 ([94, p. 155])
Let E be a real separable Hilbert space and $p \in [1,\infty[$. The operator $\mathrm{D}\colon L^p_{\mathscr{G}}(\Omega;E) \supset \mathscr{S}(\Omega;E) \to L^p_{\mathscr{G}}(\Omega;\mathrm{H}_E)$ *is closable.*

The space $\mathscr{D}^p(\Omega;\mathbb{R})$ from Definition III.6 extends to E-valued random variables as follows.

Definition III.15 ([113, p.31])
Let E be a separable Hilbert space and $p \in [1,\infty[$. The closure of $\mathscr{S}(\Omega;E)$ with respect to the graph norm

$$\|\cdot\|_{\mathscr{D}^p(\Omega;E)} \coloneqq \Big(\|\cdot\|^p_{L^p_{\mathscr{G}}(\Omega;E)} + \|\mathrm{D}\cdot\|^p_{L^p_{\mathscr{G}}(\Omega;\mathrm{H}_E)}\Big)^{\frac{1}{p}} \tag{III.11}$$

is denoted by $\mathscr{D}^p(\Omega;E) \subset L^p_{\mathscr{G}}(\Omega;E)$.

The closed extension $\overline{\mathrm{D}}$ of operator $\mathrm{D}\colon L^p_{\mathscr{G}}(\Omega;E) \supset \mathscr{S}(\Omega;E) \to L^p_{\mathscr{G}}(\Omega;\mathrm{H}_E)$ to the set $\mathscr{D}^p(\Omega;E)$ will again be denoted by D. Considering $p,q \in [1,\infty[$ with $p \le q$, Hölder's inequality and inequality (II.6) imply

$$\begin{aligned}\Big(\|\cdot\|^p_{L^p_{\mathscr{G}}(\Omega;E)} + \|\mathrm{D}\cdot\|^p_{L^p_{\mathscr{G}}(\Omega;\mathrm{H}_E)}\Big)^{\frac{1}{p}} &\le \Big(\|\cdot\|^p_{L^q_{\mathscr{G}}(\Omega;E)} + \|\mathrm{D}\cdot\|^p_{L^q_{\mathscr{G}}(\Omega;\mathrm{H}_E)}\Big)^{\frac{q}{p}\frac{1}{q}} \\ &\le 2^{\frac{q-p}{p}}\Big(\|\cdot\|^q_{L^q_{\mathscr{G}}(\Omega;E)} + \|\mathrm{D}\cdot\|^q_{L^q_{\mathscr{G}}(\Omega;\mathrm{H}_E)}\Big)^{\frac{1}{q}},\end{aligned}$$

and thus $\|\cdot\|_{\mathscr{D}^p(\Omega;E)} \le 2^{\frac{q-p}{p}}\|\cdot\|_{\mathscr{D}^q(\Omega;E)}$. This yields the inclusion $\mathscr{D}^q(\Omega;E) \subset \mathscr{D}^p(\Omega;E)$, cf. [113, p. 27].

Similarly to Remark III.7 in case of $E = \mathbb{R}$, we make some remarks on the measurability of the Malliavin derivative, cf. [113, p. 42].

Remark III.16
Let $p \in [1, \infty[$ and $F \in \mathscr{D}^p(\Omega; E)$. A representative $\mathrm{D}F \in L^p_{\mathscr{G}}(\Omega; \mathrm{H}_E)$ is always assumed to be a $\mathscr{B}([t_0, T]) \otimes \mathscr{G}/\mathscr{B}(\mathrm{H}_E)$-measurable process, cf. Remark III.7.

In the following, consider the case of $E = L^2(A; E_2)$, where $A \subseteq [t_0 - \tau, T]$ is an interval, and E_2 is a real separable Hilbert space. Thus, we consider $F \in \mathscr{D}^p(\Omega; L^2(A; E_2))$, and hence $\mathrm{D}F \in L^p_{\mathscr{G}}(\Omega; \mathrm{H}_{L^2(A;E_2)})$, where $\mathrm{H}_{L^2(A;E_2)} = L^2([t_0, T]; L_{HS}(\mathbb{R}^m; L^2(A; E_2)))$.

Let $\iota\colon L_{HS}(\mathbb{R}^m; L^2(A; E_2)) \to L^2(A; L_{HS}(\mathbb{R}^m; E_2))$ be the isometric isomorphism defined by $(\iota(F)(t))(x) := (F\,x)(t)$, $t \in A$ and $x \in \mathbb{R}^m$, see [132, Proposition 2.6] or [131, Proposition 13.5 and Theorem 13.6]. According to isomorphism ι as well as [37, Theorem III.11.17] or [61, Proposition 1.2.25], there exists a $\mathscr{B}(A) \otimes \mathscr{B}([t_0, T]) \otimes \mathscr{G}/\mathscr{B}(L_{HS}(\mathbb{R}^m; E_2))$-measurable function $Z\colon A \times [t_0, T] \times \Omega \to L_{HS}(\mathbb{R}^m; E_2)$ such that $Z(\cdot, t, \omega) = \mathrm{D}_t F(\omega)$ for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(t, \omega) \in [t_0, T] \times \Omega$.

This two-parameter process Z is unique in the sense that if there exists another $\mathscr{B}(A) \otimes \mathscr{B}([t_0, T]) \otimes \mathscr{G}/\mathscr{B}(L_{HS}(\mathbb{R}^m; E_2))$-measurable function $\hat{Z}$ with $\hat{Z}(\cdot, t, \omega) = \mathrm{D}_t F(\omega)$ for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(t, \omega) \in [t_0, T] \times \Omega$, it holds $Z = \hat{Z}$ $\lambda|_A \otimes \lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost everywhere.

Therefore, a representative $\mathrm{D}F \in L^p_{\mathscr{G}}(\Omega; \mathrm{H}_{L^2(A;E_2)})$ is always assumed to be this $\mathscr{B}(A) \otimes \mathscr{B}([t_0, T]) \otimes \mathscr{G}/\mathscr{B}(L_{HS}(\mathbb{R}^m; E_2))$-measurable function Z and

$$\mathrm{D}_t F_s(\omega) := Z(s, t, \omega)$$

for all $(s, t, \omega) \in A \times [t_0, T] \times \Omega$. Moreover, let $\mathrm{D}^j F\colon A \times [t_0, T] \times \Omega \to E_2$ denote the E_2-valued measurable two-parameter process defined by

$$\mathrm{D}^j_t F_s(\omega) := Z^j(s, t, \omega) := Z(s, t, \omega) e_j$$

for all $(s, t, \omega) \in A \times [t_0, T] \times \Omega$ and all $j \in \{1, \ldots, m\}$, where $(e_j)_{j \in \{1,\ldots,m\}}$ is the canonical orthonormal basis of $\mathbb{R}^m$, cf. formula (III.9). If furthermore $E_2 = L_{HS}(\mathbb{R}^m; E_3)$, where E_3 is a real separable Hilbert space, define

$$\mathrm{D}^j_t F^l_s(\omega) := \mathrm{D}^j_t F_s(\omega) e_l$$

for all $(s, t, \omega) \in A \times [t_0, T] \times \Omega$ and all $j, l \in \{1, \ldots, m\}$.

The spaces considered in Remark III.16 occur for example when we apply the Malliavin derivative to integrands $F\colon [t_0, T] \times \Omega \to L_{HS}(\mathbb{R}^m; \mathbb{R})$ of Itô integrals with $F \in \mathscr{D}^p(\Omega; \mathrm{H})$.

Example III.17
Let $F \in \mathscr{D}^p(\Omega; \mathrm{H})$, and recall that

$$\mathrm{H}_\mathrm{H} = L^2([t_0, T]; L_{HS}(\mathbb{R}^m; L^2([t_0, T]; L_{HS}(\mathbb{R}^m; \mathbb{R})))),$$

see formula (III.2), where $\mathrm{H} = \mathrm{H}_\mathbb{R}$. Then, the representative $\mathrm{D}F \in L^p_{\mathscr{G}}(\Omega; \mathrm{H}_\mathrm{H})$ is a $\mathscr{B}([t_0, T]) \otimes \mathscr{B}([t_0, T]) \otimes \mathscr{G}/\mathscr{B}(L_{HS}(\mathbb{R}^m; L_{HS}(\mathbb{R}^m; \mathbb{R})))$-measurable two-parameter process. Further, for $\lambda|_{[t_0,T]}$-almost all $t \in [t_0, T]$ and all $j \in \{1, \ldots, m\}$ the representative $\mathrm{D}^j_t F \in L^p_{\mathscr{G}}(\Omega; \mathrm{H})$ is a $\mathscr{B}([t_0, T]) \otimes \mathscr{G}/\mathscr{B}(L_{HS}(\mathbb{R}^m; \mathbb{R}))$-measurable process, and $\mathrm{D}^j_t F^l_s$ is a $\mathscr{G}/\mathscr{B}(\mathbb{R})$-measurable random variable for $\lambda|_{[t_0,T]} \otimes \lambda|_{[t_0,T]}$-almost all $(t, s) \in [t_0, T] \times [t_0, T]$ and all $j, l \in \{1, \ldots, m\}$.

Due to the following remark, the random variable $\mathrm{D}_t^j F_s^l$ defined in Remark III.16 is well-defined since it does not depend on the order of taking the Malliavin derivative and evaluating the stochastic process in time.

Remark III.18
Let $F \in \mathscr{D}^p(\Omega; L^2([t_0,T]; E))$, where E is a real separable Hilbert space. It holds

$$\mathrm{D}_t^j F_s(\omega) = \mathrm{D}_t^j (F_s)(\omega)$$

for $\lambda|_{[t_0,T]} \otimes \lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(s,t,\omega) \in [t_0,T] \times [t_0,T] \times \Omega$. Hence, the order of the evaluation in time and the differentiation of F does not matter. Indeed, since F belongs to $\mathscr{D}^p(\Omega; L^2([t_0,T]; E))$, it can be approximated by a sequence $(F_n)_{n\in\mathbb{N}}$ where $F_n \in \mathscr{S}(\Omega; L^2([t_0,T]; E))$ with

$$F_n = \sum_{k=1}^{K_n} G_{n,k} \cdot h_{n,k},$$

$G_{n,k} \in \mathscr{S}(\Omega; \mathbb{R})$ and $h_{n,k} \in L^2([t_0,T]; E)$. By definition, we have

$$\mathrm{D}_t^j F_n(\omega) = \sum_{k=1}^{K_n} \mathrm{D}_t^j G_{n,k}(\omega)\, h_{n,k}$$

for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(t,\omega) \in [t_0,T] \times \Omega$ and all $j \in \{1,\dots,m\}$. Hence, evaluating $\mathrm{D}_t^j F_n(\omega) \in L^2([t_0,T]; E)$ yields

$$\mathrm{D}_t^j F_n(s)(\omega) = \sum_{k=1}^{K_n} \mathrm{D}_t^j G_{n,k}(\omega)\, h_{n,k}(s)$$

for $\lambda|_{[t_0,T]} \otimes \lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(s,t,\omega) \in [t_0,T] \times [t_0,T] \times \Omega$ and all $j \in \{1,\dots,m\}$, cf. Remark III.16. On the other hand, the evaluation of F_n yields

$$F_n(s) = \sum_{k=1}^{K_n} G_{n,k} \cdot h_{n,k}(s) \in \mathscr{S}(\Omega; E)$$

for $\lambda|_{[t_0,T]}$-almost all $s \in [t_0,T]$, and hence,

$$\mathrm{D}_t^j (F_n(s))(\omega) = \sum_{k=1}^{K_n} \mathrm{D}_t^j G_{n,k}(\omega)\, h_{n,k}(s)$$

for $\lambda|_{[t_0,T]} \otimes \lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(s,t,\omega) \in [t_0,T] \times [t_0,T] \times \Omega$ and all $j \in \{1,\dots,m\}$. That is, for all $n \in \mathbb{N}$, we have

$$\mathrm{D}_t^j F_n(s)(\omega) = \mathrm{D}_t^j (F_n(s))(\omega)$$

for $\lambda|_{[t_0,T]} \otimes \lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(s,t,\omega) \in [t_0,T] \times [t_0,T] \times \Omega$ and all $j \in \{1,\dots,m\}$. Since $F_n \to F$ in $\mathscr{D}^p(\Omega; L^2([t_0,T]; E))$ as $n \to \infty$, it also holds that

$$\mathrm{D}_t^j F(s)(\omega) = \mathrm{D}_t^j (F(s))(\omega)$$

for $\lambda|_{[t_0,T]} \otimes \lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(s,t,\omega) \in [t_0,T] \times [t_0,T] \times \Omega$ and all $j \in \{1,\dots,m\}$.

The spaces $\mathscr{D}^p(\Omega;\mathbb{R}^d)$ and $\mathscr{D}^p(\Omega;\mathbb{R})$ are connected in the following sense.

Remark III.19
Considering $x = (x^1, \dots, x^d)^{\mathrm{T}} \in \mathbb{R}^d$, it holds for the Euclidean norm of x that

$$\max_{i\in\{1,\dots,d\}} |x^i| \le \|x\| \le \sqrt{d} \max_{i\in\{1,\dots,d\}} |x^i| \tag{III.12}$$

for $x = (x^1, \dots, x^d)^{\mathrm{T}} \in \mathbb{R}^d$. Let $p \in [1,\infty[$. According to inequality (III.12)*, we obtain*

$$\max_{i\in\{1,\dots,d\}} \|F^i\|_{\mathscr{D}^p(\Omega;\mathbb{R})} \le \|F\|_{\mathscr{D}^p(\Omega;\mathbb{R}^d)} \le \sqrt{d} \max_{i\in\{1,\dots,d\}} \|F^i\|_{\mathscr{D}^p(\Omega;\mathbb{R})} \tag{III.13}$$

for all $F \in \mathscr{D}^p(\Omega;\mathbb{R}^d)$. That is, $F \in \mathscr{D}^p(\Omega;\mathbb{R}^d)$ if and only if $F^i \in \mathscr{D}^p(\Omega;\mathbb{R})$ for all $i \in \{1,\dots,d\}$. Due to this, the chain rule from Theorem III.9, for example, does also apply to $\mathbb{R}^d$-valued random variables in $\mathscr{D}^p(\Omega;\mathbb{R}^d)$.

By using the Malliavin derivative for H-valued random variables, some properties of the divergence operator δ are presented.

Lemma III.20 ([113, Proposition 1.3.1])
It holds $\mathscr{D}^2(\Omega;\mathrm{H}) \subset \operatorname{dom}\delta$ and

$$\mathrm{E}[\delta(F)\delta(G)] = \mathrm{E}\Big[\sum_{j=1}^m \int_{t_0}^T F_t^j G_t^j \,\mathrm{d}t\Big] + \mathrm{E}\Big[\sum_{j,l=1}^m \int_{t_0}^T \int_{t_0}^T (\mathrm{D}_s^j F_t^l)\,(\mathrm{D}_t^l G_s^j)\,\mathrm{d}s\,\mathrm{d}t\Big] \tag{III.14}$$

for all $F, G \in \mathscr{D}^2(\Omega;\mathrm{H})$.

For the representation of equation (III.14), we also refer to [112, Equation (1.54)]. Considering the second term on the right-hand side of equation (III.14) with $G = F \in \mathscr{D}^2(\Omega;\mathrm{H})$, we P-almost surely have

$$\Big|\sum_{j,l=1}^m \int_{t_0}^T \int_{t_0}^T (\mathrm{D}_s^j F_t^l)(\mathrm{D}_t^l F_s^j)\,\mathrm{d}s\,\mathrm{d}t\Big| \le \sum_{j,l=1}^m \int_{t_0}^T \int_{t_0}^T (\mathrm{D}_s^j F_t^l)^2\,\mathrm{d}s\,\mathrm{d}t = \|\mathrm{D}F\|_{\mathrm{H_H}}^2.$$

Thus, we obtain

$$\|\delta(F)\|_{L^2_{\mathscr{G}}(\Omega;\mathbb{R})} \le \big(\|F\|^2_{L^2_{\mathscr{G}}(\Omega;\mathrm{H})} + \|\mathrm{D}F\|^2_{L^2_{\mathscr{G}}(\Omega;\mathrm{H_H})}\big)^{\frac{1}{2}} = \|F\|_{\mathscr{D}^2(\Omega;\mathrm{H})}, \tag{III.15}$$

cf. [113, Equation (1.47)]. Next, we state a highly valuable property of the divergence operator.

Proposition III.21 ([113, Proposition 1.3.3])
Let $F \in \mathscr{D}^2(\Omega;\mathbb{R})$ and $G \in \operatorname{dom}\delta$ such that $FG \in L^2(\Omega;\mathrm{H})$. If $F\delta(G), \langle \mathrm{D}F, G\rangle_{\mathrm{H}} \in L^2(\Omega;\mathbb{R})$, then $FG \in \operatorname{dom}\delta$, and it holds

$$F\delta(G) = \delta(FG) + \langle \mathrm{D}F, G\rangle_{\mathrm{H}}$$

P-*almost surely.*

The divergence operator δ and the Itô integral are related as follows.

Proposition III.22 ([113, Proposition 1.3.11])
Let $F \in L^2_{\mathscr{G}}(\Omega; \mathrm{H})$ be adapted to filtration $(\mathscr{G}_t)_{t\in[t_0,T]}$. Then, $F \in \operatorname{dom}\delta$, and it P-almost surely holds

$$\delta(F) = \sum_{j=1}^{m} \int_{t_0}^{T} F_t^j \,\mathrm{d}W_t^j,$$

that is, the divergence of F coincides with the Itô integral of F.

Due to this, the divergence operator δ can be understand as an extension of the Itô integral to anticipative stochastic processes and is also called *Skorohod integral*, cf. [128]. For $F \in \operatorname{dom}\delta$, define

$$\sum_{j=1}^{m} \int_{t_0}^{T} F_t^j \,\delta W_t^j := \int_{t_0}^{T} F_t \,\delta W_t := \delta(F).$$

Thus, equation (III.14) can be seen as an extension of the Itô isometry, cf. [113, p. 42], and is referred to as covariance between Skorohod integrals, cf. [112, p. 39]. As in the case of the Itô stochastic integral, the notation

$$\int_{t_0}^{T} G_t \,\delta W_t := (\delta(G^1), \ldots, \delta(G^d))^{\mathrm{T}} =: \delta(G) \tag{III.16}$$

for $G = (G^1, \ldots, G^d)^{\mathrm{T}}$ is used, where $G^i \in \operatorname{dom}\delta$ for $i \in \{1, \ldots, d\}$.

The following proposition and lemma focus on the Malliavin derivative of an Itô integral and an integral over time, respectively. These properties are needed in order to derive the Malliavin differentiability of the solution of SDDE (II.1).

Proposition III.23 ([113, Lemma 1.3.4])
Let $F \in L^2_{\mathscr{G}}(\Omega; \mathrm{H})$ be adapted to filtration $(\mathscr{G}_t)_{t\in[t_0,T]}$, and consider the stochastic process $(G_t)_{t\in[t_0,T]}$ given by $G_t = \sum_{j=1}^{m} \int_{t_0}^{t} F_s^j \,\mathrm{d}W_s^j$. Then, it holds $G_T \in \mathscr{D}^2(\Omega; \mathbb{R})$ if and only if $F \in \mathscr{D}^2(\Omega; \mathrm{H})$. In this case, $(G_t)_{t\in[t_0,T]} \in \mathscr{D}^2(\Omega; L^2([t_0,T]; \mathbb{R}))$, and it holds for all $t \in [t_0, T]$ and $j \in \{1, \ldots, m\}$ that

$$\mathrm{D}_s^l \Bigg(\sum_{j=1}^{m} \int_{t_0}^{t} F_r^j \,\mathrm{d}W_r^j \Bigg)(\omega) = F_s^l(\omega) + \sum_{j=1}^{m} \int_{s}^{t} \mathrm{D}_s^l F_r^j \,\mathrm{d}W_r^j(\omega)$$

for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(s,\omega) \in [t_0, t] \times \Omega$ and all $l \in \{1, \ldots, m\}$ as well as that $\mathrm{D}_s^l G_t(\omega) = 0$ for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(s,\omega) \in \,]t, T] \times \Omega$.

The following lemma is for example used in [113, Equation (2.49)] without proof.

Lemma III.24
Let $F \in \mathscr{D}^p(\Omega; L^2([t_0,T]; \mathbb{R}^d))$ for some $p \in [1, \infty[$ and $G = \int_{t_0}^{T} F_s \,\mathrm{d}s$. Then, it holds $G \in \mathscr{D}^p(\Omega; \mathbb{R}^d)$ and

$$\mathrm{D}_t^j \int_{t_0}^{T} F_s \,\mathrm{d}s(\omega) = \int_{t_0}^{T} \mathrm{D}_t^j F_s(\omega) \,\mathrm{d}s$$

for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$*-almost all* $(t,\omega) \in [t_0,T] \times \Omega$ *and all* $j \in \{1,\dots,m\}$*. If in addition* $(F_t)_{t\in[t_0,T]}$ *is* $(\mathscr{G}_t)_{t\in[t_0,T]}$*-progressively measurable, it further holds*

$$\mathrm{D}_t^j \int_{t_0}^T F_s \,\mathrm{d}s(\omega) = \int_t^T \mathrm{D}_t^j F_s(\omega)\,\mathrm{d}s \tag{III.17}$$

for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$*-almost all* $(t,\omega) \in [t_0,T] \times \Omega$ *and all* $j \in \{1,\dots,m\}$*.*

Proof. The proof is stated in Section III.3, see p. 47. □

In order to show that term $\mathcal{R}_5$ in the proof of Theorem IV.9 is of order $\mathcal{O}(h)$ as $h \to 0$, we further need the boundedness of the Skorohod integral in $L^p_{\mathscr{G}}(\Omega;\mathbb{R})$.

Proposition III.25 ([113, Proposition 1.5.4])
Let $p \in \,]1,\infty[$*. The divergence operator* δ *is continuous from* $\mathscr{D}^p(\Omega;\mathrm{H})$ *to* $L^p_{\mathscr{G}}(\Omega;\mathbb{R})$*, and hence, there exists a constant* $c_{\delta,p} > 0$ *so that*

$$\|\delta(F)\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R})} \le c_{\delta,p}\|F\|_{\mathscr{D}^p(\Omega;\mathrm{H})} \tag{III.18}$$

for all $F \in \mathscr{D}^p(\Omega;\mathrm{H})$*.*

In view of Lemma III.20 and inequality (III.15), we have $c_{\delta,2} = 1$ in inequality (III.18). Considering the Skorohod integral as an extension of the Itô integral, inequality (III.18) can be regarded as the counterpart to Burkholder's inequality from Theorem II.4.

Proposition III.25 is only stated for Skorohod integrals that take values in $\mathbb{R}$. Using definition (III.16) and the triangle inequality, we can easily extend inequality (III.18) to Skorohod integrals that take values in $\mathbb{R}^d$. These considerations are detailed in the following because the resulting inequality is used the proof of Theorem IV.9, see term $\mathcal{R}_5'$ in formula (IV.116) in particular.

Let $F = (F^1,\dots,F^d)^{\mathrm{T}}$ with $F^\iota \in \mathscr{D}^p(\Omega;\mathrm{H})$ for $\iota \in \{1,\dots,d\}$. The triangle inequality implies

$$\|\delta(F)\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} = \left\| \sum_{\iota=1}^d |\delta(F^\iota)|^2 \right\|^{\frac{1}{2}}_{L^{\frac{p}{2}}_{\mathscr{G}}(\Omega;\mathbb{R})} \le \left(\sum_{\iota=1}^d \|\delta(F^\iota)\|^2_{L^p_{\mathscr{G}}(\Omega;\mathbb{R})} \right)^{\frac{1}{2}},$$

and then, Proposition III.25 yields

$$\|\delta(F)\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \le c_{\delta,p}\left(\sum_{\iota=1}^d \|F^\iota\|^2_{\mathscr{D}^p(\Omega;\mathrm{H})} \right)^{\frac{1}{2}}. \tag{III.19}$$

Next, we insert the definition of the norm $\|\cdot\|_{\mathscr{D}^p(\Omega;\mathrm{H})}$, see equation (III.11), and further estimate inequality (III.19). Using the inequality $(c_1^q + c_2^q)^{\frac{1}{q}} \le c_1 + c_2$ for $c_1, c_2 \in [0,\infty[$ and $q \in [1,\infty[$

additionally, we obtain

$$
\begin{aligned}
&\|\delta(F)\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \\
&\leq c_{\delta,p}\bigg(\sum_{\iota=1}^{d}\big(\|F^\iota\|^p_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})}+\|\mathrm{D}F^\iota\|^p_{L^p_{\mathscr{G}}(\Omega;\mathrm{H}_\mathrm{H})}\big)^{\frac{2}{p}}\bigg)^{\frac{1}{2}} \\
&= c_{\delta,p}\Bigg(\sum_{\iota=1}^{d}\Bigg(\mathrm{E}\Bigg[\bigg(\sum_{j=1}^{m}\int_{t_0}^{T}|F_u^{\iota,j}|^2\,\mathrm{d}u\bigg)^{\frac{2}{p}}\Bigg]+\mathrm{E}\Bigg[\bigg(\sum_{j_1,j_2=1}^{m}\int_{t_0}^{T}\int_{t_0}^{T}|\mathrm{D}_v^{j_2}F_u^{\iota,j_1}|^2\,\mathrm{d}v\,\mathrm{d}u\bigg)^{\frac{2}{p}}\Bigg]\Bigg)^{\frac{2}{p}}\Bigg)^{\frac{1}{2}} \\
&= c_{\delta,p}\Bigg(\sum_{\iota=1}^{d}\Bigg(\bigg\|\sum_{j=1}^{m}\int_{t_0}^{T}|F_u^{\iota,j}|^2\,\mathrm{d}u\bigg\|^{\frac{p}{2}}_{L^{\frac{p}{2}}_{\mathscr{G}}(\Omega;\mathbb{R})}+\bigg\|\sum_{j_1,j_2=1}^{m}\int_{t_0}^{T}\int_{t_0}^{T}|\mathrm{D}_v^{j_2}F_u^{\iota,j_1}|^2\,\mathrm{d}v\,\mathrm{d}u\bigg\|^{\frac{p}{2}}_{L^{\frac{p}{2}}_{\mathscr{G}}(\Omega;\mathbb{R})}\Bigg)^{\frac{2}{p}}\Bigg)^{\frac{1}{2}} \\
&\leq c_{\delta,p}\Bigg(\sum_{\iota=1}^{d}\bigg\|\sum_{j=1}^{m}\int_{t_0}^{T}|F_u^{\iota,j}|^2\,\mathrm{d}u\bigg\|_{L^{\frac{p}{2}}_{\mathscr{G}}(\Omega;\mathbb{R})}+\bigg\|\sum_{j_1,j_2=1}^{m}\int_{t_0}^{T}\int_{t_0}^{T}|\mathrm{D}_v^{j_2}F_u^{\iota,j_1}|^2\,\mathrm{d}v\,\mathrm{d}u\bigg\|_{L^{\frac{p}{2}}_{\mathscr{G}}(\Omega;\mathbb{R})}\Bigg)^{\frac{1}{2}}.
\end{aligned}
\tag{III.20}
$$

III.2. Malliavin Derivative of Stochastic Delay Differential Equations

The Malliavin derivative of the solution of SDDE (II.1) is studied in this section, see Theorem III.26 below. Similar results have been obtained by Yan, see [137, Proposition 7.4] and [60, Proposition 3.1], and earlier by Hirsch [58, Theorem 3.1].

The statements in [60, 137] are however not entirely true. The initial condition of the considered SDDE is assumed to be random and stochastically independent of the Wiener process, see [137, Equation (7.1)] and [60, Equation (1.6)]. But the solution of the SDDE is not differentiable then, in the sense of Malliavin, as it is not $\mathscr{G}$-measurable, cf. definitions III.1, III.6, III.11, and III.15. Thus, the initial condition considered in [137, Proposition 7.4] and [60, Proposition 3.1] should be deterministic.

We refer to [113, Theorem 2.2.1] for the Malliavin derivative of solutions of SODEs. The proofs of [137, Proposition 7.4] and [60, Proposition 3.1] are based on the proof of [113, Theorem 2.2.1]. We present a similar but different proof, which is based on more elementary techniques, namely we do not use [112, Proposition 1.5.5 and Lemma 1.5.4]. Moreover, we state the upper bound in the $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$-norm of the Malliavin derivative of the solution of SDDE (II.1) by greater detail on the constants than in [137, Proposition 7.4] and [60, Proposition 3.1] or [113, Theorem 2.2.1] in case of SODEs.

In the theorem below, the initial condition is assumed to be deterministic. In the analysis of term $\mathcal{R}_5$ in proof of Theorem IV.9, we transfer SDDE (II.1) with stochastic initial condition to an SDDE with deterministic initial condition in order to apply Theorem III.26 below. For more details on that, we refer to the proof of Theorem IV.9 and in particular to Lemma IV.19.

Theorem III.26
Consider SDDE (II.1) *with initial condition* $\xi = x\colon [t_0-\tau,t_0]\to\mathbb{R}^d$ *being a deterministic càdlàg function. Let the Borel-measurable drift* a *and diffusion* b^j, $j\in\{1,\dots,m\}$, *of SDDE* (II.1)

satisfy the global Lipschitz and linear growth conditions (II.8)*,* (II.9)*,* (II.10)*, and* (II.11)*. Further, let* $a(t, t-\tau_1, \dots, t-\tau_D, \cdot, \dots, \cdot)$, $b^j(t, t-\tau_1, \dots, t-\tau_D, \cdot, \dots, \cdot) \in C^1(\mathbb{R}^{d\times(D+1)}; \mathbb{R}^d)$ *for all* $j \in \{1, \dots, m\}$ *and* $t \in [t_0, T]$*.*

Then, it holds for the solution X *of SDDE* (II.1) *with initial condition* x *that* $X_t \in \mathscr{D}^p(\Omega; \mathbb{R}^d)$ *for all* $t \in [t_0 - \tau, T]$ *and all* $p \in [2, \infty[$*. For all* $s \in [t_0, T]$ *and* $j \in \{1, \dots, m\}$*, the Malliavin derivative* $\mathrm{D}_s^j X = (\mathrm{D}_s^j X^1, \dots, \mathrm{D}_s^j X^d)^{\mathrm{T}}$ *is the unique strong solution of the d-dimensional linear SDDE*

$$\mathrm{D}_s^j X_t = \begin{cases} 0, & t \in [t_0 - \tau, s[, \\ b^j(\mathcal{T}(s, X_s)) + \int_s^t \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(u, X_u))\, \mathrm{D}_s^j X_{u-\tau_l}^i \,\mathrm{d}u & \\ \qquad + \sum_{k=1}^{m} \int_s^t \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} b^k(\mathcal{T}(u, X_u))\, \mathrm{D}_s^j X_{u-\tau_l}^i \,\mathrm{d}W_u^k, & t \in [s, T], \end{cases} \tag{III.21}$$

and it holds

$$\max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \le C_{\mathrm{D},p} \big(1 + \|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \tag{III.22}$$

for all $p \in [2, \infty[$*, where*

$$C_{\mathrm{D},p} := \sqrt{2} K_b \mathrm{e}^{d(D+1)^2 \left(\sqrt{T-t_0} L_a + \frac{p L_b \sqrt{m}}{\sqrt{p-1}}\right)^2 (T-t_0)}.$$

Proof. The proof is stated in Section III.3, see p. 49. □

III.3. Proofs

Proof of Theorem III.9

Proof of Theorem III.9. The Malliavin derivative can only be calculated for smooth random variables explicitly yet, cf. Definition III.3. Other Malliavin derivatives of random variables in the closure $\mathscr{D}^p(\Omega; \mathbb{R})$ are then obtained as the limit of smooth random variables. Thus, the random variable $\varphi(F)$ has to be approximated by smooth random variables in order to show the assertion of Theorem III.9.

In the following, we approximate random variable $F_l \in \mathscr{D}^p(\Omega; \mathbb{R})$ by a sequence $(F_{l,n})_{n\in\mathbb{N}}$ with $F_{l,n} \in \mathscr{S}(\Omega; \mathbb{R})$ for all $l \in \{1, \dots, L\}$ and function φ by a sequence of functions in $C_p^\infty(\mathbb{R}^L; \mathbb{R})$ as indicated in [113, p. 28]. The latter can be done by mollification, see e. g. [40, Section C.5].

Let $\psi \in C^\infty(\mathbb{R}^L; \mathbb{R})$ be compactly supported with

$$\int_{\mathbb{R}^L} \psi(x)\,\mathrm{d}x = 1.$$

For example, one can choose

$$\psi(x) = \begin{cases} c\, \mathrm{e}^{\frac{1}{\|x\|^2-1}}, & \|x\| < 1, \\ 0 & \|x\| \geq 1, \end{cases}$$

where $c > 0$ is a constant such that $\int_{\mathbb{R}^L} \psi(x)\, \mathrm{d}x = 1$. Further, for $\varepsilon > 0$ and $x \in \mathbb{R}^L$, we define

$$\psi_\varepsilon(x) := \varepsilon^{-L}\psi\big(\frac{1}{\varepsilon}x\big),$$

where $\int_{\mathbb{R}^L} \psi_\varepsilon(x)\, \mathrm{d}x = 1$ and $\psi_\varepsilon \in C^\infty(\mathbb{R}^L;\mathbb{R})$, cf. [40, p. 713]. Following [40, p. 714], we consider the convolution $\varphi_\varepsilon := \varphi * \psi_\varepsilon \colon \mathbb{R}^L \to \mathbb{R}$ defined by

$$(\varphi * \psi_\varepsilon)(x) := \int_{\mathbb{R}^L} \varphi(y)\psi_\varepsilon(x-y)\, \mathrm{d}y = \int_{\mathbb{R}^L} \varphi(x-y)\psi_\varepsilon(y)\, \mathrm{d}y = (\psi_\varepsilon * \varphi)(x).$$

Since ψ_ε belongs to $C^\infty(\mathbb{R}^L;\mathbb{R})$, we have $\varphi_\varepsilon \in C^\infty(\mathbb{R}^L;\mathbb{R})$. Because ψ is compactly supported, the continuity of φ, and because of

$$\varphi_\varepsilon(x) - \varphi(x) = \int_{\mathbb{R}^L} \psi_\varepsilon(y)\big(\varphi(x-y) - \varphi(x)\big)\, \mathrm{d}y = \int_{\mathbb{R}^L} \psi(y)\big(\varphi(x-\varepsilon y) - \varphi(x)\big)\, \mathrm{d}y,$$

where the first equality follows from

$$\int_{\mathbb{R}^L} \psi(x)\, \mathrm{d}x = \int_{\mathbb{R}^L} \psi_\varepsilon(x)\, \mathrm{d}x = 1$$

and the second equality from a substitution, it follows by the dominated convergence theorem that $\lim_{\varepsilon\to 0} \varphi_\varepsilon(x) = \varphi(x)$ for all $x \in \mathbb{R}^L$. Hence, ψ is a mollifier. Similarly, since

$$\partial_{x_l}\varphi_\varepsilon = \partial_{x_l}(\varphi * \psi_\varepsilon) = (\partial_{x_l}\varphi) * \psi_\varepsilon$$

and $\partial_{x_l}\varphi$ is continuous, it also holds for all $l \in \{1,\dots,L\}$ and $x \in \mathbb{R}^L$ that

$$\lim_{\varepsilon\to 0} \partial_{x_l}\varphi_\varepsilon(x) = \partial_{x_l}\varphi(x).$$

In the following, we show that φ_ε is of polynomial growth in order to obtain $\varphi_\varepsilon \in C_p^\infty(\mathbb{R}^L;\mathbb{R})$. According to the assumption $|\partial_{x_l}\varphi(x)| \leq C\big(1+\|x\|^2\big)^{\frac{\chi}{2}}$ for all $x \in \mathbb{R}^L$, we claim that $|\varphi(x)| \leq \tilde{C}\big(1+\|x\|^2\big)^{\frac{\chi+1}{2}}$ for all $x \in \mathbb{R}^L$, where $\tilde{C} > 0$ is a constant. In fact, using the triangle inequality, the mean value theorem [57, p. 278], and using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|\varphi(x)| - |\varphi(0)| &\leq |\varphi(x) - \varphi(0)| \\
&= \Big|\sum_{l=1}^{L} \int_0^1 \partial_{x_l}\varphi(\theta x)\, \mathrm{d}\theta\, x_l\Big| \\
&\leq \sum_{l=1}^{L} \int_0^1 \big|\partial_{x_l}\varphi(\theta x)\big|\, \mathrm{d}\theta\, |x_l| \\
&\leq C\big(1+\|x\|^2\big)^{\frac{\chi}{2}} \sum_{l=1}^{L} |x_l| \\
&\leq C\big(1+\|x\|^2\big)^{\frac{\chi}{2}} L^{\frac{1}{2}} \|x\| \\
&\leq C\big(1+\|x\|^2\big)^{\frac{\chi}{2}} L^{\frac{1}{2}} \big(1+\|x\|^2\big)^{\frac{1}{2}} \\
&\leq C L^{\frac{1}{2}} \big(1+\|x\|^2\big)^{\frac{\chi+1}{2}},
\end{aligned}$$

and it follows

$$|\varphi(x)| \le \big(|\varphi(0)| + CL^{\frac{1}{2}}\big)\big(1 + \|x\|^2\big)^{\frac{\chi+1}{2}} \tag{III.23}$$

for all $x \in \mathbb{R}^L$. Let $\psi_\varepsilon^{(k)}$ be the kth derivative, $k \in \mathbb{N}_0$, of ψ_ε. For a function $g \in C(\mathbb{R}^L;\mathbb{R})$ with $|g(x)| \le \hat{C}\big(1 + \|x\|^2\big)^{\frac{\nu}{2}}$ for all $x \in \mathbb{R}^L$ and an exponent $\nu \in [0,\infty[$, it holds

$$\begin{aligned}
|(\psi_\varepsilon^{(k)} * g)(x)| &\le \int_{\mathbb{R}^L} \big|\psi_\varepsilon^{(k)}(y)\big||g(x-y)|\,\mathrm{d}y \\
&\le \hat{C}\int_{\mathbb{R}^L} \big|\psi_\varepsilon^{(k)}(y)\big|\big(1 + \|x-y\|^2\big)^{\frac{\nu}{2}}\,\mathrm{d}y \\
&\le \hat{C}\int_{\mathbb{R}^L} \big|\psi_\varepsilon^{(k)}(y)\big|\big(1 + 2\|x\|^2 + 2\|y\|^2\big)^{\frac{\nu}{2}}\,\mathrm{d}y \\
&= \hat{C}\int_{[-\varepsilon,\varepsilon]^L} \big|\psi_\varepsilon^{(k)}(y)\big|\big(1 + 2\|x\|^2 + 2\|y\|^2\big)^{\frac{\nu}{2}}\,\mathrm{d}y \\
&\le \hat{C}\int_{[-\varepsilon,\varepsilon]^L} \big|\psi_\varepsilon^{(k)}(y)\big|\big(1 + 2\|x\|^2 + 2L\varepsilon^2\big)^{\frac{\nu}{2}}\,\mathrm{d}y \\
&\le \hat{C}\big(2(1+L\varepsilon^2)\big)^{\frac{\nu}{2}}\int_{[-\varepsilon,\varepsilon]^L} \big|\psi_\varepsilon^{(k)}(y)\big|\,\mathrm{d}y\big(1 + \|x\|^2\big)^{\frac{\nu}{2}} \\
&= \hat{C}\big(2(1+L\varepsilon^2)\big)^{\frac{\nu}{2}}\int_{[-\varepsilon,\varepsilon]^L} \varepsilon^{-(L+k)}\left|\psi^{(k)}\Big(\frac{y}{\varepsilon}\Big)\right|\mathrm{d}y\big(1 + \|x\|^2\big)^{\frac{\nu}{2}} \\
&= \hat{C}\big(2(1+L\varepsilon^2)\big)^{\frac{\nu}{2}}\int_{[-1,1]^L} \varepsilon^{-k}\big|\psi^{(k)}(z)\big|\,\mathrm{d}z\big(1 + \|x\|^2\big)^{\frac{\nu}{2}} \\
&= \hat{C}\big(2(1+L\varepsilon^2)\big)^{\frac{\nu}{2}}\int_{[-1,1]^L} \big|\psi^{(k)}(z)\big|\,\mathrm{d}z\,\varepsilon^{-k}\big(1 + \|x\|^2\big)^{\frac{\nu}{2}},
\end{aligned}$$

that is, for every fixed $\varepsilon > 0$, the mollification $\psi_\varepsilon^{(k)} * g$ has polynomial growth of order ν. Thus, we have $\varphi_\varepsilon \in C_p^\infty(\mathbb{R}^L;\mathbb{R})$, and in particular, we obtain, using the considerations above with $k = 0$, that

$$|\varphi_\varepsilon(x)| \le \tilde{C}\big(2(1+L\varepsilon^2)\big)^{\frac{\chi+1}{2}}\big(1 + \|x\|^2\big)^{\frac{\chi+1}{2}} \tag{III.24}$$

and

$$|\partial_{x_l}\varphi_\varepsilon(x)| \le C\big(2(1+L\varepsilon^2)\big)^{\frac{\chi}{2}}\big(1 + \|x\|^2\big)^{\frac{\chi}{2}}$$

for all $x \in \mathbb{R}^L$.

We continue with the approximation of $\varphi(F)$ with smooth random variables. Since $F_l \in \mathscr{D}^p(\Omega;\mathbb{R})$, there exist sequences $(F_{l,n})_{n\in\mathbb{N}}$ with $F_{l,n} \in \mathscr{S}(\Omega;\mathbb{R})$ and $\lim_{n\to} F_{l,n} = F_l$ in $\mathscr{D}^p(\Omega;\mathbb{R})$ for all $l \in \{1,\dots,L\}$. Moreover, $F_{l,n} \in \mathscr{S}(\Omega;\mathbb{R})$ has a representation

$$F_{l,n} = f_{l,n}\bigg(\int_{t_0}^T h_{1,l,n}(s)\,\mathrm{d}W_s, \dots, \int_{t_0}^T h_{K_{l,n},l,n}(s)\,\mathrm{d}W_s\bigg),$$

where $f_{l,n} \in C_p^\infty(\mathbb{R}^{K_{l,n}};\mathbb{R})$ and $h_{k,l,n} \in \mathrm{H}$ for all $k \in \{1,\dots,K_{l,n}\}$, $l \in \{1,\dots,L\}$, and $n \in \mathbb{N}$. Because of $\varphi_\varepsilon \in C_p^\infty(\mathbb{R}^L;\mathbb{R})$, the function

$$\begin{aligned}
&(y_{1,1,n},\dots,y_{K_{1,n},1,n},\dots,y_{1,L,n},\dots,y_{K_{L,n},L,n}) \\
&\quad\mapsto \varphi_\varepsilon \circ \Big(f_{1,n}(y_{1,1,n},\dots,y_{K_{1,n},1,n}),\dots,f_{L,n}(y_{1,L,n},\dots,y_{K_{L,n},L,n})\Big)
\end{aligned}$$

belongs to $C_p^\infty(\mathbb{R}^{\sum_{l=1}^L K_{l,n}};\mathbb{R})$, and thus, we have $\varphi_\varepsilon(F_n) = \varphi_\varepsilon(F_{1,n},\dots,F_{L,n}) \in \mathscr{S}(\Omega;\mathbb{R})$. Due to this, we can calculate the Malliavin derivative of $\varphi_\varepsilon(F_n)$ using Definition III.3 and obtain

$$\begin{aligned} \mathrm{D}\varphi_\varepsilon(F_n) &= \sum_{l=1}^L \sum_{k_l=1}^{K_{l,n}} \partial_{x_l}\varphi_\varepsilon(F_n)\partial_{y_{k_l}} f_{l,n}\bigg(\int_{t_0}^T h_{1,l,n}(s)\,\mathrm{d}W_s,\dots,\int_{t_0}^T h_{K_{l,n},l,n}(s)\,\mathrm{d}W_s\bigg)h_{k_l,l,n} \\ &= \sum_{l=1}^L \partial_{x_l}\varphi_\varepsilon(F_n) \sum_{k_l=1}^{K_{l,n}} \partial_{y_{k_l}} f_{l,n}\bigg(\int_{t_0}^T h_{1,l,n}(s)\,\mathrm{d}W_s,\dots,\int_{t_0}^T h_{K_{l,n},l,n}(s)\,\mathrm{d}W_s\bigg)h_{k_l,l,n} \\ &= \sum_{l=1}^L \partial_{x_l}\varphi_\varepsilon(F_n)\,\mathrm{D}F_{l,n}. \end{aligned}$$

Now, it is left to prove that

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}\|\varphi_\varepsilon(F_n)-\varphi(F)\|_{\mathscr{D}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})} = 0. \tag{III.25}$$

Without loss of generality, let $0<\varepsilon\le 1$. Then, using the polynomial growth of order $\chi+1$ of φ and φ_ε, see inequalities (III.23) and (III.24), we have

$$|\varphi(F)| \le \tilde{C}\big(1+\|F\|^2\big)^{\frac{\chi+1}{2}} \in L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R}),$$

$$|\varphi_\varepsilon(F)| \le \tilde{C}(2+2L)^{\frac{\chi+1}{2}}\big(1+\|F\|^2\big)^{\frac{\chi+1}{2}} \in L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R}),$$

and

$$|\varphi_\varepsilon(F_n)| \le \tilde{C}(2+2L)^{\frac{\chi+1}{2}}\big(1+\|F_n\|^2\big)^{\frac{\chi+1}{2}} \in L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})$$

because $F_l, F_{l,n} \in L_{\mathscr{G}}^p(\Omega;\mathbb{R})$. The triangle inequality and the dominated convergence theorem imply

$$\begin{aligned} &\lim_{n\to\infty}\lim_{\varepsilon\to 0}\|\varphi_\varepsilon(F_n)-\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})} \\ &\quad\le \lim_{n\to\infty}\lim_{\varepsilon\to 0}\|\varphi_\varepsilon(F_n)-\varphi_\varepsilon(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})} + \lim_{\varepsilon\to 0}\|\varphi_\varepsilon(F)-\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})} \\ &\quad= \lim_{n\to\infty}\Big\|\lim_{\varepsilon\to 0}\varphi_\varepsilon(F_n)-\varphi_\varepsilon(F)\Big\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})} + \Big\|\lim_{\varepsilon\to 0}\varphi_\varepsilon(F)-\varphi(F)\Big\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})} \\ &\quad= \lim_{n\to\infty}\|\varphi(F_n)-\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})} + 0. \end{aligned}$$

Due to Vitali's convergence theorem, see e. g. [38, p. 262] or [74, Proposition 4.12], $\lim_{n\to\infty} F_n = F$ in $L_{\mathscr{G}}^p(\Omega;\mathbb{R}^L)$ is equivalent to $\lim_{n\to\infty} F_n = F$ in probability and $(\|F_n\|^p)_{n\in\mathbb{N}}$ is uniformly integrable. Since φ is continuous, it also holds $\lim_{n\to\infty}\varphi(F_n) = \varphi(F)$ in probability [67, Theorem 17.5]. According to the growth condition of φ, we P-almost surely have

$$|\varphi(F_n)|^{\frac{p}{\chi+1}} \le \tilde{C}\big(1+\|F_n\|^2\big)^{\frac{p}{2}},$$

and the uniformly integrability of family $(\|F_n\|^p)_{n\in\mathbb{N}}$ implies that $(|\varphi(F_n)|^{\frac{p}{\chi+1}})_{n\in\mathbb{N}}$ is uniformly integrable as well [76, Theorem 6.18]. Using again Vitali's convergence theorem, it follows

$$\lim_{n\to\infty}\|\varphi(F_n)-\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})} = 0,$$

and thus

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}\|\varphi_\varepsilon(F_n)-\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})}=0. \tag{III.26}$$

In the following, we show that $\lim_{n\to\infty}\lim_{\varepsilon\to 0}\sum_{l=1}^{L}\partial_{x_l}\varphi_\varepsilon(F_n)\,\mathrm{D}F_{l,n}=\sum_{l=1}^{L}\partial_{x_l}\varphi(F)\,\mathrm{D}F_l$ in $L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})$. Let $L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})=L_{\mathscr{G}}^{\infty}(\Omega;\mathbb{R})$ in case of $\chi=0$, where $L_{\mathscr{G}}^{\infty}(\Omega;\mathbb{R})$ is the Banach space of all essentially bounded and $\mathscr{G}/\mathscr{B}(\mathbb{R})$-measurable random variables $Z\colon\Omega\to\mathbb{R}$. That is

$$\|Z\|_{L_{\mathscr{G}}^{\infty}(\Omega;\mathbb{R})}:=\|Z\|_{L^{\infty}(\Omega;\mathbb{R})}=\operatorname*{ess\,sup}_{\omega\in\Omega}|Z(\omega)|=\inf_{\mathcal{N}\subset\Omega:\mathrm{P}|_{\mathscr{G}}[\mathcal{N}]=0}\ \sup_{\omega\in\Omega\setminus\mathcal{N}}|Z(\omega)|<\infty.$$

Using the growth condition on $\partial_{x_l}\varphi$ and since $F_l\in\mathscr{D}^p(\Omega;\mathbb{R})$ for $l\in\{1,\dots,L\}$, we obtain

$$\|\partial_{x_l}\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})}\le\left\|C\big(1+\|F\|^2\big)^{\frac{\chi}{2}}\right\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})}<\infty,$$

and thus, we have, using the triangle and Hölder's inequality with $\frac{\chi+1}{p}=\frac{\chi}{p}+\frac{1}{p}$, that

$$\begin{aligned}\left\|\sum_{l=1}^{L}\partial_{x_l}\varphi(F)\,\mathrm{D}F_l\right\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})}&\le\sum_{l=1}^{L}\|\partial_{x_l}\varphi(F)\,\mathrm{D}F_l\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})}\\&=\sum_{l=1}^{L}\big\|\|\partial_{x_l}\varphi(F)\,\mathrm{D}F_l\|_{\mathrm{H}}\big\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})}\\&=\sum_{l=1}^{L}\big\||\partial_{x_l}\varphi(F)|\|\mathrm{D}F_l\|_{\mathrm{H}}\big\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})}\\&\le\sum_{l=1}^{L}\big\||\partial_{x_l}\varphi(F)|\big\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})}\big\|\|\mathrm{D}F_l\|_{\mathrm{H}}\big\|_{L_{\mathscr{G}}^{p}(\Omega;\mathbb{R})}\\&=\sum_{l=1}^{L}\|\partial_{x_l}\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})}\|\mathrm{D}F_l\|_{L_{\mathscr{G}}^{p}(\Omega;\mathrm{H})}<\infty.\end{aligned}$$

Similar considerations as the ones above provide $\partial_{x_l}\varphi(F_n)$, $\partial_{x_l}\varphi_\varepsilon(F)$, $\partial_{x_l}\varphi_\varepsilon(F_n)\in L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})$ as well as $\sum_{l=1}^{L}\partial_{x_l}\varphi_\varepsilon(F_n)\,\mathrm{D}F_{l,n}\in L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})$.

In order to show the convergence in formula (III.25), we now consider the $L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})$-norm within the graph norm $\|\cdot\|_{\mathscr{D}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})}$. The convergence in $L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})$ is already stated in equation (III.26). Using the triangle inequality, it holds

$$\begin{aligned}&\left\|\sum_{l=1}^{L}\partial_{x_l}\varphi_\varepsilon(F_n)\,\mathrm{D}F_{l,n}-\sum_{l=1}^{L}\partial_{x_l}\varphi(F)\,\mathrm{D}F_l\right\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})}\\&\quad\le\sum_{l=1}^{L}\|\partial_{x_l}\varphi_\varepsilon(F_n)(\mathrm{D}F_{l,n}-\mathrm{D}F_l)\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})}+\|(\partial_{x_l}\varphi_\varepsilon(F_n)-\partial_{x_l}\varphi(F_n))\,\mathrm{D}F_l\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})}\\&\qquad+\|(\partial_{x_l}\varphi(F_n)-\partial_{x_l}\varphi(F))\,\mathrm{D}F_l\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})}.\end{aligned} \tag{III.27}$$

At first, consider the case of $\chi = 0$ in the following. Then, we have $|\partial_{x_l}\varphi(x)| \leq C$ and $|\partial_{x_l}\varphi_\varepsilon(x)| \leq C$ for all $x \in \mathbb{R}^L$. Hence, it follows

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}\|\partial_{x_l}\varphi_\varepsilon(F_n)(\mathrm{D}F_{l,n} - \mathrm{D}F_l)\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})} \leq \lim_{n\to\infty} C\|\mathrm{D}F_{l,n} - \mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})} = 0.$$

Using $\|(\partial_{x_l}\varphi_\varepsilon(F_n) - \partial_{x_l}\varphi(F_n))\,\mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})} \leq 2C\|\mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})}$, the dominated convergence theorem implies

$$\begin{aligned}
&\lim_{n\to\infty}\lim_{\varepsilon\to 0}\|(\partial_{x_l}\varphi_\varepsilon(F_n) - \partial_{x_l}\varphi(F_n))\,\mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})} \\
&\quad = \lim_{n\to\infty}\left\|\lim_{\varepsilon\to 0}(\partial_{x_l}\varphi_\varepsilon(F_n) - \partial_{x_l}\varphi(F_n))\,\mathrm{D}F_l\right\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})} \\
&\quad = 0.
\end{aligned}$$

Since $\lim_{n\to\infty} F_n = F$ in $L^p_{\mathscr{G}}(\Omega;\mathbb{R}^L)$, it also holds $\lim_{n\to\infty} F_n = F$ in probability, and this implies, together with the continuity of $\partial_{x_l}\varphi$, that $\lim_{n\to\infty}\partial_{x_l}\varphi(F_n) = \partial_{x_l}\varphi(F)$ in probability [67, Theorem 17.5]. Using that $\|(\partial_{x_l}\varphi(F_n) - \partial_{x_l}\varphi(F))\,\mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})} \leq 2C\|\mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})}$ for all $n \in \mathbb{N}$, [67, Theorem 17.4] yields

$$\lim_{n\to\infty}\|(\partial_{x_l}\varphi(F_n) - \partial_{x_l}\varphi(F))\,\mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})} = 0.$$

In the case of $\chi = 0$, we thus have

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}\left\|\sum_{l=1}^L \partial_{x_l}\varphi_\varepsilon(F_n)\,\mathrm{D}F_{l,n} - \sum_{l=1}^L \partial_{x_l}\varphi(F)\,\mathrm{D}F_l\right\|_{L^{\frac{p}{\chi+1}}_{\mathscr{G}}(\Omega;\mathrm{H})} = 0. \tag{III.28}$$

In the following, let $\chi \in \,]0, p-1]$. Using inequality (III.27) and Hölder's inequality, it holds

$$\begin{aligned}
&\left\|\sum_{l=1}^L \partial_{x_l}\varphi_\varepsilon(F_n)\,\mathrm{D}F_{l,n} - \sum_{l=1}^L \partial_{x_l}\varphi(F)\,\mathrm{D}F_l\right\|_{L^{\frac{p}{\chi+1}}_{\mathscr{G}}(\Omega;\mathrm{H})} \\
&\quad \leq \sum_{l=1}^L \|\partial_{x_l}\varphi_\varepsilon(F_n)\|_{L^{\frac{p}{\chi}}_{\mathscr{G}}(\Omega;\mathbb{R})}\|\mathrm{D}F_{l,n} - \mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})} \\
&\qquad + \|\partial_{x_l}\varphi_\varepsilon(F_n) - \partial_{x_l}\varphi(F_n)\|_{L^{\frac{p}{\chi}}_{\mathscr{G}}(\Omega;\mathbb{R})}\|\mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})} \\
&\qquad + \|\partial_{x_l}\varphi(F_n) - \partial_{x_l}\varphi(F)\|_{L^{\frac{p}{\chi}}_{\mathscr{G}}(\Omega;\mathbb{R})}\|\mathrm{D}F_l\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H})}.
\end{aligned} \tag{III.29}$$

Without loss of generality, let $0 < \varepsilon \leq 1$ again. Since we have

$$\begin{aligned}
|\partial_{x_l}\varphi_\varepsilon(F_n)| &\leq C\big(2(1 + L\varepsilon^2)\big)^{\frac{\chi}{2}}\big(1 + \|F_n\|^2\big)^{\frac{\chi}{2}} \\
&\leq C(2 + 2L)^{\frac{\chi}{2}}\big(1 + \|F_n\|^2\big)^{\frac{\chi}{2}} \in L^{\frac{p}{\chi}}_{\mathscr{G}}(\Omega;\mathbb{R}),
\end{aligned}$$

we obtain by the dominated convergence theorem that

$$\lim_{\varepsilon\to 0}\|\partial_{x_l}\varphi_\varepsilon(F_n)\|_{L^{\frac{p}{\chi}}_{\mathscr{G}}(\Omega;\mathbb{R})} = \|\lim_{\varepsilon\to 0}\partial_{x_l}\varphi_\varepsilon(F_n)\|_{L^{\frac{p}{\chi}}_{\mathscr{G}}(\Omega;\mathbb{R})} = \|\partial_{x_l}\varphi(F_n)\|_{L^{\frac{p}{\chi}}_{\mathscr{G}}(\Omega;\mathbb{R})}.$$

According to Vitali's convergence theorem, due to the continuity of $\partial_{x_l}\varphi$, and because

$$|\partial_{x_l}\varphi(F_n)|^{\frac{p}{\chi}} \leq C^{\frac{p}{\chi}}\big(1 + \|F_n\|^2\big)^{\frac{p}{2}}$$

P-almost surely, it follows $\lim_{n\to\infty} \partial_{x_l}\varphi(F_n) = \partial_{x_l}\varphi(F)$ in probability and $(|\partial_{x_l}\varphi_\varepsilon(F_n)|^{\frac{p}{\chi}})_{n\in\mathbb{N}}$ is uniformly integrable. Thus, we obtain

$$\lim_{n\to\infty} \|\partial_{x_l}\varphi(F_n) - \partial_{x_l}\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})} = 0, \tag{III.30}$$

and therewith

$$\begin{aligned}
&\lim_{n\to\infty} \|\partial_{x_l}\varphi(F_n)\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})} \|\mathrm{D}F_{l,n} - \mathrm{D}F_l\|_{L_{\mathscr{G}}^{p}(\Omega;\mathrm{H})} \\
&\quad = \lim_{n\to\infty} \|\partial_{x_l}\varphi(F_n)\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})} \lim_{n\to\infty} \|\mathrm{D}F_{l,n} - \mathrm{D}F_l\|_{L_{\mathscr{G}}^{p}(\Omega;\mathrm{H})} \\
&\quad = \|\partial_{x_l}\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})} \cdot 0 \\
&\quad = 0.
\end{aligned}$$

Since $0 < \varepsilon \leq 1$, we have

$$\begin{aligned}
|\partial_{x_l}\varphi_\varepsilon(F_n) - \partial_{x_l}\varphi(F_n)| &\leq C\Big(\big(2(1+L\varepsilon^2)\big)^{\frac{\chi}{2}} + 1\Big)\big(1+\|F_n\|^2\big)^{\frac{\chi}{2}} \\
&\leq C\big((2+2L)^{\frac{\chi}{2}} + 1\big)\big(1+\|F_n\|^2\big)^{\frac{\chi}{2}} \in L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R}),
\end{aligned}$$

and the dominated convergence theorem implies

$$\lim_{\varepsilon\to 0} \|\partial_{x_l}\varphi_\varepsilon(F_n) - \partial_{x_l}\varphi(F_n)\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})} = \|\lim_{\varepsilon\to 0} \partial_{x_l}\varphi_\varepsilon(F_n) - \partial_{x_l}\varphi(F_n)\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})} = 0.$$

Further, we obtain, using equation (III.30), that

$$\lim_{n\to\infty} \|\partial_{x_l}\varphi(F_n) - \partial_{x_l}\varphi(F)\|_{L_{\mathscr{G}}^{\frac{p}{\chi}}(\Omega;\mathbb{R})} \|\mathrm{D}F_l\|_{L_{\mathscr{G}}^{p}(\Omega;\mathrm{H})} = 0,$$

and thus, the right-hand side of inequality (III.29) converges to zero as $\varepsilon \to 0$ and $n \to \infty$. With equation (III.28), we have

$$\lim_{n\to\infty} \lim_{\varepsilon\to 0} \left\| \sum_{l=1}^{L} \partial_{x_l}\varphi_\varepsilon(F_n)\, \mathrm{D}F_{l,n} - \sum_{l=1}^{L} \partial_{x_l}\varphi(F)\, \mathrm{D}F_l \right\|_{L_{\mathscr{G}}^{\frac{p}{\chi+1}}(\Omega;\mathrm{H})} = 0$$

for all $\chi \in [0, p-1]$ in conclusion, that is, $\big((\varphi_\varepsilon(F_n))_{\varepsilon>0}\big)_{n\in\mathbb{N}}$ converges in $\mathscr{D}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})$ as $\varepsilon \to 0$ and $n \to \infty$. Thus, it holds $\varphi(F) \in \mathscr{D}^{\frac{p}{\chi+1}}(\Omega;\mathbb{R})$ and $\mathrm{D}\varphi(F) = \sum_{l=1}^{L} \partial_{x_l}\varphi(F)\, \mathrm{D}F_l$ P-almost surely according to the closeness of operator D. □

Proof of Lemma III.24

Proof of Lemma III.24. According to the assumption $F \in \mathscr{D}^p(\Omega; L^2([t_0,T];\mathbb{R}^d))$, there exists a sequence of $L^2([t_0,T];\mathbb{R}^d)$-valued smooth random variables $(F_k)_{k\in\mathbb{N}}$ such that $F_k \to F$ in $\mathscr{D}^p(\Omega; L^2([t_0,T];\mathbb{R}^d))$ as $k \to \infty$. Since $F_k \in \mathscr{S}(\Omega; L^2([t_0,T];\mathbb{R}^d))$, we can assume that $F_k = \sum_{l=1}^{n_k} \tilde{F}_{k,l} \cdot h_{k,l}$ where $\tilde{F}_{k,l} \in \mathscr{S}(\Omega;\mathbb{R})$ and $h_{k,l} \in L^2([t_0,T];\mathbb{R}^d)$.

Defining $G_k := \int_{t_0}^T F_k(s)\,\mathrm{d}s$, we have

$$G_k = \sum_{l=1}^{n_k} \tilde{F}_{k,l} \cdot \int_{t_0}^T h_{k,l}(s)\,\mathrm{d}s \in \mathscr{S}(\Omega;\mathbb{R}^d). \tag{III.31}$$

Since $F_k \to F$ in $\mathscr{D}^p(\Omega; L^2([t_0,T];\mathbb{R}^d)) \subset L^p_{\mathscr{G}}(\Omega; L^2([t_0,T];\mathbb{R}^d))$ as $k \to \infty$, the Cauchy-Schwarz inequality implies $G_k \to G$ in $L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)$ as $k \to \infty$.

According to equation (III.31) and Definition III.13, we obtain by linearity

$$\begin{aligned}
\mathrm{D}_t^j G_k(\omega) &= \mathrm{D}_t^j\Big(\sum_{l=1}^{n_k} \tilde{F}_{k,l} \cdot \int_{t_0}^T h_{k,l}(s)\,\mathrm{d}s\Big)(\omega) = \sum_{l=1}^{n_k} \mathrm{D}_t^j \tilde{F}_{k,l}(\omega) \int_{t_0}^T h_{k,l}(s)\,\mathrm{d}s \\
&= \int_{t_0}^T \sum_{l=1}^{n_k} \mathrm{D}_t^j \tilde{F}_{k,l}(\omega) h_{k,l}(s)\,\mathrm{d}s = \int_{t_0}^T \mathrm{D}_t^j F_k(s)(\omega)\,\mathrm{d}s
\end{aligned}$$

for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(t,\omega) \in [t_0,T] \times \Omega$ and all $j \in \{1,\dots,m\}$. Now, we use the convergence $\mathrm{D}F_k \to \mathrm{D}F$ in $L^p_{\mathscr{G}}(\Omega; \mathrm{H}_{L^2([t_0,T];\mathbb{R}^d)})$ as $k \to \infty$ in order to show $\mathrm{D}G_k \to \int_{t_0}^T \mathrm{D}F(s)\,\mathrm{d}s$ in $L^p_{\mathscr{G}}(\Omega; \mathrm{H}_{\mathbb{R}^d})$ as $k \to \infty$. By rewriting the norms as well as applying the triangle inequality and the Cauchy-Schwarz inequality, it holds

$$\begin{aligned}
&\Big\| \int_{t_0}^T \mathrm{D}F_k(s)\,\mathrm{d}s - \int_{t_0}^T \mathrm{D}F(s)\,\mathrm{d}s \Big\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H}_{\mathbb{R}^d})} \\
&= \Big\| \int_{t_0}^T \mathrm{D}F_k(s)\,\mathrm{d}s - \int_{t_0}^T \mathrm{D}F(s)\,\mathrm{d}s \Big\|_{L^p_{\mathscr{G}}(\Omega;L^2([t_0,T];L_{HS}(\mathbb{R}^m;\mathbb{R}^d)))} \\
&= \Big\| \int_{t_0}^T \mathrm{D}F_k(s) - \mathrm{D}F(s)\,\mathrm{d}s \Big\|_{L^p_{\mathscr{G}}(\Omega;L^2([t_0,T];L_{HS}(\mathbb{R}^m;\mathbb{R}^d)))} \\
&= \left(\mathrm{E}\left[\left|\int_{t_0}^T \sum_{j=1}^m \Big\| \int_{t_0}^T \mathrm{D}_t F_k(s) - \mathrm{D}_t F(s)\,\mathrm{d}s\, e_j \Big\|^2 \mathrm{d}t\right|^{\frac{p}{2}}\right]\right)^{\frac{1}{p}} \\
&= \left(\mathrm{E}\left[\left|\int_{t_0}^T \sum_{j=1}^m \Big\| \int_{t_0}^T \big(\mathrm{D}_t F_k(s) - \mathrm{D}_t F(s)\big)\, e_j\,\mathrm{d}s \Big\|^2 \mathrm{d}t\right|^{\frac{p}{2}}\right]\right)^{\frac{1}{p}} \\
&\le \sqrt{T-t_0}\left(\mathrm{E}\left[\left|\int_{t_0}^T \sum_{j=1}^m \int_{t_0}^T \big\|\big(\mathrm{D}_t F_k(s) - \mathrm{D}_t F(s)\big)\, e_j\big\|^2 \mathrm{d}s\,\mathrm{d}t\right|^{\frac{p}{2}}\right]\right)^{\frac{1}{p}} \\
&= \sqrt{T-t_0}\left(\mathrm{E}\left[\left|\int_{t_0}^T \sum_{j=1}^m \big\|\big(\mathrm{D}_t F_k - \mathrm{D}_t F\big)\, e_j\big\|^2_{L^2([t_0,T];\mathbb{R}^d)} \mathrm{d}t\right|^{\frac{p}{2}}\right]\right)^{\frac{1}{p}} \\
&= \sqrt{T-t_0}\left(\mathrm{E}\left[\left|\int_{t_0}^T \|\mathrm{D}_t F_k - \mathrm{D}_t F\|^2_{L_{HS}(\mathbb{R}^m;L^2([t_0,T];\mathbb{R}^d))} \mathrm{d}t\right|^{\frac{p}{2}}\right]\right)^{\frac{1}{p}} \\
&= \sqrt{T-t_0}\|\mathrm{D}F_k - \mathrm{D}F\|_{L^p_{\mathscr{G}}(\Omega;L^2([t_0,T];L_{HS}(\mathbb{R}^m;L^2([t_0,T];\mathbb{R}^d))))} \\
&= \sqrt{T-t_0}\|\mathrm{D}F_k - \mathrm{D}F\|_{L^p_{\mathscr{G}}(\Omega;\mathrm{H}_{L^2([t_0,T];\mathbb{R}^d)})},
\end{aligned}$$

where $(e_j)_{j\in\{1,\dots,m\}}$ is the canonical orthonormal basis of $\mathbb{R}^m$. Letting $k \to \infty$, the right-hand side of the inequality above converges to zero. Thus, we have $G_k \to G$ in $\mathscr{D}^p(\Omega;\mathbb{R}^d)$ as

$k \to \infty$ and $\mathrm{D}_t^j G(\omega) = \int_{t_0}^T \mathrm{D}_t^j F(s)(\omega)\,\mathrm{d}s$ for $\lambda|_{[t_0,T]} \otimes \mathrm{P}|_{\mathscr{G}}$-almost all $(t,\omega) \in [t_0,T] \times \Omega$ and all $j \in \{1,\dots,m\}$. Equation (III.17) finally follows from Corollary III.8. □

Proof of Theorem III.26

The proof is similar to the proofs of [137, Proposition 7.4] and [60, Proposition 3.1] as well as [113, Theorem 2.2.1] in case of SODEs. However, we waive the use of [112, Proposition 1.5.5 and Lemma 1.5.4] and utilize simpler facts instead.

Proof of Theorem III.26. At first, we show the existence and uniqueness of the solution of SDDE (III.21) for arbitrary fixed $s \in [t_0,T]$ and $j \in \{1,\dots,m\}$ using Theorem II.12.

Choosing

$$\xi_t = \begin{cases} 0, & t \in [t_0 - \tau, s[, \\ b^j(\mathcal{T}(s, X_s)), & t = s, \end{cases}$$

$$\begin{aligned} &A(t, t-\tau_1, \dots, t-\tau_D, z_0, z_1, \dots, z_D, x_0, x_1, \dots, x_D) \\ &\quad = \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} a(t, t-\tau_1, \dots, t-\tau_D, z_0, z_1, \dots, z_D) x_l^i, \end{aligned}$$

and

$$\begin{aligned} &B^k(t, t-\tau_1, \dots, t-\tau_D, z_0, z_1, \dots, z_D, x_0, x_1, \dots, x_D) \\ &\quad = \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} b^k(t, t-\tau_1, \dots, t-\tau_D, z_0, z_1, \dots, z_D) x_l^i \end{aligned}$$

for all $t \in [s,T]$, $k \in \{1,\dots,m\}$, and $x_l, z_l \in \mathbb{R}^d$, we recover SDDE (II.13). Using the Lipschitz conditions (II.8) and (II.9) as well as the Cauchy-Schwarz inequality, it holds

$$\begin{aligned} &\sup_{\substack{t \in [s,T] \\ z_l \in \mathbb{R}^d\colon\, l \in \{0,1,\dots,D\}}} \left\| \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} a(t, t-\tau_1, \dots, t-\tau_D, z_0, z_1, \dots, z_D)(x_l^i - y_l^i) \right\| \\ &\quad \le L_a \sum_{l=0}^{D} \sum_{i=1}^{d} |x_l^i - y_l^i| \\ &\quad \le L_a \sqrt{d} \sum_{l=0}^{D} \|x_l - y_l\| \\ &\quad \le L_a \sqrt{d}(D+1) \sup_{l \in \{0,1,\dots,D\}} \|x_l - y_l\| \end{aligned} \tag{III.32}$$

for all $x_l, y_l \in \mathbb{R}^d$, $l \in \{0,1,\ldots,D\}$. Choosing $y_l = 0$ for all $l \in \{0,1,\ldots,D\}$ in inequality (III.32) above, we obtain

$$\begin{aligned}
&\sup_{\substack{t\in[s,T]\\ z_l\in\mathbb{R}^d\colon\, l\in\{0,1,\ldots,D\}}} \left\| \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} a(t, t-\tau_1, \ldots, t-\tau_D, z_0, z_1, \ldots, z_D) x_l^i \right\| \\
&\quad \leq L_a \sqrt{d}(D+1) \sup_{l\in\{0,1,\ldots,D\}} \|x_l\| \\
&\quad \leq L_a \sqrt{d}(D+1) \sup_{l\in\{0,1,\ldots,D\}} \left(1 + \|x_l\|^2\right)^{\frac{1}{2}}
\end{aligned} \tag{III.33}$$

for all $x_l \in \mathbb{R}^d$, $l \in \{0,1,\ldots,D\}$. Similarly, we have

$$\begin{aligned}
&\sup_{\substack{t\in[s,T]\\ z_l\in\mathbb{R}^d\colon\, l\in\{0,1,\ldots,D\}}} \max_{k\in\{1,\ldots,m\}} \left\| \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} b^k(t, t-\tau_1, \ldots, t-\tau_D, z_0, z_1, \ldots, z_D)(x_l^i - y_l^i) \right\| \\
&\quad \leq L_b \sqrt{d}(D+1) \sup_{l\in\{0,1,\ldots,D\}} \|x_l - y_l\|
\end{aligned} \tag{III.34}$$

and

$$\begin{aligned}
&\sup_{\substack{t\in[s,T]\\ z_l\in\mathbb{R}^d\colon\, l\in\{0,1,\ldots,D\}}} \max_{k\in\{1,\ldots,m\}} \left\| \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} b^k(t, t-\tau_1, \ldots, t-\tau_D, z_0, z_1, \ldots, z_D) x_l^i \right\| \\
&\quad \leq L_b \sqrt{d}(D+1) \sup_{l\in\{0,1,\ldots,D\}} \|x_l\| \\
&\quad \leq L_b \sqrt{d}(D+1) \sup_{l\in\{0,1,\ldots,D\}} \left(1 + \|x_l\|^2\right)^{\frac{1}{2}}
\end{aligned} \tag{III.35}$$

for all $x_l, y_l \in \mathbb{R}^d$, $l \in \{0,1,\ldots,D\}$. Thus, the global Lipschitz and linear growth conditions (II.14), (II.15), (II.16), and (II.17) of SDDE (II.13) are fulfilled with $L_A = K_A = L_a\sqrt{d}(D+1)$ and $L_B = K_B = L_b\sqrt{d}(D+1)$. Using further the linear growth condition (II.11) and that the initial condition x is deterministic, we obtain

$$\|\xi\|_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} = \|b^j(\mathcal{T}(s, X_s))\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \leq K_b\left(1 + \|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\right)^{\frac{1}{2}}. \tag{III.36}$$

Then, Theorem II.12 provides the existence and uniqueness of a solution of SDDE (III.21) for arbitrary fixed $s \in [t_0, T]$ and $j \in \{1,\ldots,m\}$. This solution further belongs to $S^p([t_0-\tau, T] \times \Omega; \mathbb{R}^d)$, cf. inequality (II.18).

In the following, we derive a sharper estimate of the solution of SDDE (III.21) than in inequality (II.18) from inequalities (III.33) and (III.35). Similarly to the estimates (II.25), (II.26), and (II.27), it holds

$$
\begin{aligned}
&\|\mathrm{D}^j_s X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq 2\|b^j(\mathcal{T}(s,X_s))\|^2_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} + 2\Bigg(\bigg\|\int_s^\cdot \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x^i_l} a(\mathcal{T}(u,X_u))\,\mathrm{D}^j_s X^i_{u-\tau_l}\,\mathrm{d}u\bigg\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad + \bigg\|\sum_{k=1}^{m}\int_s^\cdot \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x^i_l} b^k(\mathcal{T}(u,X_u))\,\mathrm{D}^j_s X^i_{u-\tau_l}\,\mathrm{d}W^k_u\bigg\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)}\Bigg)^2 \\
&\quad\leq 2\|b^j(\mathcal{T}(s,X_s))\|^2_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \\
&\qquad + 2d(D+1)^2\left(\sqrt{T-t_0}L_a + \frac{pL_b\sqrt{m}}{\sqrt{p-1}}\right)^2 \int_s^T \|\mathrm{D}^j_s X\|^2_{S^p([t_0-\tau,u]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}u,
\end{aligned}
$$

and Gronwall's Lemma II.7 implies

$$
\begin{aligned}
&\max_{j\in\{1,\dots,m\}}\sup_{s\in[t_0,T]}\|\mathrm{D}^j_s X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq 2\max_{j\in\{1,\dots,m\}}\sup_{s\in[t_0,T]}\|b^j(\mathcal{T}(s,X_s))\|^2_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \mathrm{e}^{2d(D+1)^2\left(\sqrt{T-t_0}L_a+\frac{pL_b\sqrt{m}}{\sqrt{p-1}}\right)^2(T-t_0)}.
\end{aligned}
\tag{III.37}
$$

Since the initial condition of SDDE (II.1) is assumed to be deterministic in this theorem, inequality (III.37) above holds true for all $p\in[2,\infty[$, cf. Theorem II.8. Then, inequality (III.22) follows by taking the square root of inequality (III.37) and using inequality (III.36).

In the following, we show that solution X of SDDE (II.1) with the deterministic initial condition x is differentiable in the sense of Malliavin and its Malliavin derivative is in fact the solution of SDDE (III.21). So far and for the time being, $\mathrm{D}^j_s X$ is only the name for the unique strong solution of SDDE (III.21), where $s\in[t_0,T]$ and $j\in\{1,\dots,m\}$ are arbitrary fixed and has nothing to do with the Malliavin derivative yet.

As in the proof of Theorem II.8, we consider the Picard's iterations

$$
X^{(0)}_t := \begin{cases} x_t, & t\in[t_0-\tau,t_0], \\ x_{t_0}, & t\in\,]t_0,T], \end{cases}
$$

and

$$
X^{(\ell+1)}_t := \begin{cases} x_t, & t\in[t_0-\tau,t_0], \\ x_{t_0} + \int_{t_0}^t a(\mathcal{T}(u,X^{(\ell)}_u))\,\mathrm{d}u + \sum_{k=1}^{m}\int_{t_0}^t b^k(\mathcal{T}(u,X^{(\ell)}_u))\,\mathrm{d}W^k_u, & t\in\,]t_0,T] \end{cases}
$$

for $\ell\in\mathbb{N}_0$. According to equation (II.29), we have

$$
\lim_{\ell\to\infty}\|X^{(\ell)}_t - X_t\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} = 0 \tag{III.38}
$$

for all $t\in[t_0-\tau,T]$ and $p\in[2,\infty[$.

In the following, we show that $X^{(\ell)}_t\in\mathscr{D}^p(\Omega;\mathbb{R}^d)$ is the Malliavin differentiability for all $t\in[t_0,T]$, $\ell\in\mathbb{N}_0$, and $p\in[2,\infty[$.

Let $p \in [2, \infty[$ be arbitrary fixed in the sequel.

Since $x_t \in \mathbb{R}^d$ for all $t \in [t_0 - \tau, t_0]$, x_t is an $\mathbb{R}^d$-valued smooth random variable, and we have $\mathrm{D}X_t^{(0)} = 0$ by Definition III.13 for all $t \in [t_0 - \tau, T]$. Thus, it holds $X^{(0)} \in \mathscr{D}^p(\Omega; \mathbb{R}^d)$ for all $t \in [t_0 - \tau, T]$.

According to Remark III.19, we also have $X_t^{(0),i} \in \mathscr{D}^p(\Omega; \mathbb{R})$ for all $t \in [t_0 - \tau, T]$ and $i \in \{1, \ldots, d\}$. Due to the assumptions on coefficients a and b^k, $k \in \{1, \ldots, m\}$, the assumptions of Theorem III.9 are fulfilled with $\chi = 0$. Using Lemma III.24, Proposition III.23 in addition, we obtain $X_t^{(1)} \in \mathscr{D}^p(\Omega; \mathbb{R}^d)$ for all $t \in [t_0 - \tau, T]$, and it holds

$$\mathrm{D}_s^j X_t^{(1)} = \begin{cases} 0, & t \in [t_0 - \tau, s[, \\ b^j(\mathcal{T}(s, X_s^{(0)})) + \displaystyle\int_s^t \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} a\big(\mathcal{T}(u, X_u^{(0)})\big)\, \mathrm{D}_s^j X_{u-\tau_l}^{(0),i} \,\mathrm{d}u & \\ \quad + \displaystyle\sum_{k=1}^{m} \int_s^t \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} b^k\big(\mathcal{T}(u, X_u^{(0)})\big)\, \mathrm{D}_s^j X_{u-\tau_l}^{(0),i} \,\mathrm{d}W_u^k, & t \in [s, T], \end{cases} \tag{III.39}$$

$$= \begin{cases} 0, & t \in [t_0 - \tau, s[, \\ b^j(\mathcal{T}(s, X_s^{(0)})), & t \in [s, T], \end{cases}$$

for all $(s, \omega) \in [t_0, T] \times \Omega$ and all $j \in \{1, \ldots, m\}$. Because of linear growth condition (II.11) and $X^{(0)}$ being a càdlàg function, same result could be obtained by considering $X_t^{(1)}$ as an $\mathbb{R}^d$-valued smooth random variable and applying Definition III.13.

Using the same arguments in the derivation of equation (III.39), it follows inductively over $\ell \in \mathbb{N}$ that $X_t^{(\ell)} \in \mathscr{D}^p(\Omega; \mathbb{R}^d)$ for all $t \in [t_0 - \tau, T]$, where

$$\mathrm{D}_s^j X_t^{(\ell+1)}(\omega) = \begin{cases} 0, \quad t \in [t_0 - \tau, s[, & \\ b^j(\mathcal{T}(s, X_s^{(\ell)}(\omega))) + \displaystyle\int_s^t \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} a\big(\mathcal{T}(u, X_u^{(\ell)}(\omega))\big)\, \mathrm{D}_s^j X_{u-\tau_l}^{(\ell),i}(\omega) \,\mathrm{d}u & \\ \quad + \Big(\displaystyle\sum_{k=1}^{m} \int_s^t \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} b^k\big(\mathcal{T}(u, X_u^{(\ell)})\big)\, \mathrm{D}_s^j X_{u-\tau_l}^{(\ell),i} \,\mathrm{d}W_u^k\Big)(\omega), & t \in [s, T], \end{cases} \tag{III.40}$$

for $\lambda|_{[t_0,T]} \times \mathrm{P}|_{\mathscr{G}}$-almost all $(s, \omega) \in [t_0, T] \times \Omega$ and all $j \in \{1, \ldots, m\}$.

In the following, we show

$$\lim_{\ell \to \infty} \|\mathrm{D}X_t^{(\ell)} - \mathrm{D}X_t\|_{L^p_{\mathscr{G}}(\Omega; \mathrm{H}_{\mathbb{R}^d})} = 0 \tag{III.41}$$

for all $t \in [t_0 - \tau, T]$ so that with equation (III.38), we obtain

$$\lim_{\ell \to \infty} \|X_t^{(\ell)} - X_t\|_{\mathscr{D}^p(\Omega; \mathbb{R}^d)} = 0 \tag{III.42}$$

for all $t \in [t_0 - \tau, T]$. It is only when the convergence in equation (III.41) holds true that $\mathrm{D}_s^j X_t$ really is the Malliavin derivative of X_t and its name meaningful. So far, we only know that the solution at point in time $t \in [t_0 - \tau, T]$ of SDDE (III.21) denoted by $\mathrm{D}_s^j X_t$ exists, where $s \in [t_0, T]$ and $j \in \{1, \dots, m\}$.

In order to show the convergence in equation (III.41), we first make the following considerations. By rewriting the norms, using triangle inequality and taking the supremum over time, we have

$$\begin{aligned}
&\|\mathrm{D}X_t^{(\ell)} - \mathrm{D}X_t\|_{L^p(\Omega;\mathrm{H}_{\mathbb{R}^d})} \\
&\quad = \|\mathrm{D}X_t^{(\ell)} - \mathrm{D}X_t\|_{L^p(\Omega;L^2([t_0,T];L_{HS}(\mathbb{R}^m;\mathbb{R}^d)))} \\
&\quad = \left(\mathrm{E}\left[\left(\sum_{j=1}^m \int_{t_0}^T \|\mathrm{D}_s^j X_t^{(\ell)} - \mathrm{D}_s^j X_t\|^2 \,\mathrm{d}s\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}} \\
&\quad = \left\|\sum_{j=1}^m \int_{t_0}^T \|\mathrm{D}_s^j X_t^{(\ell)} - \mathrm{D}_s^j X_t\|^2 \,\mathrm{d}s\right\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})}^{\frac{1}{2}} \\
&\quad \le \left(\sum_{j=1}^m \int_{t_0}^T \|\mathrm{D}_s^j X_t^{(\ell)} - \mathrm{D}_s^j X_t\|_{L^p(\Omega;\mathbb{R}^d)}^2 \,\mathrm{d}s\right)^{\frac{1}{2}} \\
&\quad \le \sqrt{T-t_0}\sqrt{m} \max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X_t^{(\ell)} - \mathrm{D}_s^j X_t\|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad \le \sqrt{T-t_0}\sqrt{m} \max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(\ell)} - \mathrm{D}_s^j X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}.
\end{aligned} \tag{III.43}$$

If the right-hand side of inequality (III.43) above converges to zero as $\ell \to \infty$, the left-hand side thus converges to zero as well. Subsequently, we show

$$\lim_{\ell\to\infty} \max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(\ell+1)} - \mathrm{D}_s^j X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = 0 \tag{III.44}$$

in order to prove equation (III.41). The same considerations used in the derivation of inequality (III.22), also cf. inequality (II.27), yield

$$\begin{aligned}
&\max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(\ell+1)}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le \sqrt{2}K_b\big(1 + \|X^{(\ell)}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}^2\big)^{\frac{1}{2}} \mathrm{e}^{d(D+1)^2\left(\sqrt{T-t_0}L_a + \frac{pL_b\sqrt{m}}{\sqrt{p-1}}\right)^2(T-t_0)}
\end{aligned} \tag{III.45}$$

for all $\ell \in \mathbb{N}_0$ and $\max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(0)}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = 0$. Inserting inequality (II.27) into inequality (III.45), we obtain

$$\begin{aligned}
&\max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(\ell)}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le \sqrt{2}K_b\Big(1 + 2\sup_{t\in[t_0-\tau,t_0]} \|x_t\|^2\Big)^{\frac{1}{2}} \mathrm{e}^{\left(\left(\sqrt{T-t_0}K_a + \frac{pK_b\sqrt{m}}{\sqrt{p-1}}\right)^2 + d(D+1)^2\left(\sqrt{T-t_0}L_a + \frac{pL_b\sqrt{m}}{\sqrt{p-1}}\right)^2\right)(T-t_0)}
\end{aligned} \tag{III.46}$$

for all $\ell \in \mathbb{N}_0$. Thus, the right-hand side of inequality (III.43) is finite for all $\ell \in \mathbb{N}_0$. Considering the norm on the right-hand side of inequality (III.43) and inserting the representations (III.40) and (III.21), we obtain by rewriting and applying the triangle inequality

$$\begin{aligned}
&\|\mathrm{D}_s^j X^{(\ell+1)} - \mathrm{D}_s^j X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq \|b^j(\mathcal{T}(s,X_s^{(\ell)})) - b^j(\mathcal{T}(s,X_s))\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \\
&\qquad+ \Big\| \int_s^{\cdot} \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(u,X_u^{(\ell)}))\big(\mathrm{D}_s^j X_{u-\tau_l}^{(\ell),i} - \mathrm{D}_s^j X_{u-\tau_l}^i\big)\,\mathrm{d}u \Big\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad+ \Big\| \int_s^{\cdot} \sum_{l=0}^{D}\sum_{i=1}^{d} \big(\partial_{x_l^i} a(\mathcal{T}(u,X_u^{(\ell)})) - \partial_{x_l^i} a(\mathcal{T}(u,X_u))\big)\,\mathrm{D}_s^j X_{u-\tau_l}^i\,\mathrm{d}u \Big\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad+ \Big\| \sum_{k=1}^{m}\int_s^{\cdot} \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} b^k(\mathcal{T}(u,X_u^{(\ell)}))\big(\mathrm{D}_s^j X_{u-\tau_l}^{(\ell),i} - \mathrm{D}_s^j X_{u-\tau_l}^i\big)\,\mathrm{d}W_u^k \Big\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad+ \Big\| \sum_{k=1}^{m}\int_s^{\cdot} \sum_{l=0}^{D}\sum_{i=1}^{d} \big(\partial_{x_l^i} b^k(\mathcal{T}(u,X_u^{(\ell)})) - \partial_{x_l^i} b^k(\mathcal{T}(u,X_u))\big)\mathrm{D}_s^j X_{u-\tau_l}^i\,\mathrm{d}W_u^k \Big\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)}
\end{aligned} \tag{III.47}$$

for all $s \in [t_0,T]$, $j \in \{1,\dots,m\}$, and $\ell \in \mathbb{N}_0$. Using the Lipschitz conditions (II.9), (III.32), and (III.34), it follows, similarly to estimates (II.22) and (II.23), by triangle inequality, Hölder's inequality, and Zakai's inequality from Theorem II.6 for the first, second, and fourth term on the right-hand side of inequality (III.47) that

$$\|b^j(\mathcal{T}(s,X_s^{(\ell)})) - b^j(\mathcal{T}(s,X_s))\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \leq L_b\|X^{(\ell)} - X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)},$$

$$\begin{aligned}
&\Big\| \int_s^{\cdot} \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(u,X_u^{(\ell)}))\big(\mathrm{D}_s^j X_{u-\tau_l}^{(\ell),i} - \mathrm{D}_s^j X_{u-\tau_l}^i\big)\,\mathrm{d}u \Big\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq L_a\sqrt{d}(D+1)\int_s^T \|\mathrm{D}_s^j X^{(\ell)} - \mathrm{D}_s^j X\|_{S^p([t_0-\tau,u]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}u \\
&\quad\leq L_a\sqrt{d}(D+1)\sqrt{T-t_0}\bigg(\int_{t_0}^T \|\mathrm{D}_s^j X^{(\ell)} - \mathrm{D}_s^j X\|^2_{S^p([t_0-\tau,u]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}u\bigg)^{\frac{1}{2}},
\end{aligned}$$

and

$$\begin{aligned}
&\Big\| \sum_{k=1}^{m}\int_s^{\cdot} \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} b^k(\mathcal{T}(u,X_u^{(\ell)}))\big(\mathrm{D}_s^j X_{u-\tau_l}^{(\ell),i} - \mathrm{D}_s^j X_{u-\tau_l}^i\big)\,\mathrm{d}W_u^k \Big\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq \frac{pL_b\sqrt{d}(D+1)\sqrt{m}}{\sqrt{p-1}}\bigg(\int_{t_0}^T \|\mathrm{D}_s^j X^{(\ell)} - \mathrm{D}_s^j X\|^2_{S^p([t_0-\tau,u]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}u\bigg)^{\frac{1}{2}}.
\end{aligned}$$

Using the triangle inequality and Zakai's inequality from Theorem II.6, we estimate the third and fifth term on the right-hand side of inequality (III.47) and obtain

$$\begin{aligned}
&\Big\| \int_s^{\cdot} \sum_{l=0}^{D}\sum_{i=1}^{d} \big(\partial_{x_l^i} a(\mathcal{T}(u,X_u^{(\ell)})) - \partial_{x_l^i} a(\mathcal{T}(u,X_u))\big)\,\mathrm{D}_s^j X_{u-\tau_l}^i\,\mathrm{d}u \Big\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq \int_{t_0}^T \Big\| \sum_{l=0}^{D}\sum_{i=1}^{d} \big(\partial_{x_l^i} a(\mathcal{T}(u,X_u^{(\ell)})) - \partial_{x_l^i} a(\mathcal{T}(u,X_u))\big)\,\mathrm{D}_s^j X_{u-\tau_l}^i \Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}\,\mathrm{d}u
\end{aligned}$$

as well as

$$\begin{aligned}&\Big\| \sum_{k=1}^{m} \int_{s}^{\cdot} \sum_{l=0}^{D} \sum_{i=1}^{d} \big(\partial_{x_l^i} b^k(\mathcal{T}(u, X_u^{(\ell)})) - \partial_{x_l^i} b^k(\mathcal{T}(u, X_u))\big) \mathrm{D}_s^j X_{u-\tau_l}^i \, \mathrm{d}W_u^k \Big\|_{S^p([s,T]\times\Omega;\mathbb{R}^d)} \\ &\quad\le \frac{p}{\sqrt{p-1}} \\ &\qquad \times \bigg(\int_{t_0}^{T} \sum_{k=1}^{m} \Big\| \sum_{l=0}^{D} \sum_{i=1}^{d} \big(\partial_{x_l^i} b^k(\mathcal{T}(u, X_u^{(\ell)})) - \partial_{x_l^i} b^k(\mathcal{T}(u, X_u))\big) \mathrm{D}_s^j X_{u-\tau_l}^i \Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^2 \mathrm{d}u \bigg)^{\frac{1}{2}}.\end{aligned}$$

Inserting the previous five estimates into inequality (III.47) and square both sides of the inequality, we have, using inequality (II.6) for all $\ell \in \mathbb{N}_0$, that

$$\begin{aligned}&\max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(\ell+1)} - \mathrm{D}_s^j X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}^2 \\ &\quad\le C^{(\ell)} + C \int_{t_0}^{T} \max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(\ell)} - \mathrm{D}_s^j X\|_{S^p([t_0-\tau,u]\times\Omega;\mathbb{R}^d)}^2 \, \mathrm{d}u\end{aligned} \tag{III.48}$$

where

$$\begin{aligned}C^{(\ell)} \coloneqq 2\Bigg(&L_b \|X^{(\ell)} - X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\ &+ \int_{t_0}^{T} \max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \Big\| \sum_{l=0}^{D} \sum_{i=1}^{d} \big(\partial_{x_l^i} a(\mathcal{T}(u, X_u^{(\ell)})) \\ &\qquad - \partial_{x_l^i} a(\mathcal{T}(u, X_u))\big) \mathrm{D}_s^j X_{u-\tau_l}^i \Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \mathrm{d}u \\ &+ \frac{p}{\sqrt{p-1}} \bigg(\int_{t_0}^{T} \sum_{k=1}^{m} \max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \Big\| \sum_{l=0}^{D} \sum_{i=1}^{d} \big(\partial_{x_l^i} b^k(\mathcal{T}(u, X_u^{(\ell)})) \\ &\qquad - \partial_{x_l^i} b^k(\mathcal{T}(u, X_u))\big) \mathrm{D}_s^j X_{u-\tau_l}^i \Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^2 \mathrm{d}u \bigg)^{\frac{1}{2}} \Bigg)^2\end{aligned}$$

and

$$C \coloneqq 2d(D+1)^2 \bigg(L_a \sqrt{T-t_0} + \frac{pL_b\sqrt{m}}{\sqrt{p-1}} \bigg)^2.$$

In the following, we show $\lim_{\ell\to\infty} C^{(\ell)} = 0$. Equation (II.29) states

$$\lim_{\ell\to\infty} \|X^{(\ell)} - X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = 0, \tag{III.49}$$

so we only need to consider the other two terms of $C^{(\ell)}$. We consider the second term first and show its convergence to zero as $\ell \to \infty$. The convergence in equation (III.49) implies

$$\sup_{t\in[t_0-\tau,T]} \|X_t^{(\ell)} - X_t\| \to 0$$

in probability as $\ell \to \infty$, that is

$$\lim_{\ell\to\infty} \mathrm{P}|_{\mathscr{G}}\Big[\omega \in \Omega : \sup_{t\in[t_0-\tau,T]} \|X_t^{(\ell)}(\omega) - X_t(\omega)\| > \varepsilon\Big] = 0$$

for all $\varepsilon > 0$, [67, Theorem 17.2]. According to the assumptions on the continuity of the derivatives $\partial_{x_l^i} a$ and $\partial_{x_l^i} b^k$, we have

$$\lim_{\ell\to\infty} \mathrm{P}|_{\mathscr{G}}\Big[\omega \in \Omega : \big\|\partial_{x_l^i} a(\mathcal{T}(u, X_u^{(\ell)}(\omega))) - \partial_{x_l^i} a(\mathcal{T}(u, X_u(\omega)))\big\| > \varepsilon\Big] = 0$$

for all $\varepsilon > 0$ and $u \in [t_0, T]$ using [67, Theorem 17.5]. Then, it holds

$$\lim_{\ell\to\infty} \mathrm{P}|_{\mathscr{G}}\Bigg[\omega \in \Omega \colon \Bigg\|\sum_{l=0}^{D}\sum_{i=1}^{d}\big(\partial_{x_l^i} a(\mathcal{T}(u, X_u^{(\ell)}(\omega))) - \partial_{x_l^i} a(\mathcal{T}(u, X_u(\omega)))\big)\mathrm{D}_s^j X_{u-\tau_l}^i(\omega)\Bigg\| > \varepsilon\Bigg] = 0$$

as well for all $\varepsilon > 0$ and $u \in [t_0, T]$. Due to the Lipschitz continuity of the drift coefficient a, its partial derivatives are bounded by L_a in the Euclidean norm. With the boundedness of $\mathrm{D}_s^j X_{u-\tau_l}^i$, see inequality (III.22) and Theorem II.8, we obtain the uniform integrability of the family

$$\left(\left\|\sum_{l=0}^{D}\sum_{i=1}^{d}\big(\partial_{x_l^i} a(\mathcal{T}(u, X_u^{(\ell)})) - \partial_{x_l^i} a(\mathcal{T}(u, X_u))\big)\,\mathrm{D}_s^j X_{u-\tau_l}^i\right\|^p\right)_{\ell\in\mathbb{N}_0}.$$

Then, Vitali's convergence theorem, see e. g. [38, p. 262] or [74, Proposition 4.12], implies

$$\lim_{\ell\to\infty} \max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \left\|\sum_{l=0}^{D}\sum_{i=1}^{d}\big(\partial_{x_l^i} a(\mathcal{T}(u, X_u^{(\ell)})) - \partial_{x_l^i} a(\mathcal{T}(u, X_u))\big)\,\mathrm{D}_s^j X_{u-\tau_l}^i\right\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} = 0$$

for all $u \in [t_0, T]$. Finally, using the dominated convergence theorem, it holds

$$\begin{aligned}\lim_{\ell\to\infty} \int_{t_0}^{T} \max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \Bigg\|\sum_{l=0}^{D}\sum_{i=1}^{d}\big(&\partial_{x_l^i} a(\mathcal{T}(u, X_u^{(\ell)})) \\ &- \partial_{x_l^i} a(\mathcal{T}(u, X_u))\big)\,\mathrm{D}_s^j X_{u-\tau_l}^i\Bigg\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \mathrm{d}u = 0.\end{aligned}$$

Following the same arguments for the third term of $C^{(\ell)}$, we obtain

$$\begin{aligned}\lim_{\ell\to\infty} \int_{t_0}^{T} \sum_{k=1}^{m} \max_{j\in\{1,\dots,m\}} \sup_{s\in[t_0,T]} \Bigg\|\sum_{l=0}^{D}\sum_{i=1}^{d}\big(&\partial_{x_l^i} b^k(\mathcal{T}(u, X_u^{(\ell)})) \\ &- \partial_{x_l^i} b^k(\mathcal{T}(u, X_u))\big)\,\mathrm{D}_s^j X_{u-\tau_l}^i\Bigg\|^2_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)} \mathrm{d}u = 0\end{aligned}$$

as well, and thus, we have

$$\lim_{\ell\to\infty} C^{(\ell)} = 0 \tag{III.50}$$

in total. We are using this in order to show the convergence in equation (III.44) in the following. According to equation (III.50), for all $\tilde{\varepsilon} > 0$ there exist an $N(\tilde{\varepsilon}) \in \mathbb{N}_0$ such that $C^{(\ell)} < \tilde{\varepsilon}$ for all $\ell \geq N(\tilde{\varepsilon})$. Due to inequality (III.48) and since $\mathrm{D}_s^j X^{(0)} = 0$ for all $s \in [t_0, T]$ and $j \in \{1, \ldots, m\}$, we inductively obtain

$$\begin{aligned}
&\max_{j\in\{1,\ldots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(\ell+1)} - \mathrm{D}_s^j X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad < \tilde{\varepsilon} + C \int_{t_0}^{T} \max_{j\in\{1,\ldots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(\ell)} - \mathrm{D}_s^j X\|^2_{S^p([t_0-\tau,u]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}u \\
&\quad \leq \tilde{\varepsilon} \max_{j\in\{1,\ldots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \sum_{n=0}^{\ell+1} \frac{\big(C(T-t_0)\big)^n}{n!} \\
&\quad \leq \tilde{\varepsilon} \max_{j\in\{1,\ldots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \,\mathrm{e}^{C(T-t_0)}
\end{aligned} \tag{III.51}$$

for all $\ell \geq N(\tilde{\varepsilon})$. If

$$\max_{j\in\{1,\ldots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = 0,$$

the convergence in equation (III.44) is evident, and we assume

$$\max_{j\in\{1,\ldots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} > 0$$

in the following. Choosing

$$\tilde{\varepsilon} = \varepsilon \bigg(\max_{j\in\{1,\ldots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \bigg)^{-1} \mathrm{e}^{-C(T-t_0)}$$

for all $\varepsilon > 0$, it holds

$$\max_{j\in\{1,\ldots,m\}} \sup_{s\in[t_0,T]} \|\mathrm{D}_s^j X^{(\ell+1)} - \mathrm{D}_s^j X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} < \varepsilon$$

for all $\ell \geq N(\tilde{\varepsilon})$ according to inequality (III.51). That is, the convergence holds true in equation (III.44) and hence also in equations (III.41) and (III.42). Since $p \in [2, \infty[$ is arbitrary fixed in the considerations above, we have $X_t \in \mathscr{D}^p(\Omega; \mathbb{R}^d)$ for all $t \in [t_0 - \tau, T]$ and all $p \in [2, \infty[$ according to the closeness of operator D. Moreover, $\mathrm{D}X_t$ is in fact the Malliavin derivative of X_t, where the stochastic process $\mathrm{D}_s^j X$ satisfies the SDDE (III.21) for all $s \in [t_0, T]$ and all $j \in \{1, \ldots, m\}$. □

IV

Numerical Approximation of Stochastic Delay Differential Equations

Only since 1999 numerical solutions of SDDEs have been investigated. Yan considered the Euler-Maruyama and the Milstein scheme for SDDEs in his dissertation [137]. A little later, in 2000, further works were published [9, 17, 82].

In [9, 17], explicit one-step methods and their strong convergence are considered for SDDEs, where the Wiener process is one-dimensional. Since the increment functions of that explicit one-step methods only depend on increments of the Wiener process, see [17, Section 3], and on nondelayed-iterated stochastic integrals, see [9, Equation (21)], their results are not suitable for general approximations of higher order such like the Milstein scheme in [137]. Higher order methods are proposed in [82]. However, their convergence analysis has not been done thoroughly, cf. [137, pp. 40–41]. Some stochastic integrals appearing in [82, Equations (10.2) and (10.5)] are not well-defined in the sense of the Itô calculus, because their integrands are not adapted to the filtration generated by the integrator, the shifted Wiener process $(W_{t-\tau})_{t\in[t_0+\tau,T]}$.

In [137], Yan circumvents this problem using a tamed Itô formula for anticipating functionals and proved the convergence in $L^2(\Omega;\mathbb{R}^d)$ of the Milstein scheme. His result is also published in [60] together with Hu and Mohammed, where the SDDE's coefficients are allowed to be dependent on time additionally. Kloeden and Shardlow present a different proof in [80], compared to [60, 137], without using an anticipating calculus. Their proof exploits the differentiability of the SDDE's solution with respect to its initial condition. However, this is closely related to the Malliavin derivative of the solution, cf. [58] and [113, p. 126].

In this chapter, we prove in particular the convergence of the Milstein approximation for SDDEs in a stronger sense, namely in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ for all $p\in[1,\infty[$, and under milder conditions than in [60, 80, 137]. In addition, we show the pathwise convergence of the Milstein scheme for SDDEs. The types of convergences are defined below.

We remark that higher order schemes for SDDEs are also considered in [104, 124]. But these schemes are not optimal in the following the sense. The first-order scheme, for example, contains a term that is globally of order one as well. To show that the term is of order one, however, is more difficult as we will see in proof of Theorem IV.9 below.

In the following, we state some definitions on types of convergences and present results on the pathwise convergence.

Let $Y^h = Y$ denote the approximation based on a discretization $\{t_0, t_1, \ldots, t_N\}$ of the interval $[t_0, T]$ with maximum step size

$$h \coloneqq \max_{n\in\{0,1,\ldots,N-1\}} (t_{n+1} - t_n), \tag{IV.1}$$

where $t_0 < t_1 < \ldots < t_N \coloneqq T$.

Definition IV.1 (Strong Convergence)
A family of approximation processes $Y^h = (Y_t^h)_{t\in[t_0-\tau,T]} \in S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ for $h \in\,]0, T-t_0]$ converges in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ *to solution X of SDDE* (II.1) *for some $p \in [1,\infty[$ if*

$$\lim_{h\to 0} \|X - Y^h\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = 0.$$

The family $(Y^h)_{h\in]0,T-t_0]}$ is further said to converge in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ with order $\alpha \in\,]0,\infty[$ *to solution X if there exists a constant $C > 0$, independent of h, and an $h^\star \in\,]0, T-t_0]$ such that*

$$\|X - Y^h\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \le C h^\alpha \tag{IV.2}$$

for all $h \in I$ with $h < h^\star$.

Often, the convergence is considered in $L^2(\Omega;\mathbb{R}^d)$ only, cf. [60, 137]. Analogously to the previous definition, the family of approximation processes converges in $L^p(\Omega;\mathbb{R}^d)$ if

$$\lim_{h\to 0} \sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^h\|_{L^p(\Omega;\mathbb{R}^d)} = 0,$$

and the convergence is of order $\alpha \in\,]0,\infty[$ if

$$\sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^h\|_{L^p(\Omega;\mathbb{R}^d)} \le C h^\alpha.$$

Note that the convergence in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ implies the convergence in $L^q(\Omega;\mathbb{R}^d) \subseteq L^p(\Omega;\mathbb{R}^d)$ for every $p, q \in [1,\infty[$ with $q \le p$.

If a family of approximation processes converges in $L^p(\Omega;\mathbb{R}^d)$ or in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ for all $p \in [1,\infty[$, we can even draw conclusions about its almost sure convergence as we will see below. The almost sure convergence of numerical solutions of SDEs is also called pathwise convergence in the literature, cf. [2, 41, 51, 77, 130].

Definition IV.2 (Pathwise Convergence)
A family of approximation processes $Y^h = (Y_t^h)_{t\in[t_0-\tau,T]} \in S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ for $h \in I \subseteq\,]0, T-t_0]$ converges pathwise *to solution X of SDDE* (II.1) *if*

$$\sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^h\| \to 0$$

converges P*-almost surely as* $h \to 0$*. The family* $(Y^h)_{h \in]0,T-t_0]}$ *is further said to* converge pathwise with order $\alpha \in]0, \infty[$ *to solution* X *if there exists a positive random variable* $Z\colon \Omega \to \mathbb{R}$*, independent of* h*, and an* $h^\star \in I$ *such that*

$$\sup_{t \in [t_0-\tau,T]} \|X_t - Y_t^h\| \leq Z h^\alpha$$

P*-almost surely for all* $h \in I$ *with* $h < h^\star$*.*

Faure stated in [41, Proposition 23] that if

$$\sup_{n \in \{0,1,\ldots,N\}} \|X_{t_n} - Y_{t_n}^{h_N}\|_{L^p(\Omega;\mathbb{R}^d)} \leq C_p h_N^\alpha \tag{IV.3}$$

for all $p \in [1, \infty[$ and $N \in \mathbb{N}$ where $h_N = \frac{T-t_0}{N}$, then we also have that

$$N^{\alpha-\varepsilon} \sup_{n \in \{0,1,\ldots,N\}} \|X_{t_n} - Y_{t_n}^{h_N}\| \to 0 \tag{IV.4}$$

converges P-almost surely as $N \to \infty$ for all $\varepsilon > 0$. He used this result in order to show that the Euler-Maruyama scheme and the Milstein scheme for SODEs satisfy the convergence in equation (IV.4) with $\alpha = \frac{1}{2}$ and $\alpha = 1$, respectively, see [41, Proposition 14] and [41, Proposition 21 and 25].

Later, Gyöngy proved the pathwise convergence of order $\alpha = \frac{1}{2} - \varepsilon$ for the Euler-Maruyama scheme of SODEs whose coefficients satisfy a local Lipschitz condition [51, Theorem 2.4]. Kloeden and Neuenkirch used an idea of Gyöngy's proof in order to show for a sequence $(\Upsilon_N)_{N \in \mathbb{N}}$ of random variables $\Upsilon_N\colon \Omega \to \mathbb{R}$ with

$$\|\Upsilon_N\|_{L^p(\Omega;\mathbb{R})} \leq C_p N^{-\alpha}$$

for all $p \in [1, \infty[$ and $N \in \mathbb{N}$ that, for all $\varepsilon > 0$, there exists a positive random variable Z_ε with $\|Z_\varepsilon\|_{L^p(\Omega;\mathbb{R})} < \infty$ for all $p \in [1, \infty[$ such that

$$\|\Upsilon_N\| \leq Z_\varepsilon N^{-(\alpha-\varepsilon)}$$

P-almost surely for all $N \in \mathbb{N}$, see [77, Lemma 2.1]. Thus, [77, Lemma 2.1] especially implies that if

$$\|X - Y^{h_N}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq C_p h_N^\alpha$$

for all $p \in [1, \infty[$ and $N \in \mathbb{N}$ where $h_N = \frac{T-t_0}{N}$, then

$$\sup_{t \in [t_0-\tau,T]} \|X_t - Y_t^{h_N}\| \leq Z_\varepsilon h_N^{\alpha-\varepsilon} \tag{IV.5}$$

P-almost surely for all $N \in \mathbb{N}$, and

$$N^{\alpha-\varepsilon} \sup_{t \in [t_0-\tau,T]} \|X_t - Y_t^{h_N}\| \to 0$$

converges P-almost surely as $N \to \infty$ for all $\varepsilon > 0$. However, if condition (IV.3) only holds true instead of condition (IV.5), we merely obtain by [77, Lemma 2.1] that

$$N^{\alpha-\varepsilon} \|X_{t_n} - Y_{t_n}^{h_N}\| \to 0$$

converges P-almost surely as $N \to \infty$ for all $\varepsilon > 0$ and $n \in \{0, 1, \dots, N\}$, which is weaker than the convergence in formula (IV.4). Thus, the result on the convergence of Faure is a little bit stronger than the one of Kloeden and Neuenkirch. However, the result of Kloeden and Neuenkirch provides the existence of a positive random variable Z_ε with $\|Z_\varepsilon\|_{L^p(\Omega;\mathbb{R})} < \infty$ for all $p \in [1, \infty[$. Our lemma below combines these both results from [41, Proposition 23] and [77, Lemma 2.1].

Lemma IV.3

Let $\alpha \in]0, \infty[$ and X be the solution of SDDE (II.1). *Consider a family of approximation processes $Y^{h_N} = (Y_t^{h_N})_{t\in[t_0-\tau,T]} \in S^p([t_0 - \tau, T] \times \Omega; \mathbb{R}^d)$ for $(h_N)_{N\in\mathbb{N}} \subset]0, T - t_0]$. For $N \in \mathbb{N}$, let $\{t_n^N : n \in \{0, 1, \dots, N\}\}$ be the discretization of $[t_0, T]$ with $t_0 =: t_0^N < t_1^N < \dots < t_N^N := T$ that corresponds to the maximum step size h_N. Let $q_\varepsilon \in [1, \infty[$ for all $\varepsilon > 0$ be independent of N and such that*

$$\sum_{N=1}^{\infty} h_N^{\varepsilon q_\varepsilon} < \infty. \tag{IV.6}$$

Further, let

$$\sup_{n\in\{0,1,\dots,N-1\}} \|X - Y^{h_N}\|_{S^p(([t_0-\tau,t_0]\cup]t_n^N,t_{n+1}^N])\times\Omega;\mathbb{R}^d)} \le C_p h_N^\alpha \tag{IV.7}$$

for all $p \in [1, \infty[$ and all $N \in \mathbb{N}$, where $C_p > 0$ is a constant independent of h_N.

Then,

$$h_N^{-(\alpha-\varepsilon)} \sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\| \to 0 \tag{IV.8}$$

converges P*-almost surely as $N \to \infty$ for all $\varepsilon > 0$, and for all $\varepsilon > 0$, there exists a positive random variable Z_ε with $\|Z_\varepsilon\|_{L^p(\Omega;\mathbb{R})} < \infty$ for all $p \in [1, \infty[$ such that*

$$\sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\| \le Z_\varepsilon h_N^{\alpha-\varepsilon} \tag{IV.9}$$

P*-almost surely for all $N \in \mathbb{N}$.*

Proof. The proof is stated in Section IV.3, see p. 79. □

The proof of Lemma IV.3 is based on the Borel-Cantelli Lemma. In order to apply the Borel-Cantelli Lemma, the condition (IV.6) is needed, cf. also [2, Lemma 3.2]. In [41] and [77], the results are only presented for equidistant discretization, that is in case of $h_N = \frac{T-t_0}{N}$. Our lemma, however, also holds true for more general discretizations, for example when $h_N = \frac{T-t_0}{\sqrt{N}}$ for $N \in \mathbb{N}$. Then, the condition (IV.6) holds true for all $q_\varepsilon > \frac{2}{\varepsilon}$. However, we do not obtain the pathwise convergence for all sequences $(h_N)_{N\in\mathbb{N}}$ that converge to zero. Consider for example the sequence with $h_N = \frac{T-t_0}{\log(N+1)}$. Then, there exists no $q_\varepsilon \in [1, \infty[$ that is independent of N and such that the condition (IV.6) holds true. Hence, the pathwise convergence cannot be obtained for all sequences $(h_N)_{N\in\mathbb{N}}$ that converge to zero as $N \to \infty$ by applying Lemma IV.3. In contrast to this, the strong convergence directly follows from inequality (IV.2) for all sequences $(h_N)_{N\in\mathbb{N}}$ that converge to zero as $N \to \infty$.

Condition (IV.7) in Lemma IV.3 above seems to be quite technical at first. Let us give two remarks on this condition.

Remark IV.4
The condition

$$\|X - Y^{h_N}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq C_p h_N^\alpha \tag{IV.10}$$

for all $p \in [1, \infty[$ *and all* $N \in \mathbb{N}$*, where* $C_p > 0$ *is a constant independent of* h_N*, clearly implies condition* (IV.7)*.*

Remark IV.5
For all $t \in [t_0 - \tau, t_0]$ *and* $N \in \mathbb{N}$*, let* $Y_t^{h_N} = X_t$ P*-almost surely for simplicity. Further, let*

$$\sup_{n\in\{1,\dots,N\}} \|X_{t_n^N} - Y_{t_n^N}^{h_N}\|_{L^p(\Omega;\mathbb{R}^d)} \leq C_p h_N^\alpha \tag{IV.11}$$

for all $p \in [1, \infty[$ *and all* $N \in \mathbb{N}$*, where* $C_p > 0$ *is a constant independent of* h_N*. Then, in order to show that condition* (IV.7) *is fulfilled, we only need to show additionally that the local errors satisfy*

$$\left\| \sup_{t\in]t_n^N, t_{n+1}^N[} \|X_t - Y_t^{h_N}\| \right\|_{L^p(\Omega;\mathbb{R})} \leq C_p h_N^\alpha \tag{IV.12}$$

for all $p \in [1, \infty[$*,* $n \in \{1, \dots, N\}$*, and* $N \in \mathbb{N}$*.*

This is especially an advantage for the proof of the pathwise convergence of higher order numerical methods for SDDEs. Here, it is much more complicated to show condition (IV.10) *than conditions* (IV.11) *and* (IV.12) *to be fulfilled. For details on this in the case of the Milstein scheme, we refer to the estimate of the term* $\mathcal{R}_5$ *in the proof of Theorem IV.9, and here, see inequality* (IV.146) *and Lemma IV.22 in particular.*

Before we begin with the analysis of the convergence of the Milstein scheme, we recall the convergence of the Euler-Maruyama scheme in the next section. Then, we can account for problems in proving the convergence of the Milstein scheme in Section IV.2.

IV.1. The Euler-Maruyama Approximation

Let $\{t_0, t_1, \dots, t_N\}$, $N \in \mathbb{N}$, be a discretization of $[t_0, T]$ where $t_0 < t_1 < \dots < t_N := T$. The Euler-Maruyama approximation Y with respect to discretization $\{t_0, t_1, \dots, t_N\}$ and SDDE (II.1) is defined by

$$Y_t = \begin{cases} \xi_t \quad \text{if } t \in [t_0 - \tau, t_0] \text{ and} \\ Y_{t_n} + a(t_n, t_n - \tau_1, \dots, t_n - \tau_D, Y_{t_n}, Y_{t_n-\tau_1}, \dots, Y_{t_n-\tau_D})(t - t_n) \\ \qquad + \sum_{j=1}^m b^j(t_n, t_n - \tau_1, \dots, t_n - \tau_D, Y_{t_n}, Y_{t_n-\tau_1}, \dots, Y_{t_n-\tau_D})(W_t^j - W_{t_n}^j) \\ \quad \text{if } t \in]t_n, t_{n+1}] \text{ where } n = 0, 1, \dots, N-1. \end{cases} \tag{IV.13}$$

Its convergence is analyzed in, among others, [2, 9, 17, 77, 82, 98, 104, 124, 137]. In this regard, the results presented in this section are not new and only serve as an introduction to the approximation of solutions of SDDEs.

In order to keep the formulas clear, we introduce some notations. At first, recall the shift operator

$$\mathcal{T}(t, Y_t) = (t, t - \tau_1, \ldots, t - \tau_D, Y_t, Y_{t-\tau_1}, \ldots, Y_{t-\tau_D})$$

defined in formula (II.3). Then, the Euler-Maruyama scheme (IV.13) can be represented as

$$Y_t = \begin{cases} \xi_t \quad \text{if } t \in [t_0 - \tau, t_0] \text{ and} \\ Y_{t_n} + a(\mathcal{T}(t_n, Y_{t_n}))(t - t_n) + \sum_{j=1}^{m} b^j(\mathcal{T}(t_n, Y_{t_n}))(W_t^j - W_{t_n}^j) \\ \quad \text{if } t \in \,]t_n, t_{n+1}] \text{ where } n = 0, 1, \ldots, N-1. \end{cases}$$

Further, define the projections $\lfloor\cdot\rfloor, \lceil\cdot\rceil \colon [t_0, T] \to \{t_0, t_1, \ldots, t_N\}$ by

$$\lfloor s \rfloor := \sum_{n=0}^{N-1} t_n \mathbb{1}_{[t_n, t_{n+1}[}(s) + t_N \mathbb{1}_{t_N}(s) \tag{IV.14}$$

and

$$\lceil s \rceil := t_0 \mathbb{1}_{t_0}(s) + \sum_{n=0}^{N-1} t_{n+1} \mathbb{1}_{]t_n, t_{n+1}]}(s), \tag{IV.15}$$

respectively, for all $s \in [t_0, T]$. Thus, we have $\lfloor s \rfloor = t_n$ for $s \in [t_n, t_{n+1}[$ and $\lceil s \rceil = t_{n+1}$ for $s \in \,]t_n, t_{n+1}]$, where $n \in \{0, 1, \ldots, N-1\}$.

Taking advantage of these notations and of the measurability of the coefficients in the Euler-Maruyama approximation, we can rewrite scheme (IV.13) to

$$Y_t = \begin{cases} \xi_t \quad \text{if } t \in [t_0 - \tau, t_0] \text{ and} \\ \xi_{t_0} + \int_{t_0}^{t} a(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor}))\,\mathrm{d}s + \sum_{j=1}^{m} \int_{t_0}^{t} b^j(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor}))\,\mathrm{d}W_s^j \quad \text{if } t \in \,]t_0, T]. \end{cases}$$

Using this notations, we now state and prove the theorem on the convergence of the Euler-Maruyama scheme. Let us note that the convergence analysis can also be done under weaker assumptions regarding the coefficients of the SDDE than presented below, cf. [52, 83, 99, 100].

Theorem IV.6 (Strong Convergence of the Euler-Maruyama Approximation)
Let the Borel-measurable drift a and diffusion b^j, $j \in \{1, \ldots, m\}$, of SDDE (II.1) *satisfy the global Lipschitz and linear growth conditions* (II.8), (II.9), (II.10), *and* (II.11) *as well as, for some growth exponents $\gamma_a, \gamma_b \in [0, \infty[$ and some constants $L_{\mathrm{t},a}, L_{\mathrm{t},b} > 0$, the conditions*

$$\begin{aligned} &\|a(t, t - \tau_1, \ldots, t - \tau_D, x_0, x_1, \ldots, x_D) \\ &\qquad - a(s, s - \tau_1, \ldots, s - \tau_D, x_0, x_1, \ldots, x_D)\| \\ &\quad \leq L_{\mathrm{t},a} \max_{l \in \{0,1,\ldots,D\}} \left(1 + \|x_l\|^2\right)^{\frac{\gamma_a}{2}} \cdot \sqrt{|t - s|} \end{aligned} \tag{IV.16}$$

and

$$\begin{aligned}\max_{j\in\{1,\ldots,m\}} &\|b^j(t,t-\tau_1,\ldots,t-\tau_D,x_0,x_1,\ldots,x_D)\\ &\quad - b^j(s,s-\tau_1,\ldots,s-\tau_D,x_0,x_1,\ldots,x_D)\|\\ &\le L_{t,b}\max_{l\in\{0,1,\ldots,D\}}\left(1+\|x_l\|^2\right)^{\frac{\gamma_b}{2}}\cdot\sqrt{|t-s|}\end{aligned} \tag{IV.17}$$

for all $s,t\in[t_0,T]$ *and* $x_l\in\mathbb{R}^d$, $l\in\{0,1,\ldots,D\}$. *Consider Euler-Maruyama approximation* (IV.13) *regarding SDDE* (II.1) *with initial condition* $\xi\in S^{(\gamma_a\vee\gamma_b\vee 1)p}([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)$ *for some* $p\in[2,\infty[$. *Let* ξ *fulfill for some constant* $L_\xi>0$ *the condition*

$$\|\xi_t-\xi_s\|_{L^p(\Omega;\mathbb{R}^d)}\le L_\xi\sqrt{|t-s|} \tag{IV.18}$$

for all $s,t\in[t_0-\tau,t_0]$.

Then, the family of Euler-Maruyama approximations $(Y^h)_{h\in]0,T-t_0]}$ *converges in* $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ *with the order* $\alpha=\frac{1}{2}$ *to solution* X *of SDDE* (II.1) *as* $h\to 0$. *That is, there exists a constant* $C_{\text{Euler}}>0$, *independent of* h, *such that*

$$\|X-Y^h\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\le C_{\text{Euler}}\sqrt{h}$$

for all $h\in]0,T-t_0]$.

Proof. For sake of simplicity, we fix an $h\in]0,T-t_0]$ and set $Y=Y^h$. We P-almost surely have

$$X_t-Y_t=\begin{cases}0 \quad \text{if } t\in[t_0-\tau,t_0] \text{ and}\\ \displaystyle\int_{t_0}^t a(\mathcal{T}(s,X_s))-a(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\,\mathrm{d}s\\ \displaystyle\quad+\sum_{j=1}^m\int_{t_0}^t b^j(\mathcal{T}(s,X_s))-b^j(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\,\mathrm{d}W_s^j \quad \text{if } t\in]t_0,T],\end{cases} \tag{IV.19}$$

for all $t\in[t_0-\tau,T]$ and use the expansions

$$\begin{aligned}&a(\mathcal{T}(s,X_s))-a(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\\ &\quad=a(\mathcal{T}(s,X_s))-a(\mathcal{T}(\lfloor s\rfloor,X_s))+a(\mathcal{T}(\lfloor s\rfloor,X_s))-a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}))\\ &\qquad+a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}))-a(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\end{aligned} \tag{IV.20}$$

and

$$\begin{aligned}&b^j(\mathcal{T}(s,X_s))-b^j(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\\ &\quad=b^j(\mathcal{T}(s,X_s))-b^j(\mathcal{T}(\lfloor s\rfloor,X_s))+b^j(\mathcal{T}(\lfloor s\rfloor,X_s))-b^j(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}))\\ &\qquad+b^j(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}))-b^j(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\end{aligned} \tag{IV.21}$$

for all $s\in[t_0,T]$. Then, we prove the convergence of order $\alpha=\frac{1}{2}$ using the triangle inequality, Zakai's inequality from Theorem II.6, and the Gronwall's Lemma II.7.

The order $\alpha = \frac{1}{2}$ results from the following estimates. Using the triangle inequality and the Lipschitz continuity (II.8) of the drift coefficient, it holds

$$\begin{aligned}
&\left\| \int_{t_0}^{\cdot} a(\mathcal{T}(\lfloor s\rfloor, X_s)) - a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor})) \,\mathrm{d}s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad\leq \left\| \sup_{t\in[t_0,T]} \int_{t_0}^{t} \| a(\mathcal{T}(\lfloor s\rfloor, X_s)) - a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}))\| \,\mathrm{d}s \right\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \int_{t_0}^{T} \| a(\mathcal{T}(\lfloor s\rfloor, X_s)) - a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}))\|_{L^p(\Omega;\mathbb{R}^d)} \,\mathrm{d}s && \text{(IV.22)} \\
&\quad\leq L_a \int_{t_0}^{T} \left\| \sup_{l\in\{0,1,\ldots,D\}} \| X_{s-\tau_l} - X_{\lfloor s\rfloor-\tau_l} \| \right\|_{L^p(\Omega;\mathbb{R})} \mathrm{d}s && \text{(IV.23)} \\
&\quad\leq L_a \int_{t_0}^{T} \sum_{l=0}^{D} \| X_{s-\tau_l} - X_{\lfloor s\rfloor-\tau_l} \|_{L^p(\Omega;\mathbb{R}^d)} \,\mathrm{d}s. && \text{(IV.24)}
\end{aligned}$$

To take initial condition ξ on the interval $[t_0 - \tau, t_0]$ into account, we write

$$X_{s-\tau_l} - X_{\lfloor s\rfloor-\tau_l} = \xi_{(s-\tau_l)\wedge t_0} - \xi_{(\lfloor s\rfloor-\tau_l)\wedge t_0} + X_{(s-\tau_l)\vee t_0} - X_{(\lfloor s\rfloor-\tau_l)\vee t_0} \tag{IV.25}$$

for all $s \in [t_0, T]$ and $l \in \{0, 1, \ldots, D\}$. Further, for all $s \in [t_0, T]$, we estimate

$$\begin{aligned}
&\sum_{l=0}^{D} \| X_{s-\tau_l} - X_{\lfloor s\rfloor-\tau_l} \|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad\leq \sum_{l=1}^{D} \| \xi_{(s-\tau_l)\wedge t_0} - \xi_{(\lfloor s\rfloor-\tau_l)\wedge t_0} \|_{L^p(\Omega;\mathbb{R}^d)} + \sum_{l=0}^{D} \| X_{(s-\tau_l)\vee t_0} - X_{(\lfloor s\rfloor-\tau_l)\vee t_0} \|_{L^p(\Omega;\mathbb{R}^d)},
\end{aligned} \tag{IV.26}$$

where

$$\xi_{(s-\tau_0)\wedge t_0} - \xi_{(\lfloor s\rfloor-\tau_0)\wedge t_0} = \xi_{s\wedge t_0} - \xi_{\lfloor s\rfloor\wedge t_0} = \xi_{t_0} - \xi_{t_0} = 0$$

for all $s \in [t_0, T]$ is used. According to condition (IV.18) and Lemma II.9, we have

$$\begin{aligned}
&\sum_{l=1}^{D} \| \xi_{(s-\tau_l)\wedge t_0} - \xi_{(\lfloor s\rfloor-\tau_l)\wedge t_0} \|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad\leq L_\xi \sum_{l=1}^{D} \sqrt{((s-\tau_l)\wedge t_0) - (\lfloor s\rfloor - \tau_l)\wedge t_0)} \\
&\quad\leq L_\xi D \sqrt{s - \lfloor s\rfloor} && \text{(IV.27)}
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{l=0}^{D} \| X_{(s-\tau_l)\vee t_0} - X_{(\lfloor s\rfloor-\tau_l)\vee t_0} \|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad\leq C_1 \sum_{l=0}^{D} \sqrt{((s-\tau_l)\wedge t_0) - (\lfloor s\rfloor - \tau_l)\wedge t_0)} \\
&\quad\leq C_1 (D+1) \sqrt{s - \lfloor s\rfloor} && \text{(IV.28)}
\end{aligned}$$

for all $s \in [t_0, T]$, where $C_1 > 0$ is a constant, cf. Lemma II.9. It thus follows

$$\sum_{l=0}^{D} \|X_{s-\tau_l} - X_{\lfloor s\rfloor-\tau_l}\|_{L^p(\Omega;\mathbb{R}^d)} \leq \big(L_\xi D + C_1(D+1)\big)\sqrt{s-\lfloor s\rfloor} \tag{IV.29}$$

for all $s \in [t_0, T]$. Inserting this into inequality (IV.24) results in

$$\begin{aligned} &\Big\| \int_{t_0}^{\cdot} a(\mathcal{T}(\lfloor s\rfloor, X_s)) - a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}))\,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\ &\quad\leq L_a\big(L_\xi D + C_1(D+1)\big) \int_{t_0}^{T} \sqrt{s-\lfloor s\rfloor}\,\mathrm{d}s. \end{aligned} \tag{IV.30}$$

Consider the integral over time in the inequality above. It holds in view of equations (IV.14) and (IV.1) that

$$\begin{aligned} \int_{t_0}^{T} \sqrt{s-\lfloor s\rfloor}\,\mathrm{d}s &= \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \sqrt{s-t_n}\,\mathrm{d}s = \sum_{n=0}^{N-1} \frac{2}{3}(t_{n+1}-t_n)^{\frac{3}{2}} \\ &\leq \frac{2}{3}\sqrt{h} \sum_{n=0}^{N-1}(t_{n+1}-t_n) = \frac{2}{3}(T-t_0)\sqrt{h}. \end{aligned} \tag{IV.31}$$

Thus, with inequality (IV.30), we obtain the estimate

$$\begin{aligned} &\Big\| \int_{t_0}^{\cdot} a(\mathcal{T}(\lfloor s\rfloor, X_s)) - a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}))\,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\ &\quad\leq L_a\big(L_\xi D + C_1(D+1)\big) \frac{2}{3}(T-t_0)\sqrt{h}. \end{aligned} \tag{IV.32}$$

A similar estimate holds for the stochastic integrals. For more details on the proof and the constants appearing in the estimates, we refer to Section IV.3, see p. 81. □

Based on [77, Lemma 2.1], Kloeden and Neuenkirch showed that the Euler-Maruyama scheme for SDDEs converges pathwise with order $\alpha = \frac{1}{2} - \varepsilon$, see [77, Theorem 2.5]. In [2] and [52], the pathwise convergence is also proved under weaker assumptions. Nevertheless, we state the pathwise convergence of the Euler-Maruyama scheme as a corollary of our theorem above and Lemma IV.3.

Corollary IV.7 (Pathwise Convergence of the Euler-Maruyama Approximation)
Let $\xi \in S^p([t_0-\tau, t_0] \times \Omega; \mathbb{R}^d)$ and the additional assumptions in Theorem IV.6 regarding SDDE (II.1) *be fulfilled for all $p \in [1,\infty[$. Consider the family of Euler-Maruyama approximations $(Y^{h_N})_{N\in\mathbb{N}}$, where $(h_N)_{N\in\mathbb{N}} \subset]0, T-t_0]$. Let $q_\varepsilon \in [1,\infty[$ for all $\varepsilon > 0$ be independent of N and such that $\sum_{N=1}^{\infty} h_N^{\varepsilon q_\varepsilon} < \infty$.*

Then, the family of Euler-Maruyama approximations $(Y^{h_N})_{N\in\mathbb{N}}$ converges pathwise with order $\alpha = \frac{1}{2} - \varepsilon$ to solution X of SDDE (II.1) *for arbitrary $\varepsilon > 0$ as $N \to \infty$. That is, for all $\varepsilon > 0$, there exists a positive random variable Z_ε, which belongs to $L^p(\Omega;\mathbb{R})$ for all $p \in [1,\infty[$, such that*

$$\sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\| \leq Z_\varepsilon\, h_N^{\frac{1}{2}-\varepsilon}$$

P-*almost surely for all $N \in \mathbb{N}$.*

IV.2. The Milstein Approximation

As we showed in the previous section, the Euler-Maruyama scheme for SDDE (II.1) converges with order $\alpha = \frac{1}{2}$ in $S^p([t_0-\tau, T]\times\Omega;\mathbb{R}^d)$. However, it is well-known that the Euler approximation converges with order $\alpha = 1$ in case of deterministic delay differential equations. In order to increase the order of convergence, the approximation needs to incorporate more information about the diffusion in general, cf. [78, 105], where SODEs are considered. The simplest first order scheme for SODEs originates from Milstein [106] and is called Milstein scheme to his honor. We introduce and study the strong convergence of the Milstein scheme for SDDE (II.1), cf. [60, 80, 137].

Whereas the Euler-Maruyama scheme for SDDEs is consistent with Euler-Maruyama scheme for SODEs in the number of incorporated terms, the Milstein scheme for SDDEs differs from its variant for SODEs if a diffusion coefficient b^j of SDDE (II.1) depends on the past history of the solution, that is in the presence of delay.

Using the notations introduced in the previous section, the Milstein approximation Y for strong solution X of SDDE (II.1) is defined by

$$Y_t = \begin{cases} \xi_t \quad \text{if } t \in [t_0-\tau, t_0] \text{ and} \\ \xi_{t_0} + \int_{t_0}^t a(\mathcal{T}(\lfloor s\rfloor, Y_{\lfloor s\rfloor}))\,\mathrm{d}s + \sum_{j=1}^m \int_{t_0}^t b^j(\mathcal{T}(\lfloor s\rfloor, Y_{\lfloor s\rfloor}))\,\mathrm{d}W_s^j \\ \quad + \sum_{l=0}^D \sum_{j_1=1}^m \int_{t_0}^t \Bigg(\sum_{i=1}^d \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s\rfloor, Y_{\lfloor s\rfloor})) \\ \quad \times \sum_{j_2=1}^m \int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j_2}(\mathcal{T}((\lfloor s\rfloor-\tau_l)\vee t_0, Y_{(\lfloor s\rfloor-\tau_l)\vee t_0}))\,\mathrm{d}W_u^{j_2} \Bigg)\,\mathrm{d}W_s^{j_1} \\ \quad \text{if } t \in \,]t_0, T]. \end{cases} \tag{IV.33}$$

The derivatives of the diffusion coefficients with respect to delay arguments, that are the summands for $l \in \{1,\ldots,D\}$, vanish if the diffusion does not depend on the past history of the solution. In this case, the scheme simplifies and is consistent with the Milstein scheme for SODEs, cf. [78, 105].

Its mean-square convergence ($p = 2$) of Milstein scheme (IV.33) has been analyzed in [60, 80, 137] under rather strong assumptions. Namely, in [137, Theorem 9.2] and [60, Theorem 5.2], the authors assume that the SDDEs' coefficients have bounded first and second spatial derivatives whereas in [80] the coefficients are not time-dependent, and the third derivatives are even assumed to be bounded in addition, cf. [80, Assumptions 3.1 and 7.1]. In this thesis, we show the strong convergence of the Milstein under weaker assumptions and make the following assumption on SDDE (II.1) for our analysis.

Assumption IV.8
Let the coefficients $a, b^j\colon \mathbb{R}^{1\times(D+1)}\times\mathbb{R}^{d\times(D+1)} \to \mathbb{R}^d$, $j \in \{1,\ldots,m\}$, *and initial condition* $\xi\colon [t_0-\tau, t_0]\times\Omega \to \mathbb{R}^d$ *of SDDE* (II.1) *fulfill the following.*

i) Drift coefficient a *and diffusion coefficient* b^j, $j \in \{1,\ldots,m\}$, *are Borel-measurable, and for all* $t \in [t_0, T]$, $a(t, t-\tau_1, \ldots, t-\tau_D, \cdot, \ldots, \cdot)$ *and* $b^j(t, t-\tau_1, \ldots, t-\tau_D, \cdot, \ldots, \cdot)\colon \mathbb{R}^{d\times(D+1)}$

$\to \mathbb{R}^d$, $j \in \{1,\dots,m\}$, belong to $C^2(\mathbb{R}^{d\times(D+1)};\mathbb{R}^d)$, that is, they are continuous and have continuous first and second partial derivatives.

ii) The global Lipschitz conditions (II.8) *and* (II.9) *hold. That is, there exist constants $L_a, L_b > 0$ such that*

$$\sup_{t\in[t_0,T]} \|a(t,t-\tau_1,\dots,t-\tau_D,x_0,x_1,\dots,x_D) - a(t,t-\tau_1,\dots,t-\tau_D,y_0,y_1,\dots,y_D)\| \\ \le L_a \max_{l\in\{0,1,\dots,D\}} \|x_l - y_l\|$$

and

$$\sup_{t\in[t_0,T]} \max_{j\in\{1,\dots,m\}} \|b^j(t,t-\tau_1,\dots,t-\tau_D,x_0,x_1,\dots,x_D) \\ - b^j(t,t-\tau_1,\dots,t-\tau_D,y_0,y_1,\dots,y_D)\| \\ \le L_b \max_{l\in\{0,1,\dots,D\}} \|x_l - y_l\|$$

for all $x_l, y_l \in \mathbb{R}^d$, where $l \in \{0,1,\dots,D\}$.

iii) There exist a constant $L_{\partial b} > 0$ and a growth exponent $\beta \in [0,\infty[$ such that the Lipschitz condition

$$\sup_{t\in[t_0,T]} \max_{\substack{j_1,j_2\in\{1,\dots,m\}\\ l\in\{0,1,\dots,D\}}} \Bigg\| \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(t,t-\tau_1,\dots,t-\tau_D,x_0,x_1,\dots,x_D) \\ \times b^{i,j_2}((t-\tau_0)\vee t_0-\tau_l,\dots,(t-\tau_D)\vee t_0-\tau_l,\tilde{x}_0,\tilde{x}_1,\dots,\tilde{x}_D) \\ - \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(t,t-\tau_1,\dots,t-\tau_D,y_0,y_1,\dots,y_D) \\ \times b^{i,j_2}((t-\tau_0)\vee t_0-\tau_l,\dots,(t-\tau_D)\vee t_0-\tau_l,\tilde{y}_0,\tilde{y}_1,\dots,\tilde{y}_D)\Bigg\| \\ \le L_{\partial b} \max_{l\in\{0,1,\dots,D\}} \left(1 + (\|x_l\|\vee\|\tilde{x}_l\|)^2 + (\|y_l\|\vee\|\tilde{y}_l\|)^2\right)^{\frac{\beta}{2}} \\ \times \max_{l\in\{0,1,\dots,D\}} \left(\|x_l-y_l\| \vee \|\tilde{x}_l - \tilde{y}_l\|\right)$$

holds for all $x_l, \tilde{x}_l, y_l, \tilde{y}_l \in \mathbb{R}^d$, where $l \in \{0,1,\dots,D\}$.

iv) The linear growth conditions (II.10) *and* (II.11) *hold. That is, there exist constants $K_a, K_b > 0$ such that*

$$\sup_{t\in[t_0,T]} \|a(t,t-\tau_1,\dots,t-\tau_D,x_0,x_1,\dots,x_D)\| \le K_a \max_{l\in\{0,1,\dots,D\}} \left(1+\|x_l\|^2\right)^{\frac{1}{2}}$$

and

$$\sup_{t\in[t_0,T]} \max_{j\in\{1,\dots,m\}} \|b^j(t,t-\tau_1,\dots,t-\tau_D,x_0,x_1,\dots,x_D)\| \le K_b \max_{l\in\{0,1,\dots,D\}} \left(1+\|x_l\|^2\right)^{\frac{1}{2}}$$

for all $x_l \in \mathbb{R}^d$, where $l \in \{0,1,\dots,D\}$.

v) *There exist constants* $K_{\partial^2 a}, K_{\partial^2 b} > 0$ *and growth exponents* $\varrho_a, \varrho_b \in [0, \infty[$ *such that the growth conditions*

$$\sup_{t \in [t_0,T]} \max_{\substack{i_1,i_2 \in \{1,\dots,d\} \\ l_1,l_2 \in \{0,1,\dots,D\}}} \|\partial_{x_{l_1}^{i_1}} \partial_{x_{l_2}^{i_2}} a(t, t-\tau_1, \dots, t-\tau_D, x_0, x_1, \dots, x_D)\| \\ \leq K_{\partial^2 a} \max_{l \in \{0,1,\dots,D\}} \left(1 + \|x_l\|^2\right)^{\frac{\varrho_a}{2}}$$

and

$$\sup_{t \in [t_0,T]} \max_{\substack{j \in \{1,\dots,m\} \\ i_1,i_2 \in \{1,\dots,d\} \\ l_1,l_2 \in \{0,1,\dots,D\}}} \|\partial_{x_{l_1}^{i_1}} \partial_{x_{l_2}^{i_2}} b^j(t, t-\tau_1, \dots, t-\tau_D, x_0, x_1, \dots, x_D)\| \\ \leq K_{\partial^2 b} \max_{l \in \{0,1,\dots,D\}} \left(1 + \|x_l\|^2\right)^{\frac{\varrho_b}{2}}$$

hold for all $x_l \in \mathbb{R}^d$, *where* $l \in \{0, 1, \dots, D\}$.

vi) *There exist constants* $L_{\mathrm{t},a}, L_{\mathrm{t},b} > 0$ *and growth exponents* $\gamma_a, \gamma_b \in [0, \infty[$ *such that the Lipschitz conditions in time*

$$\|a(t, t-\tau_1, \dots, t-\tau_D, x_0, x_1, \dots, x_D) - a(s, s-\tau_1, \dots, s-\tau_D, x_0, x_1, \dots, x_D)\| \\ \leq L_{\mathrm{t},a} \max_{l \in \{0,1,\dots,D\}} \left(1 + \|x_l\|^2\right)^{\frac{\gamma_a}{2}} |t-s|$$

and

$$\max_{j \in \{1,\dots,m\}} \|b^j(t, t-\tau_1, \dots, t-\tau_D, x_0, x_1, \dots, x_D) \\ - b^j(s, s-\tau_1, \dots, s-\tau_D, x_0, x_1, \dots, x_D)\| \\ \leq L_{\mathrm{t},b} \max_{l \in \{0,1,\dots,D\}} \left(1 + \|x_l\|^2\right)^{\frac{\gamma_b}{2}} |t-s|$$

hold for all $s, t \in [t_0, T]$ *and* $x_l \in \mathbb{R}^d$, *where* $l \in \{0, 1, \dots, D\}$.

vii) *Let the growth exponents* $\beta, \varrho_a, \varrho_b, \gamma_a, \gamma_b \in [0, \infty[$ *be specified by assumptions iii), v), and vi). The initial condition* ξ *belongs to* $S^{\tilde{p}}([t_0 - \tau, t_0] \times \Omega; \mathbb{R}^d)$ *where*

$$\tilde{p} = p \cdot \max\{\gamma_a, \gamma_b, 2\beta + 2, \varrho_a + 2, \varrho_b + 2\},$$

and its realizations are P*-almost surely continuous. In addition, there exists a constant* $L_\xi > 0$ *such that*

$$\|\xi_t - \xi_s\|_{L^{((\varrho_a \vee \varrho_b)+2)p}(\Omega;\mathbb{R}^d)} \leq L_\xi |t-s|$$

holds for all $s, t \in [t_0 - \tau, t_0]$.

Before we state our results on the Milstein scheme's convergence, we elucidate problems that arise in comparison to SODEs, and we elaborate on those results in [60, 80, 137].

In case of SODEs, the convergence analysis of numerical schemes of higher order is usually done by applying Itô's formula to the SODEs' coefficients, cf. [78, 105]. This standard technique does however not apply to coefficients that also depend on the past history of the SDDE's solution because the $\mathbb{R}^{d\times(D+1)}$-valued and $(\mathscr{F}_t)_{t\in[t_0,T]}$-adapted process

$$\big((X_t, X_{t-\tau_1}, \ldots, X_{t-\tau_D})\big)_{t\in[t_0,T]} \tag{IV.34}$$

is not a semimartingale with respect to filtration $(\mathscr{F}_t)_{t\in[t_0,T]}$ or any natural filtration, cf. [60, pp. 269–270].

In [60, 137], the authors develop an Itô formula for functionals on processes of the form (IV.34) using the Malliavin calculus. The integrals occurring there are however not defined in the sense of Itô anymore but are Skorohod integrals, see [137, Theorem 4.7 on p. 57, p. 110 and term R_3^p on p. 116] and [60, Theorem 2.1 on p. 271, p. 294 and term R_3^π on p. 303].

Although Hu, Mohammed, and Yan claim to prove the convergence of the Milstein approximation in $L^p(\Omega;\mathbb{R}^d)$ for all $p\in[1,\infty[$, where the initial condition of the SDDE under consideration is random, see [60, p. 269], they only prove convergence in $L^2(\Omega;\mathbb{R}^d)$ with order $\alpha=1$ for deterministic initial conditions, see [60, Theorem 5.2].

In [137, Theorem 9.2], Yan also states the convergence in $L^2(\Omega;\mathbb{R}^d)$ of the Milstein approximation for random initial conditions. However, his proof of [137, Theorem 9.2] only holds true for deterministic initial conditions as well. As already mentioned in Section III.2, if the initial condition is random and $\mathscr{F}_{t_0}$-measurable, the solution X of the SDDE is not $\mathscr{G}$-measurable. Thus, X_t cannot belong to the space $\mathscr{D}^2(\Omega;\mathbb{R}^d)$ for any $t\in[t_0-\tau,T]$. Then, [137, Proposition 7.4] does not hold true, and also [137, Theorem 4.7] cannot be applied, where the initial condition is assumed to be a deterministic and continuous function.

In summary, [60, Theorem 5.2] and [137, Theorem 9.2] state the convergence in $L^2(\Omega;\mathbb{R}^d)$ with order $\alpha=1$ of the Milstein scheme for SDDEs with initial conditions in $C([t_0-\tau,t_0];\mathbb{R}^d)$.

Just in case of linear drift and diffusion coefficients, Yan provide the convergence of the Milstein approximation regarding linear SDDEs with random initial conditions in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ for arbitrary $p\in[1,\infty[$, see [137, pp. 119–120]. However, he only obtains the order of convergence $\alpha=\frac{1}{2}+\frac{1}{p}$, and not $\alpha=1$, as we would expect, see [137, Theorem 9.3].

Instead of using a generalized Itô formula like in [60, 137], Kloeden and Shardlow applied the deterministic Taylor's formula to the SDDE's coefficients in [80]. The occurring stochastic integrals are then all well-defined in the sense of Itô. In [80, Theorem 7.4], they claim the convergence in $S^2([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ of the Milstein approximation. However, there is gap in the proof of [80, Lemma 5.1] on which [80, Theorem 7.4] is based. We discuss this gap in the following.

Kloeden and Shardlow claim in [80, Proof of Lemma 5.1 on p. 190]: "If $S_k=\sum_{j=0}^{k-1} r_{j+1}$, where r_k are $\mathbf{R}^d$ valued $\mathscr{F}_{t_k}$ measurable random variables, then $S_k-\mathbf{E}S_k$ is a discrete martingale, and Doob's maximal inequality gives $\mathbf{E}\sup_{k\le n}\|S_k-\mathbf{E}S_k\|^2_{\mathbf{R}^d} \le 2\mathbf{E}\|S_k-\mathbf{E}S_k\|^2_{\mathbf{R}^d} \le \ldots$". Here, the symbol $\mathbf{E}=\mathrm{E}[\cdot]$ denotes the expectation on $(\Omega,\mathscr{F},\mathrm{P})$, and we have $k\in\{1,\ldots,N\}$. Apart from the fact that Doob's maximal inequality only holds true with a factor 4 instead of factor 2, cf. [35, Theorem 3.4 on p. 317] or [67, Theorem 26.3], the time-discrete process $(S_k-\mathrm{E}[S_k])_{k\in\{1,\ldots,N\}}$ is in general not a discrete martingale nor a submartingale with respect to filtration $(\mathscr{F}_{t_k})_{k\in\{1,\ldots,N\}}$. Thus, Doob's maximal inequality cannot even be applied. We provide an example where the discrete martingale property does not hold. Let $d=m$, and set

$r_k = W_{t_k}$ for $k \in \{1, \ldots, N\}$. Then, it holds $\mathrm{E}[S_k] = 0$ for all $k \in \{1, \ldots, N\}$, and we P-almost surely have

$$\begin{aligned}\mathrm{E}\big[S_k - \mathrm{E}[S_k]\big|\mathscr{F}_{t_{k-1}}\big] &= \sum_{j=1}^{k-1} \mathrm{E}[W_{t_{j+1}}|\mathscr{F}_{t_{k-1}}] = S_{k-1} + \mathrm{E}[W_{t_k}|\mathscr{F}_{t_{k-1}}] \\ &= S_{k-1} - \mathrm{E}[S_{k-1}] + W_{t_{k-1}}\end{aligned}$$

for all $k \in \{2, \ldots, N\}$. That is, $(S_k - \mathrm{E}[S_k])_{k\in\{1,\ldots,N\}}$ is not a discrete martingale. Further, the time-discrete process $\big(\sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j})\big)_{k\in\{1,\ldots,N\}}$ from [80, Equation (5.1) from Lemma 5.1] seems not be a discrete martingale as well if the drift coefficient depends on the past history of the solution, cf. inequality (IV.43) below. In addition, if the time-discrete process $\big(\sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j})\big)_{k\in\{1,\ldots,N\}}$ from [80, Equation (5.1) from Lemma 5.1] would be a discrete martingale, then, we cannot just apply Doob's maximal inequality as stated in [80, Proof of Lemma 5.1 on p. 190] but also the discrete Burkholder inequality, see Theorem II.3, and the technical considerations in [80, Lemma 7.3] as well as in the proof of [80, Theorem 7.4] would not be needed, and the standard Itô calculus would be sufficient.

Thus, considering the proof of [80, Lemma 5.1], there is a gap in estimating the supremum over time inside the expectation, and we only obtain the convergence in $L^2(\Omega; \mathbb{R}^d)$ with order $\alpha = 1$ of the Milstein scheme in [80, Theorem 7.4] like in [60, 137]. However, according to the title of article [80] of Kloeden and Shardlow , the main contribution of their work is not the proof of convergence of the Milstein scheme for SDDE in $S^2([t_0 - \tau, T] \times \Omega; \mathbb{R}^d)$, it is rather providing a proof that does not involve Skorohod integrals and techniques from the Malliavin calculus, also see [80, p. 182]. After applying deterministic Taylor expansions to the SDDE's coefficients, their analysis is decisively based on the inner product of $L^2(\Omega; \mathbb{R}^d)$ and on the differentiability of the SDDE's solution with respect to its initial condition, see [80, Lemma 7.3 and Theorem 7.4].

As already mentioned in the introduction of this chapter, the latter is closely related to the Malliavin derivative of the SDDE's solution, see [58] and [113, p. 126]. Thus, in the use of the Malliavin derivative of the SDDE's solution, see [60, Proposition 3.1] and [137, Proposition 7.4], and in the use of the derivative of the SDDE's solution with respect to its initial condition, see [80, Theorem 3.5], the proofs in [60, 137] and [80] are based on similar results.

Similarly to Kloeden and Shardlow in [80], we apply deterministic Taylor expansions to the coefficients of SDDE (II.1) in order to analyze the order of convergence of the Milstein scheme in $S^p([t_0 - \tau, T] \times \Omega; \mathbb{R}^d)$ for $p \in [1, \infty[$. In the following, we develop these expansions. Afterwards, the difficulty encountered in proving the order of convergence $\alpha = 1$ in case of SDDEs compared to SODEs is elucidated on the basis of these expansions.

As a starting point, we consider formulas (IV.20) and (IV.21) from the previous section. Let either $f = a$ or $f = b^j$ in the sequel. Using Taylor's formula [57, p. 284] on term $f(\mathcal{T}(\lfloor s \rfloor, X_s)) -$

$f(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}))$ in expansions (IV.20) and (IV.21), we obtain

$$\begin{aligned} &f(\mathcal{T}(\lfloor s\rfloor, X_s)) - f(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}))\\ &\quad = \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} f(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}))(X^i_{s-\tau_l} - X^i_{\lfloor s\rfloor-\tau_l})\\ &\qquad + \sum_{l_1,l_2=0}^{D}\sum_{i_1,i_2=1}^{d} \int_0^1 \partial_{x_{l_1}^{i_1}}\partial_{x_{l_2}^{i_2}} f(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor} + \theta(X_s - X_{\lfloor s\rfloor})))(1-\theta)\,\mathrm{d}\theta\\ &\qquad\quad \times (X^{i_1}_{s-\tau_{l_1}} - X^{i_1}_{\lfloor s\rfloor-\tau_{l_1}})(X^{i_2}_{s-\tau_{l_2}} - X^{i_2}_{\lfloor s\rfloor-\tau_{l_2}}) \end{aligned} \tag{IV.35}$$

for all $s \in [t_0, T]$ where

$$\begin{aligned} &\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor} + \theta(X_s - X_{\lfloor s\rfloor}))\\ &\quad := \big(\lfloor s\rfloor, \lfloor s\rfloor - \tau_1, \dots, \lfloor s\rfloor - \tau_D, X_{\lfloor s\rfloor} + \theta(X_s - X_{\lfloor s\rfloor}),\\ &\qquad\quad X_{\lfloor s\rfloor-\tau_1} + \theta(X_{s-\tau_1} - X_{\lfloor s\rfloor-\tau_1}), \dots, X_{\lfloor s\rfloor-\tau_D} + \theta(X_{s-\tau_D} - X_{\lfloor s\rfloor-\tau_D})\big). \end{aligned}$$

Using equation (IV.25) and that X is the strong solution of SDDE (II.1), it holds

$$\begin{aligned} &X_{s-\tau_l} - X_{\lfloor s\rfloor-\tau_l}\\ &\quad = \xi_{(s-\tau_l)\wedge t_0} - \xi_{(\lfloor s\rfloor-\tau_l)\wedge t_0}\\ &\qquad + \int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} a(\mathcal{T}(u, X_u))\,\mathrm{d}u + \sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^j(\mathcal{T}(u, X_u))\,\mathrm{d}W_u^j \end{aligned} \tag{IV.36}$$

for all $s \in [t_0, T]$ P-almost surely. Inserting this into expansion (IV.35), we obtain, similarly to formula (IV.20), the expansion

$$\begin{aligned} &a(\mathcal{T}(s, X_s))\\ &\quad = a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor})) + a(\mathcal{T}(s, X_s)) - a(\mathcal{T}(\lfloor s\rfloor, X_s))\\ &\qquad + \sum_{l=1}^{D}\sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}))(\xi^i_{(s-\tau_l)\wedge t_0} - \xi^i_{(\lfloor s\rfloor-\tau_l)\wedge t_0})\\ &\qquad + \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor})) \int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} a^i(\mathcal{T}(u, X_u))\,\mathrm{d}u\\ &\qquad + \sum_{l=0}^{D}\sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor})) \sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u))\,\mathrm{d}W_u^j\\ &\qquad + \sum_{l_1,l_2=0}^{D}\sum_{i_1,i_2=1}^{d} \int_0^1 \partial_{x_{l_1}^{i_1}}\partial_{x_{l_2}^{i_2}} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor} + \theta(X_s - X_{\lfloor s\rfloor})))(1-\theta)\,\mathrm{d}\theta\\ &\qquad\quad \times (X^{i_1}_{s-\tau_{l_1}} - X^{i_1}_{\lfloor s\rfloor-\tau_{l_1}})(X^{i_2}_{s-\tau_{l_2}} - X^{i_2}_{\lfloor s\rfloor-\tau_{l_2}}) \end{aligned} \tag{IV.37}$$

for all $s \in [t_0, T]$ P-almost surely.

The same expansion holds for the diffusion coefficients, too. However, we further expand the integrand $u \mapsto b^{i,j}(\mathcal{T}(u, X_u))$ of the stochastic integral in equation (IV.36). By substituting $\lfloor s\rfloor$

with $(\lfloor s \rfloor - \tau_l) \vee t_0$ in the expansion (IV.21), we have

$$\begin{aligned} b^j(\mathcal{T}(u, X_u)) = b^j(\mathcal{T}(u, X_u)) &- b^j(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_u)) \\ &+ b^j(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_u)) - b^j(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_{(\lfloor s \rfloor - \tau_l) \vee t_0})) \\ &+ b^j(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_{(\lfloor s \rfloor - \tau_l) \vee t_0})) \end{aligned} \tag{IV.38}$$

for all $u, s \in [t_0, T]$ and $j \in \{1, \ldots, m\}$. Inserting above expansion (IV.38) into equation (IV.36) and this, in turn, into equation (IV.35), the expansion of the diffusion coefficients results in

$$\begin{aligned} &b^{j_1}(\mathcal{T}(s, X_s)) \\ &\quad = b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) + b^{j_1}(\mathcal{T}(s, X_s)) - b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_s)) \\ &\quad + \sum_{l=1}^{D} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}))(\xi^i_{(s-\tau_l)\wedge t_0} - \xi^i_{(\lfloor s \rfloor - \tau_l)\wedge t_0}) \\ &\quad + \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} a^i(\mathcal{T}(u, X_u)) \,\mathrm{d}u \\ &\quad + \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j_2=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_{(\lfloor s \rfloor - \tau_l)\vee t_0})) \,\mathrm{d}W_u^{j_2} \\ &\quad + \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j_2=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j_2}(\mathcal{T}(u, X_u)) \\ &\qquad\qquad - b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_u)) \,\mathrm{d}W_u^{j_2} \\ &\quad + \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j_2=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_u)) \\ &\qquad\qquad - b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_{(\lfloor s \rfloor - \tau_l)\vee t_0})) \,\mathrm{d}W_u^{j_2} \\ &\quad + \sum_{l_1,l_2=0}^{D} \sum_{i_1,i_2=1}^{d} \int_0^1 \partial_{x_{l_1}^{i_1}} \partial_{x_{l_2}^{i_2}} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor} + \theta(X_s - X_{\lfloor s \rfloor})))(1-\theta) \,\mathrm{d}\theta \\ &\qquad \times (X^{i_1}_{s-\tau_{l_1}} - X^{i_1}_{\lfloor s \rfloor - \tau_{l_1}})(X^{i_2}_{s-\tau_{l_2}} - X^{i_2}_{\lfloor s \rfloor - \tau_{l_2}}) \end{aligned} \tag{IV.39}$$

for all $s \in [t_0, T]$ and all $j_1 \in \{1, \ldots, m\}$ P-almost surely.

In expansions (IV.37) and (IV.39), all occurring stochastic integrals are well-defined in the sense of Itô as in [80] and in contrast to [60, 137].

However, the analysis of the Milstein scheme still needs more sophisticated techniques in order to obtain convergence of order $\alpha = 1$. To see this, we first consider the Euler-Maruyama scheme again. Its convergence results from, among others, inequality (IV.24) and Lemma (II.9). Using the triangle inequality first as in inequality (IV.22) and applying expansion (IV.37) on the right-hand side of inequality (IV.22) does however not improve the order of convergence due to the irregularity of the Wiener process as we will see in the sequel. Proceeding in this way, we would obtain the term

$$\int_{t_0}^{T} \Big\| \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l) \vee t_0} b^{i,j}(\mathcal{T}(u, X_u)) \, \mathrm{d}W_u^j \Big\|_{L^p(\Omega;\mathbb{R}^d)} \mathrm{d}s, \tag{IV.40}$$

and we have to show that its integrand is of order $\mathcal{O}(h)$ to obtain the order of convergence $\alpha = 1$. But in this regard, we get the following. Using the triangle inequality, Assumption IV.8 *ii)*, and Assumption IV.8 *iv)* as well as the Cauchy-Schwarz inequality and Theorem II.6, it holds, similarly to inequality (II.26), that

$$\begin{aligned}
&\Big\| \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l) \vee t_0} b^{i,j}(\mathcal{T}(u, X_u)) \, \mathrm{d}W_u^j \Big\|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad \leq \Big\| \sum_{l=0}^{D} \sum_{i=1}^{d} \| \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \| \Big| \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l) \vee t_0} b^{i,j}(\mathcal{T}(u, X_u)) \, \mathrm{d}W_u^j \Big| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq L_a \Big\| \sum_{l=0}^{D} \sum_{i=1}^{d} \Big| \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l) \vee t_0} b^{i,j}(\mathcal{T}(u, X_u)) \, \mathrm{d}W_u^j \Big| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq L_a \sqrt{d} \sum_{l=0}^{D} \Big\| \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l) \vee t_0} b^{j}(\mathcal{T}(u, X_u)) \, \mathrm{d}W_u^j \Big\|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad \leq L_a \sqrt{d} \sqrt{p-1} \sum_{l=0}^{D} \bigg(\int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l) \vee t_0} \Big\| \sum_{j=1}^{m} \| b^{j}(\mathcal{T}(u, X_u)) \|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \mathrm{d}u \bigg)^{\frac{1}{2}} \\
&\quad \leq L_a \sqrt{d} \sqrt{p-1} K_b \sqrt{m} (D+1) \big(1 + \|X\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \sqrt{h}.
\end{aligned} \tag{IV.41}$$

Thus, proceeding like this, we only obtain an order of convergence $\alpha = \frac{1}{2}$. This means, we have to analyze the process $(Z_t)_{t\in[t_0,T]}$ defined by

$$Z_t := \int_{t_0}^{t} \sum_{l=0}^{D} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l) \vee t_0} b^{i,j}(\mathcal{T}(u, X_u)) \, \mathrm{d}W_u^j \, \mathrm{d}s \tag{IV.42}$$

as a whole in the $S^p([t_0,T]\times\Omega;\mathbb{R}^d)$-norm. In doing so, we can take into account the dependencies between the intervals $[t_n, t_{n+1}]$, $n \in \{0, 1, \ldots, N-1\}$, of the discretization. So far, the same problem occurs in the analysis of the Milstein scheme for SODEs, see [78, Section 10.8] or [105, p. 17]. In case of SODEs – consider formula (IV.42) with $l = 0$ – and the analysis in $L^2(\Omega;\mathbb{R}^d)$, the higher order of convergence is obtained by utilizing the inner product of $L^2(\Omega;\mathbb{R}^d)$ and

$$\begin{aligned}
&\mathrm{E}\Big[\sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j=1}^{m} \int_{\lfloor s \rfloor}^{s} b^{i,j}(\mathcal{T}(u, X_u)) \, \mathrm{d}W_u^j \Big| \mathscr{F}_{\lfloor s \rfloor} \Big] \\
&\quad = \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \mathrm{E}\Big[\sum_{j=1}^{m} \int_{\lfloor s \rfloor}^{s} b^{i,j}(\mathcal{T}(u, X_u)) \, \mathrm{d}W_u^j \Big| \mathscr{F}_{\lfloor s \rfloor} \Big] \\
&\quad = 0
\end{aligned}$$

P-almost surely, see [78, Section 10.8]. That is, we employ the discrete martingale property of the time-discrete process $(Z_{t_n})_{n\in\{0,1,\ldots,N\}}$ in case of $l = 0$. This idea on the Hilbert space $L^2(\Omega;\mathbb{R}^d)$ can be transferred to $L^p(\Omega;\mathbb{R}^d)$ and $S^p([t_0,T]\times\Omega;\mathbb{R}^d)$, $p\in[2,\infty[$, using the Burkholder inequality.

However, if the drift coefficient depends on the past history of the solution in case of SDDEs, this technique cannot be used, because for $l\in\{1,\dots,D\}$, we P-almost surely have

$$\begin{aligned}
&\mathrm{E}\Bigg[\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}))\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j\Bigg|\mathscr{F}_{\lfloor s\rfloor}\Bigg]\\
&=\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}))\mathrm{E}\Bigg[\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j\Bigg|\mathscr{F}_{\lfloor s\rfloor}\Bigg]\\
&\neq 0
\end{aligned}\tag{IV.43}$$

in general. Moreover, one cannot simply move the random variable $\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}))$ into the Itô integral because it is $\mathscr{F}_{\lfloor s\rfloor}/\mathscr{B}(\mathbb{R}^d)$-measurable, but in general not $\mathscr{F}_{(\lfloor s\rfloor-\tau_l)\vee t_0}/\mathscr{B}(\mathbb{R}^d)$-measurable. Thus, the time-discrete process $(Z_{t_n})_{n\in\{0,1,\dots,N\}}$ is lacking the martingale property in general when $l\in\{1,\dots,D\}$.

In [60, 137], the authors solve this problem using the Skorohod integral and applying [113, Proposition 1.3.1], see [60, p. 303] and [137, pp. 118–119]. However, they do not consider the supremum over time to be inside the $L^2(\Omega;\mathbb{R}^d)$-norm. Recall that Doob's maximal inequality cannot be applied because process Z defined in equation (IV.42) is in general not a martingale nor a submartingale with respect to $(\mathscr{F}_t)_{t\in[t_0,T]}$ in continuous time nor with respect to $(\mathscr{F}_{t_n})_{n\in\{0,1,\dots,N\}}$ in discrete time. One may generalize the considerations in [60, p. 303] and [137, pp. 118–119] to the convergence in $S^p([t_0,T]\times\Omega;\mathbb{R}^d)$ for $p\in\,]2,\infty[$ using a maximal inequality for Skorohod integrals developed by Alòs and Nualart, see [3, Theorem 3.1]. But in order to directly apply [3, Theorem 3.1], stronger assumptions than in Assumption IV.8 must be made.

Inspired by the proofs of [3, Theorem 3.1], [60, Theorem 5.2], and [137, Theorem 9.2], we prove convergence of order $\alpha=1$ for Milstein approximation regarding SDDE (II.1) using the Malliavin calculus as well.

According to Assumption IV.8 *vii)*, let us emphasize that initial condition ξ of SDDE (II.1) is considered to be a random process in contrast to the results in [60, 80, 137]. In the first instance, the Malliavin calculus can just deal with random variables and processes that are measurable with respect to σ-algebra $\mathscr{G}$, see Chapter III. Since initial condition ξ is adapted to filtration $(\mathscr{F}_t)_{t\in[t_0-\tau,t_0]}$, for all $t\in[t_0-\tau,t_0]$, random variable ξ_t is in particular $\mathscr{F}_{t_0}/\mathscr{B}(\mathbb{R}^d)$-measurable and independent of the Wiener process $(W_t)_{t\in[t_0,T]}$. Thus, the Malliavin calculus cannot be applied to functionals of solution X of SDDE (II.1) when ξ is not deterministic.

One way to deal with random variables that are independent of σ-algebra $\mathscr{G}$ in the Malliavin calculus is the following. Let E be a real separable Hilbert space. Consider an E-valued random variable in $L^p(\Omega;E)$, $p\in[2,\infty[$, where $(\Omega,\mathscr{F},\mathrm{P})=(\Omega_1\times\Omega_2,\mathscr{G}\otimes\mathscr{F}_{t_0},\mathrm{P}_1\otimes\mathrm{P}_2)$ is a product probability space for a moment. Then, this random variable can be understood as a random variable in

$$L^p((\Omega_1,\mathscr{G},\mathrm{P}_1);L^p(\Omega_2,\mathscr{F}_{t_0},\mathrm{P}_2);(E,\mathscr{B}(E))),$$

that is, as a random variable that takes values in the Banach space $L^p((\Omega_2, \mathscr{F}_{t_0}, \mathrm{P}_2); (E, \mathscr{B}(E)))$. But proceeding in this way and doing the analysis of convergence thoroughly, we must presuppose this product structure of the underlying probability space. This is related to the partial Malliavin calculus, see e. g. [85]. Moreover, we have to deal with random variables that take values in Banach spaces. This further leads to stochastic integration and Malliavin calculus of Banach space valued random variables and stochastic processes. We refer to [93, 94, 117, 132, 133] for literature on these topics.

Our analysis of the convergence of the Milstein scheme regarding SDDE (II.1) with a random initial condition is not restricted to product probability spaces and uses simpler arguments. For more details on that, we refer to the proof of Theorem IV.9 and in particular to Lemma IV.19.

Let us now state the main result on the strong convergence of the Milstein approximation.

Theorem IV.9 (Strong Convergence of the Milstein Approximation)
Let SDDE (II.1) *fulfill Assumption IV.8 for some* $p \in]2, \infty[$, *and consider Milstein approximation* (IV.33) *regarding SDDE* (II.1).

Then, the family of Milstein approximations $(Y^h)_{h \in]0,T-t_0]}$ *converges in* $S^p([t_0-\tau, T] \times \Omega; \mathbb{R}^d)$ *with order* $\alpha = 1$ *to solution* X *of SDDE* (II.1) *as* $h \to 0$. *That is, there exists a constant* $C_{\text{Milstein}} > 0$, *independent of* h, *such that*

$$\|X - Y^h\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq C_{\text{Milstein}}\, h$$

for all $h \in]0, T-t_0]$.

Proof. For the proof of this theorem and details on the constant C_{Milstein}, we refer to Section IV.3, see p. 85. □

Remark IV.10
The Lipschitz continuity in Assumption IV.8 iii) can be seen as an extension of the third assumption in [78, Formula (10.3.21)] in case of SODEs. Using Taylor's formula, this Assumption IV.8 iii) can be neglected in view of Assumption IV.8 v). But then we have to assume that

$$\tilde{p} = p \cdot \max\{\gamma_a, \gamma_b, 2\beta+3, \varrho_a+2, \varrho_b+2\}$$

in Assumption IV.8 vii). Even in case of $\gamma_a = \gamma_b = \beta = \varrho_a = \varrho_b = 0$, *we consequently require* $\xi \in S^{3p}([t_0-\tau, t_0] \times \Omega; \mathbb{R}^d)$ *instead of* $\xi \in S^{2p}([t_0-\tau, t_0] \times \Omega; \mathbb{R}^d)$.

Using Hölder's inequality, we also obtain convergence of order $\alpha = 1$ in $S^q([t_0-\tau, T] \times \Omega; \mathbb{R}^d)$ of the Milstein scheme for all $q \in [1, p]$ where $p \in]2, \infty[$.

Corollary IV.11
Let SDDE (II.1) *fulfill Assumption IV.8 for some* $p \in]2, \infty[$, *and consider Milstein approximation* (IV.33) *regarding SDDE* (II.1).

Then, for all $q \in [1, p]$, *the family of Milstein approximations* $(Y^h)_{h \in]0,T-t_0]}$ *converges in* $S^q([t_0-\tau, T] \times \Omega; \mathbb{R}^d)$ *with order* $\alpha = 1$ *to solution* X *of SDDE* (II.1) *as* $h \to 0$. *That is, there exists a constant* $C_{\text{Milstein}} > 0$, *independent of* h, *such that*

$$\|X - Y^h\|_{S^q([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq C_{\text{Milstein}}\, h$$

for all $h \in \,]0, T - t_0]$ *and all* $q \in [1, p]$.

According to this corollary, if SDDE (II.1) fulfills Assumption IV.8 for all $p \in \,]2, \infty[$, the Milstein approximation converges in $S^p([t_0 - \tau, T] \times \Omega; \mathbb{R}^d)$ for all $p \in [1, \infty[$. Using Lemma IV.3, we obtain the following result on the pathwise convergence.

Corollary IV.12 (Pathwise Convergence of the Milstein Approximation)
Let SDDE (II.1) *fulfill Assumption IV.8 for all* $p \in \,]2, \infty[$. *Consider the family of Milstein approximations* $(Y^{h_N})_{N \in \mathbb{N}}$ *regarding SDDE* (II.1) *from equation* (IV.33), *where* $(h_N)_{N \in \mathbb{N}} \subset \,]0, T - t_0]$. *Let* $q_\varepsilon \in [1, \infty[$ *for all* $\varepsilon > 0$ *be independent of* N *and such that* $\sum_{N=1}^{\infty} h_N^{\varepsilon q_\varepsilon} < \infty$.

Then, the family of Milstein approximations $(Y^{h_N})_{N \in \mathbb{N}}$ *converges pathwise with order* $\alpha = 1 - \varepsilon$ *to solution* X *of SDDE* (II.1) *for arbitrary* $\varepsilon > 0$ *as* $N \to \infty$. *That is, for all* $\varepsilon > 0$, *there exists a positive random variable* Z_ε, *which belongs to* $L^p(\Omega; \mathbb{R})$ *for all* $p \in [1, \infty[$, *such that*

$$\sup_{t \in [t_0 - \tau, T]} \|X_t - Y_t^{h_N}\| \leq Z_\varepsilon \, h_N^{1-\varepsilon}$$

P-*almost surely for all* $N \in \mathbb{N}$.

Using the results from Theorem IV.9 and Corollary IV.12, we can improve the results on the convergence of the Euler-Maruyama scheme (IV.13) under certain conditions. If the diffusion coefficient b^j at most depends on time t and not on the solution X for all $j \in \{1, \dots, m\}$, Milstein scheme (IV.33) simplifies to Euler-Maruyama scheme (IV.13). In this case, the noise of the SDDE is *additive*, and we have

$$b^j(t, t - \tau_1, \dots, t - \tau_D, X_t, X_{t-\tau_1}, \dots, X_{t-\tau_D}) = b^j(t, t - \tau_1, \dots, t - \tau_D)$$

for all $t \in [t_0, T]$ and $j \in \{1, \dots, m\}$. According to Theorem IV.9, the Euler-Maruyama scheme then converges strongly with order $\alpha = 1$, cf. [78, p. 341] in case of SODEs.

Corollary IV.13
Let SDDE (II.1) *have* additive noise *and fulfill Assumption IV.8 for some* $p \in \,]2, \infty[$, *where* $L_b = L_{\partial b} = \beta = K_b = K_{\partial^2 b} = \varrho_b = 0$ *consequently. Consider Euler-Maruyama approximation* (IV.13) *regarding SDDE* (II.1).

Then, the family Euler-Maruyama approximations $(Y^h)_{h \in]0, T - t_0]}$ *converges in* $S^p([t_0 - \tau, T] \times \Omega; \mathbb{R}^d)$ *with order* 1 *to solution* X *of SDDE* (II.1) *as* $h \to 0$. *That is, there exists a constant* $C_{\text{Euler}} > 0$, *independent of* h, *such that*

$$\|X - Y^h\|_{S^p([t_0 - \tau, T] \times \Omega; \mathbb{R}^d)} \leq C_{\text{Euler}} \, h$$

for all $h \in \,]0, T - t_0]$.

Further, we obtain a similar result for the pathwise convergence by Corollary IV.12.

Corollary IV.14
Let SDDE (II.1) *have* additive noise *and fulfill Assumption IV.8 for all* $p \in \,]2, \infty[$, *where* $L_b = L_{\partial b} = \beta = K_b = K_{\partial^2 b} = \varrho_b = 0$ *consequently. Consider the family of Euler-Maruyama*

approximations $(Y^{h_N})_{N\in\mathbb{N}}$ *regarding SDDE* (II.1) *from equation* (IV.13)*, where* $(h_N)_{N\in\mathbb{N}} \subset]0, T-t_0]$*. Let* $q_\varepsilon \in [1,\infty[$ *for all* $\varepsilon > 0$ *be independent of* N *and such that* $\sum_{N=1}^{\infty} h_N^{\varepsilon q_\varepsilon} < \infty$*.*

Then, the family of Euler-Maruyama approximations $(Y^{h_N})_{N\in\mathbb{N}}$ *converges pathwise with order* $1-\varepsilon$ *to solution* X *of SDDE* (II.1) *for arbitrary* $\varepsilon > 0$ *as* $N \to \infty$*. That is, for all* $\varepsilon > 0$*, there exists a positive random variable* Z_ε*, which belongs to* $L^p(\Omega;\mathbb{R})$ *for all* $p \in [1,\infty[$*, such that*

$$\sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\| \le Z_\varepsilon \, h_N^{1-\varepsilon}$$

P*-almost surely for all* $N \in \mathbb{N}$*.*

IV.3. Proofs

Proof of Lemma IV.3

Proof of Lemma IV.3. We follow the proofs of [41, Proposition 23] and [77, Lemma 2.1], and we generalize their concepts to nonequidistant discretizations, cf. [2, Theorem 3.2].

Let us fix an $\varepsilon > 0$ and a

$$p > 2q_\varepsilon + \frac{1}{\varepsilon} \tag{IV.44}$$

with q_ε from assumption (IV.6). At first, it holds

$$\Big(h_N^{-(\alpha-\varepsilon)} \sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\|\Big)^p \le h_N^{-p(\alpha-\varepsilon)} \sum_{n=0}^{N-1} \sup_{t\in[t_0-\tau,t_0]\cup]t_n^N,t_{n+1}^N]} \|X_t - Y_t^{h_N}\|^p, \tag{IV.45}$$

and further, condition (IV.6) implies $h_N \to 0$ as $N \to \infty$. In the following, we use the Borel-Cantelli Lemma in order to show the P-almost surely convergence in (IV.8), cf. [43, Satz 1.11.8]. Using inequality (IV.45) and Markov's inequality, we obtain

$$\begin{aligned}
&\sum_{N=1}^{\infty} \mathrm{P}\left[h_N^{-(\alpha-\varepsilon)} \sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\| > h_N^{\frac{1}{p}}\right] \\
&\quad\le \sum_{N=1}^{\infty} h_N^{-1} h_N^{-p(\alpha-\varepsilon)} \mathrm{E}\left[\sum_{n=0}^{N-1} \sup_{t\in[t_0-\tau,t_0]\cup]t_n^N,t_{n+1}^N]} \|X_t - Y_t^{h_N}\|^p\right] \\
&\quad= \sum_{N=1}^{\infty} h_N^{-1} h_N^{-p(\alpha-\varepsilon)} \sum_{n=0}^{N-1} \|X - Y^{h_N}\|^p_{S^p(([t_0-\tau,t_0]\cup]t_n^N,t_{n+1}^N])\times\Omega;\mathbb{R}^d)} \\
&\quad\le \sum_{N=1}^{\infty} h_N^{-1} h_N^{-p(\alpha-\varepsilon)} \sum_{n=0}^{N-1} C_p^p h_N^{\alpha p} \\
&\quad= C_p^p \sum_{N=1}^{\infty} N h_N^{p\varepsilon-1}.
\end{aligned} \tag{IV.46}$$

According to assumption (IV.6), it holds $h_N^{\varepsilon q_\varepsilon} = \mathcal{O}(N^{-1})$ as $N \to \infty$. This implies $N h_N^{\varepsilon q_\varepsilon} = \mathcal{O}(1)$, and for all $\varepsilon > 0$, there exists a constant $K_\varepsilon > 0$ such that $N h_N^{\varepsilon q_\varepsilon} \leq K_\varepsilon$ for all $N \in \mathbb{N}$. In view of condition (IV.44), we then obtain by inequality (IV.46) that

$$\begin{aligned} &\sum_{N=1}^{\infty} \mathrm{P}\Big[h_N^{-(\alpha-\varepsilon)} \sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\| > h_N^{\frac{1}{p}}\Big] \\ &\quad \leq C_p^p K_\varepsilon \sum_{N=1}^{\infty} h_N^{p\varepsilon-1-\varepsilon q_\varepsilon} \\ &\quad \leq C_p^p K_\varepsilon (T-t_0)^{p\varepsilon-1-2\varepsilon q_\varepsilon} \sum_{N=1}^{\infty} h_N^{\varepsilon q_\varepsilon} \\ &\quad < \infty, \end{aligned} \tag{IV.47}$$

and hence,

$$h_N^{-(\alpha-\varepsilon)} \sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\| \to 0$$

converges P-almost surely as $N \to \infty$ for all $\varepsilon > 0$.

Next, we show the existence of random variable Z_ε in inequality (IV.9). We set

$$Z_\varepsilon \coloneqq \sup_{N\in\mathbb{N}} h_N^{-(\alpha-\varepsilon)} \sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\|.$$

If $\sup_{N\in\mathbb{N}}\|X_t - Y_t^{h_N}\| = 0$ P-almost surely, inequality (IV.9) is clearly true, so let $\sup_{N\in\mathbb{N}}\|X_t - Y_t^{h_N}\| > 0$ P-almost surely in the following.

Note that $Z_\varepsilon \colon \Omega \to \mathbb{R}$ is positive and $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable. Moreover, using inequality (IV.45) and the monotone convergence theorem, we obtain

$$\begin{aligned} \mathrm{E}\big[|Z_\varepsilon|^p\big] &\leq \mathrm{E}\Big[\sup_{N\in\mathbb{N}} h_N^{-p(\alpha-\varepsilon)} \sum_{n=0}^{N-1} \sup_{t\in[t_0-\tau,t_0]\cup]t_n^N,t_{n+1}^N]} \|X_t - Y_t^{h_N}\|^p\Big] \\ &\leq \mathrm{E}\Big[\sum_{N=1}^{\infty} h_N^{-p(\alpha-\varepsilon)} \sum_{n=0}^{N-1} \sup_{t\in[t_0-\tau,t_0]\cup]t_n^N,t_{n+1}^N]} \|X_t - Y_t^{h_N}\|^p\Big] \\ &= \sum_{N=1}^{\infty} h_N^{-p(\alpha-\varepsilon)} \sum_{n=0}^{N-1} \|X - Y^{h_N}\|^p_{S^p(([t_0-\tau,t_0]\cup]t_n^N,t_{n+1}^N])\times\Omega;\mathbb{R}^d)}. \end{aligned}$$

Then, similarly to the considerations in inequalities (IV.46) and (IV.47), it follows $\mathrm{E}[|Z_\varepsilon|^p] < \infty$. Since $\|Z_\varepsilon\|_{L^p(\Omega;\mathbb{R})} < \infty$ for all $p \in\,]2q_\varepsilon + \frac{1}{\varepsilon}, \infty[$, cf. condition (IV.44), Hölder's inequality implies $\|Z_\varepsilon\|_{L^q(\Omega;\mathbb{R})} < \infty$ for all $q \in [1, \infty[$. Finally, inequality (IV.9) follows by

$$\sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\| \leq \Big(\sup_{N\in\mathbb{N}} h_N^{-(\alpha-\varepsilon)} \sup_{t\in[t_0-\tau,T]} \|X_t - Y_t^{h_N}\|\Big) h_N^{\alpha-\varepsilon} = Z_\varepsilon h_N^{\alpha-\varepsilon}$$

P-almost surely for all $\varepsilon > 0$, and the proof is complete. □

Proof of Theorem IV.6

In order to show the strong convergence of the Euler-Maruyama scheme, we first have to ensure the boundedness of its moments.

Lemma IV.15
Let the Borel-measurable drift a and diffusion b^j, $j \in \{1, \dots, m\}$, of SDDE (II.1) *satisfy the global Lipschitz and linear growth conditions* (II.8), (II.9), (II.10), *and* (II.11). *Further, let initial condition ξ belong to $S^p([t_0-\tau, t_0] \times \Omega; \mathbb{R}^d)$ for some $p \in [2, \infty[$.*

Considering Euler-Maruyama approximation Y from formula (IV.13) *regarding SDDE* (II.1), *it holds*

$$1 + \|Y\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq \big(1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\big)\, \mathrm{e}^{2\left(K_a\sqrt{T-t_0}+\frac{p}{\sqrt{p-1}}K_b\sqrt{m}\right)^2(T-t_0)}.$$

Proof. Since $\xi \in S^p([t_0-\tau, t_0] \times \Omega; \mathbb{R}^d)$, we have $\|Y\|_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} < \infty$. We assume that

$$\|Y\|_{S^p([t_0-\tau,t_\nu]\times\Omega;\mathbb{R}^d)} < \infty$$

has been proven for all $\nu \in \{0, 1, \dots, n-1\}$ where $n \in \{1, \dots, N\}$. For all $n \in \{1, \dots, N\}$, inequality (II.6) and the triangle inequality imply

$$\begin{aligned} 1 &+ \|Y\|^2_{S^p([t_0-\tau,t_n]\times\Omega;\mathbb{R}^d)} \\ &\leq 1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} + 2\bigg(\Big\|\int_{t_0}^{\cdot} a(\mathcal{T}(\lfloor s\rfloor, Y_{\lfloor s\rfloor}))\,\mathrm{d}s\Big\|_{S^p([t_0-\tau,t_n]\times\Omega;\mathbb{R}^d)} \\ &\qquad + \Big\|\sum_{j=1}^m \int_{t_0}^{\cdot} b^j(\mathcal{T}(\lfloor s\rfloor, Y_{\lfloor s\rfloor}))\,\mathrm{d}W_s^j\Big\|_{S^p([t_0-\tau,t_n]\times\Omega;\mathbb{R}^d)}\bigg)^2. \end{aligned} \tag{IV.48}$$

In the following, we estimate the two last $S^p([t_0, T] \times \Omega; \mathbb{R}^d)$-norms on the right-hand side of inequality (IV.48) above. Using the triangle inequality and linear growth condition (II.10), we obtain

$$\begin{aligned} \Big\|&\int_{t_0}^{\cdot} a(\mathcal{T}(\lfloor s\rfloor, Y_{\lfloor s\rfloor}))\,\mathrm{d}s\Big\|_{S^p([t_0-\tau,t_n]\times\Omega;\mathbb{R}^d)} \\ &\leq \int_{t_0}^{t_n} \|a(\mathcal{T}(\lfloor s\rfloor, Y_{\lfloor s\rfloor}))\|_{L^p(\Omega;\mathbb{R}^d)}\,\mathrm{d}s \\ &\leq K_a \int_{t_0}^{t_n} \Big\|\sup_{l\in\{0,1,\dots,D\}} \big(1 + \|Y_{\lfloor s\rfloor-\tau_l}\|^2\big)^{\frac{1}{2}}\Big\|_{L^p(\Omega;\mathbb{R})}\,\mathrm{d}s \\ &\leq K_a \int_{t_0}^{t_n} \big(1 + \|Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\,\mathrm{d}s \\ &\leq K_a\sqrt{t_n - t_0}\bigg(\int_{t_0}^{t_n} 1 + \|Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s\bigg)^{\frac{1}{2}}, \end{aligned} \tag{IV.49}$$

where the Cauchy-Schwarz inequality is used in the last step. Similarly to this estimate, the Zakai's inequality from Theorem II.6 and linear growth condition (II.11) lead to

$$\begin{aligned}
&\Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor}))\,\mathrm{d}W_s^j \Big\|_{S^p([t_0-\tau,t_n]\times\Omega;\mathbb{R}^d)} \\
&\quad \le \frac{p}{\sqrt{p-1}} \Big(\int_{t_0}^{t_n} \Big\| \sum_{j=1}^{m} \|b^j(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor}))\|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \,\mathrm{d}s \Big)^{\frac{1}{2}} \\
&\quad \le \frac{p}{\sqrt{p-1}} K_b \sqrt{m} \Big(\int_{t_0}^{t_n} 1 + \|Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s \Big)^{\frac{1}{2}}.
\end{aligned} \tag{IV.50}$$

Inserting the results from inequalities (IV.49) and (IV.50) into inequality (IV.48), we obtain

$$\begin{aligned}
&1 + \|Y\|^2_{S^p([t_0-\tau,t_n]\times\Omega;\mathbb{R}^d)} \\
&\quad \le 1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} \\
&\qquad + 2\Big(K_a \sqrt{t_n - t_0} + \frac{p}{\sqrt{p-1}} K_b \sqrt{m} \Big)^2 \int_{t_0}^{t_n} 1 + \|Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s.
\end{aligned}$$

Then, Gronwall's Lemma II.7 implies

$$1 + \|Y\|^2_{S^p([t_0-\tau,t_n]\times\Omega;\mathbb{R}^d)} \le \big(1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\big)\, \mathrm{e}^{2\left(K_a\sqrt{t_n-t_0} + \frac{p}{\sqrt{p-1}} K_b \sqrt{m}\right)^2 (t_n - t_0)}$$

for all $n \in \{1, \dots, N\}$, and, by taking the maximum over $n \in \{1, \dots, N\}$ on both sides of the inequality above, we finally have

$$1 + \|Y\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \le \big(1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\big)\, \mathrm{e}^{2\left(K_a\sqrt{T-t_0} + \frac{p}{\sqrt{p-1}} K_b \sqrt{m}\right)^2 (T - t_0)}.$$

□

Proof of Theorem IV.6. Consider the difference of the solution X and approximation Y in formula (IV.19). Substituting the integrands in formula (IV.19) by their expansions (IV.20) and (IV.21), we have, after the application of the triangle inequality, that

$$\begin{aligned}
&\|X - Y\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le \Big\| \int_{t_0}^{\cdot} a(\mathcal{T}(s, X_s)) - a(\mathcal{T}(\lfloor s \rfloor, X_s))\,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad + \Big\| \int_{t_0}^{\cdot} a(\mathcal{T}(\lfloor s \rfloor, X_s)) - a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}))\,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad + \Big\| \int_{t_0}^{\cdot} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) - a(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor}))\,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad + \Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(s, X_s)) - b^j(\mathcal{T}(\lfloor s \rfloor, X_s))\,\mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad + \Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(\lfloor s \rfloor, X_s)) - b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}))\,\mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\qquad + \Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) - b^j(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor}))\,\mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}.
\end{aligned} \tag{IV.51}$$

We estimate the terms on the right-hand side of the inequality (IV.51) above term by term in the following. We start with the first term. Using the triangle inequality and the assumption from inequality (IV.16), it holds

$$\begin{aligned}
&\Big\| \int_{t_0}^{\cdot} a(\mathcal{T}(s,X_s)) - a(\mathcal{T}(\lfloor s \rfloor, X_s)) \,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le \int_{t_0}^{T} \| a(\mathcal{T}(s,X_s)) - a(\mathcal{T}(\lfloor s \rfloor, X_s)) \|_{L^p(\Omega;\mathbb{R}^d)} \,\mathrm{d}s \\
&\quad \le L_{\mathrm{t},a} \int_{t_0}^{T} \Big\| \sup_{l\in\{0,1,\dots,D\}} \big(1+\|X_{s-\tau_l}\|^2\big)^{\frac{\gamma_a}{2}} \Big\|_{L^p(\Omega;\mathbb{R})} \sqrt{s-\lfloor s \rfloor} \,\mathrm{d}s \\
&\quad \le L_{\mathrm{t},a} \Big\| \big(1+\|X\|^2\big)^{\frac{\gamma_a}{2}} \Big\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R})} \int_{t_0}^{T} \sqrt{s-\lfloor s \rfloor} \,\mathrm{d}s \\
&\quad \le L_{\mathrm{t},a} \big(1+\|X\|^2_{S^{(\gamma_a\vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R})}\big)^{\frac{\gamma_a}{2}} \int_{t_0}^{T} \sqrt{s-\lfloor s \rfloor} \,\mathrm{d}s \\
&\quad \le L_{\mathrm{t},a} \big(1+\|X\|^2_{S^{(\gamma_a\vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_a}{2}} \frac{2}{3}(T-t_0)\sqrt{h},
\end{aligned} \tag{IV.52}$$

where inequality (IV.31) is used in the last step. Here, we take the maximum $\gamma_a \vee 1$ in order to ensure that $(\gamma_a \vee 1)p \ge 1$, and hence, $\|\cdot\|_{S^{(\gamma_a\vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}$ is actually a norm. We already estimated the second term of the right-hand side of inequality (IV.51) in inequality (IV.32), where

$$C_1 = \big(K_a\sqrt{T-t_0} + \sqrt{p-1}K_b\sqrt{m}\big)\big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}$$

by Lemma II.9. Continuing with the third term, the triangle inequality and the Lipschitz condition (II.8), cf. inequality (IV.23), imply

$$\begin{aligned}
&\Big\| \int_{t_0}^{\cdot} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) - a(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor})) \,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le L_a \int_{t_0}^{T} \|X-Y\|_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s \\
&\quad \le L_a\sqrt{T-t_0}\Big(\int_{t_0}^{T} \|X-Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s\Big)^{\frac{1}{2}},
\end{aligned} \tag{IV.53}$$

where we used the Cauchy-Schwarz inequality in the last step. The stochastic integrals in inequality (IV.51) can be estimated using similar arguments as above after applying Theorem II.6. We obtain

$$\begin{aligned}
&\Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(s,X_s)) - b^j(\mathcal{T}(\lfloor s \rfloor, X_s)) \,\mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \le \frac{p}{\sqrt{p-1}} \Big(\int_{t_0}^{T} \Big\| \sum_{j=1}^{m} \| b^j(\mathcal{T}(s,X_s)) - b^j(\mathcal{T}(\lfloor s \rfloor, X_s)) \|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \,\mathrm{d}s \Big)^{\frac{1}{2}} \\
&\quad \le \frac{p}{\sqrt{p-1}} L_{\mathrm{t},b}\sqrt{m} \Big(\int_{t_0}^{T} \Big\| \sup_{l\in\{0,1,\dots,D\}} \big(1+\|X_{s-\tau_l}\|^2\big)^{\frac{\gamma_b}{2}} \Big\|^2_{L^p(\Omega;\mathbb{R})} (s-\lfloor s \rfloor) \,\mathrm{d}s \Big)^{\frac{1}{2}} \\
&\quad \le \frac{p}{\sqrt{p-1}} L_{\mathrm{t},b}\sqrt{m} \big(1+\|X\|^2_{S^{(\gamma_b\vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_b}{2}} \frac{1}{\sqrt{2}}\sqrt{T-t_0}\sqrt{h},
\end{aligned} \tag{IV.54}$$

$$
\begin{aligned}
&\Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(\lfloor s \rfloor, X_s)) - b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}))\,\mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \leq \frac{p}{\sqrt{p-1}} \bigg(\int_{t_0}^{T} \Big\| \sum_{j=1}^{m} \| b^j(\mathcal{T}(\lfloor s \rfloor, X_s)) - b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \,\mathrm{d}s \bigg)^{\frac{1}{2}} \\
&\quad \leq \frac{p}{\sqrt{p-1}} L_b \sqrt{m} \bigg(\int_{t_0}^{T} \Big(\sum_{l=0}^{D} \| X_{s-\tau_l} - X_{\lfloor s \rfloor - \tau_l} \|_{L^p(\Omega;\mathbb{R})} \Big)^2 \,\mathrm{d}s \bigg)^{\frac{1}{2}} \\
&\quad \leq \frac{p}{\sqrt{p-1}} L_b \sqrt{m} \big(L_\xi D + C_1(D+1) \big) \bigg(\int_{t_0}^{T} (s - \lfloor s \rfloor)\,\mathrm{d}s \bigg)^{\frac{1}{2}} \\
&\quad \leq \frac{p}{\sqrt{p-1}} L_b \sqrt{m} \big(L_\xi D + C_1(D+1) \big) \frac{1}{\sqrt{2}} \sqrt{T - t_0} \sqrt{h},
\end{aligned}
\tag{IV.55}
$$

and

$$
\begin{aligned}
&\Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) - b^j(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor}))\,\mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \leq \frac{p}{\sqrt{p-1}} \bigg(\int_{t_0}^{T} \Big\| \sum_{j=1}^{m} \| b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) - b^j(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor})) \|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \,\mathrm{d}s \bigg)^{\frac{1}{2}} \\
&\quad \leq \frac{p}{\sqrt{p-1}} L_b \sqrt{m} \bigg(\int_{t_0}^{T} \| X - Y \|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s \bigg)^{\frac{1}{2}}.
\end{aligned}
\tag{IV.56}
$$

We insert inequalities (IV.52), (IV.32), (IV.53), (IV.54), (IV.55), and (IV.56) into inequality (IV.51), and thus, we have in total

$$
\begin{aligned}
&\| X - Y \|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \\
&\quad \leq \bigg(\frac{2L_{\mathrm{t},a}}{3} \big(1 + \| X \|^2_{S^{(\gamma_a \vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \big)^{\frac{\gamma_a}{2}} \sqrt{T - t_0} \\
&\qquad\quad + \frac{p L_{\mathrm{t},b} \sqrt{m}}{\sqrt{2}\sqrt{p-1}} \big(1 + \| X \|^2_{S^{(\gamma_b \vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \big)^{\frac{\gamma_b}{2}} \\
&\qquad\quad + \bigg(\frac{2L_a}{3} \sqrt{T - t_0} + \frac{p L_b \sqrt{m}}{\sqrt{2}\sqrt{p-1}} \bigg) \big(L_\xi D + C_1(D+1) \big) \bigg) \sqrt{T - t_0} \sqrt{h} \\
&\qquad + \bigg(L_a \sqrt{T - t_0} + \frac{p L_b \sqrt{m}}{\sqrt{p-1}} \bigg) \bigg(\int_{t_0}^{T} \| X - Y \|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \,\mathrm{d}s \bigg)^{\frac{1}{2}}.
\end{aligned}
\tag{IV.57}
$$

It follows by inequality (II.6) that

$$\begin{aligned}
&\|X-Y\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\\
&\quad\le 2\bigg(\frac{2L_{\mathrm{t},a}}{3}\big(1+\|X\|^2_{S^{(\gamma_a\vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_a}{2}}\sqrt{T-t_0}\\
&\qquad\qquad+\frac{pL_{\mathrm{t},b}\sqrt{m}}{\sqrt{2}\sqrt{p-1}}\big(1+\|X\|^2_{S^{(\gamma_b\vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_b}{2}}\\
&\qquad\qquad+\bigg(\frac{2L_a}{3}\sqrt{T-t_0}+\frac{pL_b\sqrt{m}}{\sqrt{2}\sqrt{p-1}}\bigg)\big(L_\xi D+C_1(D+1)\big)\bigg)^2(T-t_0)h\\
&\qquad+2\bigg(L_a\sqrt{T-t_0}+\frac{pL_b\sqrt{m}}{\sqrt{p-1}}\bigg)^2\int_{t_0}^{T}\|X-Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s,
\end{aligned}$$

and Gronwall's Lemma II.7 yields

$$\begin{aligned}
&\|X-Y\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\\
&\quad\le 2\bigg(\frac{2L_{\mathrm{t},a}}{3}\big(1+\|X\|^2_{S^{(\gamma_a\vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_a}{2}}\sqrt{T-t_0}\\
&\qquad\qquad+\frac{pL_{\mathrm{t},b}\sqrt{m}}{\sqrt{2}\sqrt{p-1}}\big(1+\|X\|^2_{S^{(\gamma_b\vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_b}{2}}\\
&\qquad\qquad+\bigg(\frac{2L_a}{3}\sqrt{T-t_0}+\frac{pL_b\sqrt{m}}{\sqrt{2}\sqrt{p-1}}\bigg)\big(L_\xi D+C_1(D+1)\big)\bigg)^2(T-t_0)\,h\\
&\qquad\times \mathrm{e}^{2\big(L_a\sqrt{T-t_0}+\frac{pL_b\sqrt{m}}{\sqrt{p-1}}\big)^2(T-t_0)}.
\end{aligned}\tag{IV.58}$$

Hence, there exists a constant $C_{\mathrm{Euler}}>0$ such that

$$\|X-Y^h\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\le C_{\mathrm{Euler}}\sqrt{h}$$

for all $h\in\,]0,T-t_0]$, and the family of Euler-Maruyama approximation $(Y^h)_{h\in]0,T-t_0]}$ converges in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ to solution X of SDDE (II.1) as $h\to 0$. □

Proof of Theorem IV.9

In order to show the convergence of the Milstein scheme, we have to ensure the boundedness of its moments first.

Lemma IV.16

Let the Borel-measurable coefficients of SDDE (II.1) *fulfill Assumption IV.8 iv), where* $b^j(t,t-\tau_1,\dots,t-\tau_D,\cdot,\dots,\cdot)\in C^1(\mathbb{R}^{d\times(D+1)};\mathbb{R}^d)$ *and its spatial partial derivatives are bounded by a constant* $L_b>0$ *for all* $t\in[t_0,T]$ *and* $j\in\{1,\dots,m\}$*. That is* $\big\|\partial_{x_l^i}b^j(t,t-\tau_1,\dots,t-\tau_D,x_0,x_1,\dots,x_D)\big\|\le L_b$ *for all* $t\in[t_0,T]$*,* $i,j\in\{1,\dots,d\}$*, and* $x_l\in\mathbb{R}^d$*, where* $l\in\{0,1,\dots,D\}$*. Further, let initial condition* ξ *belong to* $S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)$ *for some* $p\in[2,\infty[$*.*

Considering Milstein approximation Y *from formula* (IV.33) *regarding SDDE* (II.1)*, it holds*

$$\begin{aligned}
&1+\|Y\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\\
&\quad\le\big(1+2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\big)\,\mathrm{e}^{2\big(K_a\sqrt{T-t_0}+\frac{p}{\sqrt{p-1}}K_b\sqrt{m}+pL_bm\sqrt{d}K_b(D+1)\sqrt{T-t_0}\big)^2(T-t_0)}.
\end{aligned}$$

Proof. The proof is similar to proof of Lemma IV.15. Therefore, we only consider the part that changes here. Inside the brackets of the right-hand side of inequality (IV.48), the term

$$\sum_{l=0}^{D}\Big\|\sum_{j_1=1}^{m}\int_{t_0}^{\cdot}\sum_{i=1}^{d}\partial_{x_l^i}b^{j_1}(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\\ \times\sum_{j_2=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j_2}(\mathcal{T}((\lfloor s\rfloor-\tau_l)\vee t_0,Y_{(\lfloor s\rfloor-\tau_l)\vee t_0}))\,\mathrm{d}W_u^{j_2}\,\mathrm{d}W_s^{j_1}\Big\|_{S^p([t_0-\tau,t_n]\times\Omega;\mathbb{R}^d)}$$

has to be added. Using Theorem II.6 twice, the boundedness of the derivatives $\partial_{x_l^i}b^{j_1}$, Assumption IV.8 *iv)*, and the Cauchy-Schwarz inequality, we obtain, similarly to inequalities (II.26) and (IV.50), the estimate

$$\begin{aligned}
&\sum_{l=0}^{D}\Big\|\sum_{j_1=1}^{m}\int_{t_0}^{\cdot}\sum_{i=1}^{d}\partial_{x_l^i}b^{j_1}(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\\
&\qquad\times\sum_{j_2=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j_2}(\mathcal{T}((\lfloor s\rfloor-\tau_l)\vee t_0,Y_{(\lfloor s\rfloor-\tau_l)\vee t_0}))\,\mathrm{d}W_u^{j_2}\,\mathrm{d}W_s^{j_1}\Big\|_{S^p([t_0-\tau,t_n]\times\Omega;\mathbb{R}^d)}\\
&\le\frac{p}{\sqrt{p-1}}\sum_{l=0}^{D}\Bigg(\int_{t_0}^{t_n}\Big\|\sum_{j_1=1}^{m}\Big\|\sum_{i=1}^{d}\partial_{x_l^i}b^{j_1}(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\\
&\qquad\times\sum_{j_2=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j_2}(\mathcal{T}((\lfloor s\rfloor-\tau_l)\vee t_0,Y_{(\lfloor s\rfloor-\tau_l)\vee t_0}))\,\mathrm{d}W_u^{j_2}\Big\|^2\Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})}\mathrm{d}s\Bigg)^{\frac{1}{2}}\\
&\le\frac{pL_b\sqrt{m}\sqrt{d}}{\sqrt{p-1}}\sum_{l=0}^{D}\Bigg(\int_{t_0}^{t_n}\Big\|\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{j}(\mathcal{T}((\lfloor s\rfloor-\tau_l)\vee t_0,Y_{(\lfloor s\rfloor-\tau_l)\vee t_0}))\,\mathrm{d}W_u^{j}\Big\|^2_{L^p(\Omega;\mathbb{R}^d)}\mathrm{d}s\Bigg)^{\frac{1}{2}}\\
&\le pL_bm\sqrt{d}K_b(D+1)\sqrt{T-t_0}\Bigg(\int_{t_0}^{t_n}1+\|Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s\Bigg)^{\frac{1}{2}}.
\end{aligned}$$

We add this additional term to the right-hand side of inequality (IV.57), and then, similarly to inequality (IV.58), Gronwall's inequality yields the assertion. □

Proof of Theorem IV.9. Consider the difference of the SDDE's solution X and the corresponding Milstein approximation Y, that is

$$X_t-Y_t=\begin{cases}0\quad\text{if } t\in[t_0-\tau,t_0]\text{ and}\\
\displaystyle\int_{t_0}^{t}a(\mathcal{T}(s,X_s))-a(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\,\mathrm{d}s\\
\displaystyle\quad+\sum_{j=1}^{m}\int_{t_0}^{t}b^j(\mathcal{T}(s,X_s))-b^j(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\,\mathrm{d}W_s^j\\
\displaystyle\quad-\sum_{l=0}^{D}\sum_{j_1=1}^{m}\int_{t_0}^{t}\sum_{i=1}^{d}\partial_{x_l^i}b^{j_1}(\mathcal{T}(\lfloor s\rfloor,Y_{\lfloor s\rfloor}))\\
\displaystyle\qquad\times\sum_{j_2=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j_2}(\mathcal{T}((\lfloor s\rfloor-\tau_l)\vee t_0,Y_{(\lfloor s\rfloor-\tau_l)\vee t_0}))\,\mathrm{d}W_u^{j_2}\,\mathrm{d}W_s^{j_1}\\
\quad\text{if } t\in\,]t_0,T]\end{cases}\tag{IV.59}$$

for all $t \in [t_0 - \tau, T]$ P-almost surely. Inserting expansions (IV.37) and (IV.39) into formula (IV.59) and applying the triangle inequality, we have

$$\|X - Y\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq \sum_{r=1}^{14} \mathcal{R}_r \tag{IV.60}$$

where

$$\mathcal{R}_1 := \left\| \int_{t_0}^{\cdot} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) - a(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor}))\,\mathrm{d}s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},$$

$$\mathcal{R}_2 := \left\| \int_{t_0}^{\cdot} a(\mathcal{T}(s, X_s)) - a(\mathcal{T}(\lfloor s \rfloor, X_s))\,\mathrm{d}s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},$$

$$\mathcal{R}_3 := \sum_{l=1}^{D} \left\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}))(\xi^i_{(s-\tau_l)\wedge t_0} - \xi^i_{(\lfloor s \rfloor-\tau_l)\wedge t_0})\,\mathrm{d}s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},$$

$$\mathcal{R}_4 := \sum_{l=0}^{D} \left\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \int_{(\lfloor s \rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} a^i(\mathcal{T}(u, X_u))\,\mathrm{d}u\,\mathrm{d}s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},$$

$$\begin{aligned}\mathcal{R}_5 &:= \sum_{l=0}^{D} \mathcal{R}_5^l \\ &:= \sum_{l=0}^{D} \left\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u))\,\mathrm{d}W_u^j\,\mathrm{d}s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},\end{aligned} \tag{IV.61}$$

$$\begin{aligned}\mathcal{R}_6 := \sum_{l_1,l_2=0}^{D} \Bigg\| \int_{t_0}^{\cdot} \sum_{i_1,i_2=1}^{d} \int_0^1 \partial_{x_{l_1}^{i_1}} \partial_{x_{l_2}^{i_2}} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor} + \theta(X_s - X_{\lfloor s \rfloor})))(1-\theta)\,\mathrm{d}\theta \\ \times (X^{i_1}_{s-\tau_{l_1}} - X^{i_1}_{\lfloor s \rfloor-\tau_{l_1}})(X^{i_2}_{s-\tau_{l_2}} - X^{i_2}_{\lfloor s \rfloor-\tau_{l_2}})\,\mathrm{d}s \Bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},\end{aligned}$$

$$\mathcal{R}_7 := \left\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) - b^j(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor}))\,\mathrm{d}W_s^j \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},$$

$$\mathcal{R}_8 := \left\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} b^j(\mathcal{T}(s, X_s)) - b^j(\mathcal{T}(\lfloor s \rfloor, X_s))\,\mathrm{d}W_s^j \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},$$

$$\mathcal{R}_9 := \sum_{l=1}^{D} \left\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}))(\xi^i_{(s-\tau_l)\wedge t_0} - \xi^i_{(\lfloor s \rfloor-\tau_l)\wedge t_0})\,\mathrm{d}W_s^j \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},$$

$$\mathcal{R}_{10} \coloneqq \sum_{l=0}^{D} \Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} a^i(\mathcal{T}(u, X_u)) \, \mathrm{d}u \, \mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},$$

$$\begin{aligned}\mathcal{R}_{11} \coloneqq \sum_{l=0}^{D} \Big\| \sum_{j_1=1}^{m} \int_{t_0}^{\cdot} \sum_{j_2=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \\ \times \sum_{i=1}^{d} \Big(\partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_{(\lfloor s \rfloor - \tau_l)\vee t_0})) \\ - \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor})) b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, Y_{(\lfloor s \rfloor - \tau_l)\vee t_0})) \Big) \mathrm{d}W_s^{j_1} \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},\end{aligned}$$

$$\begin{aligned}\mathcal{R}_{12} \coloneqq \sum_{l=0}^{D} \Big\| \sum_{j_1=1}^{m} \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j_2=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \Big(b^{i,j_2}(\mathcal{T}(u, X_u)) \\ - b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_u)) \Big) \, \mathrm{d}W_u^{j_2} \, \mathrm{d}W_s^{j_1} \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},\end{aligned}$$

$$\begin{aligned}\mathcal{R}_{13} \coloneqq \sum_{l=0}^{D} \Big\| \sum_{j_1=1}^{m} \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) \sum_{j_2=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \Big(b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_u)) \\ - b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_{(\lfloor s \rfloor - \tau_l)\vee t_0})) \Big) \, \mathrm{d}W_u^{j_2} \, \mathrm{d}W_s^{j_1} \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},\end{aligned}$$

and

$$\begin{aligned}\mathcal{R}_{14} \coloneqq \Big\| \sum_{j=1}^{m} \int_{t_0}^{\cdot} \sum_{l_1,l_2=0}^{D} \sum_{i_1,i_2=1}^{d} \int_0^1 \partial_{x_{l_1}^{i_1}} \partial_{x_{l_2}^{i_2}} b^j(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor} + \theta(X_s - X_{\lfloor s \rfloor})))(1-\theta) \, \mathrm{d}\theta \\ \times (X_{s-\tau_{l_1}}^{i_1} - X_{\lfloor s \rfloor - \tau_{l_1}}^{i_1})(X_{s-\tau_{l_2}}^{i_2} - X_{\lfloor s \rfloor - \tau_{l_2}}^{i_2}) \, \mathrm{d}W_s^j \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}.\end{aligned}$$

In this proof, we estimate the terms $\mathcal{R}_r$, $r \in \{1, \dots, 14\}$, separately and show that there exist constants $C_1, C_2 > 0$ such that

$$\|X - Y\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \le C_1 h + C_2 \Big(\int_{t_0}^{T} \|X - Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \, \mathrm{d}s \Big)^{\frac{1}{2}}. \tag{IV.62}$$

Considering inequality (IV.62) as being satisfied and using inequality (II.6), we would obtain

$$\|X - Y\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \le 2C_1^2 h^2 + 2C_2^2 \int_{t_0}^{T} \|X - Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \, \mathrm{d}s, \tag{IV.63}$$

and the assertion of Theorem IV.9 follows by Gronwall's Lemma II.7.

In the following, we estimate the terms $\mathcal{R}_r$, $r \in \{1, \dots, 14\}$, from inequality (IV.60), where we proceed in lexicographical order.

In order to keep the overview, we provide the following Table IV.17. It contains the number and the page number of the inequality where the estimate of term $\mathcal{R}_r$, $r \in \{1, \dots, 14\}$, is stated as well as the information whether the term is bounded by a constant times h or a constant times $\big(\int_{t_0}^T \|X - Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}$. The latter information indicates whether term $\mathcal{R}_r$ contributes to constant C_1 or constant C_2 in inequality (IV.62).

The first term $\mathcal{R}_1$ is already estimated in inequality (IV.53), which is utilized the proof of convergence of the Euler-Maruyama scheme. It equally holds

$$\mathcal{R}_1 \leq L_a \sqrt{T - t_0} \bigg(\int_{t_0}^T \|X - Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)} \, \mathrm{d}s \bigg)^{\frac{1}{2}}. \tag{IV.64}$$

Similarly to inequality (IV.52), we obtain, using Assumption IV.8 *vi)* and

$$\begin{aligned} \int_{t_0}^T (s - \lfloor s \rfloor) \, \mathrm{d}s &= \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} (s - t_n) \, \mathrm{d}s = \sum_{n=0}^{N-1} \frac{1}{2} (t_{n+1} - t_n)^2 \\ &\leq \frac{1}{2} h \sum_{n=0}^{N-1} (t_{n+1} - t_n) = \frac{1}{2} (T - t_0) \, h, \end{aligned} \tag{IV.65}$$

Table IV.17. Consider the terms $\mathcal{R}_r$, $r \in \{1, \dots, 14\}$, of the right-hand side of inequality (IV.60). The number and the page number of the inequality stating the estimate of term $\mathcal{R}_r$ are presented. The last two columns represent whether term $\mathcal{R}_r$ is estimated by a constant times h or a constant times $\big(\int_{t_0}^T \|X - Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}$. The latter indicates whether term $\mathcal{R}_r$ contributes to constant C_1 or C_2 in inequality (IV.62).

	Inequality	Page	h	$\big(\int_{t_0}^T \|X - Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}$
$\mathcal{R}_1$	(IV.64)	89		✓
$\mathcal{R}_2$	(IV.66)	90	✓	
$\mathcal{R}_3$	(IV.68)	90	✓	
$\mathcal{R}_4$	(IV.69)	90	✓	
$\mathcal{R}_5^0$	(IV.79)	93	✓	
$\mathcal{R}_5^l$	(IV.145)	118	✓	
$\mathcal{R}_5$	(IV.146)	118	✓	
$\mathcal{R}_6$	(IV.153)	121	✓	
$\mathcal{R}_7$	(IV.154)	121		✓
$\mathcal{R}_8$	(IV.155)	122	✓	
$\mathcal{R}_9$	(IV.157)	122	✓	
$\mathcal{R}_{10}$	(IV.158)	122	✓	
$\mathcal{R}_{11}$	(IV.166)	125	✓	✓
$\mathcal{R}_{12}$	(IV.169)	126	✓	
$\mathcal{R}_{13}$	(IV.170)	127	✓	
$\mathcal{R}_{14}$	(IV.171)	127	✓	

that

$$\mathcal{R}_2 \le L_{\mathrm{t},a}\big(1 + \|X\|^2_{S^{(\gamma_a \vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_a}{2}} \frac{1}{2}(T - t_0)\, h. \tag{IV.66}$$

The estimate of term $\mathcal{R}_3$ makes use of the Lipschitz continuity of the drift coefficient a, see Assumption IV.8 *ii)*. Due to the Lipschitz continuity, we have $\|\partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}(\omega)))\| \le L_a$ for all $(s,\omega) \in [t_0,T]\times\Omega$. This inequality will be frequently used in the following, only with the reference to the Lipschitz condition in Assumption IV.8 *ii)*. Moreover, the Cauchy-Schwarz inequality implies the inequality

$$\sum_{i=1}^{d} |x^i| \le \sqrt{d}\|x\| \tag{IV.67}$$

for $x \in \mathbb{R}^d$ that is also often used below. Using the triangle inequality, Assumption IV.8 *ii)*, inequality (IV.67), Assumption IV.8 *vii)*, and inequality (IV.65), we obtain

$$\begin{aligned}
\mathcal{R}_3 &\le \sum_{l=1}^{D} \int_{t_0}^{T} \Big\| \sum_{i=1}^{d} \|\partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}))\| |\xi^i_{(s-\tau_l)\wedge t_0} - \xi^i_{(\lfloor s\rfloor-\tau_l)\wedge t_0}| \Big\|_{L^p(\Omega;\mathbb{R})} \mathrm{d}s \\
&\le L_a \sqrt{d} \sum_{l=1}^{D} \int_{t_0}^{T} \|\xi_{(s-\tau_l)\wedge t_0} - \xi_{(\lfloor s\rfloor-\tau_l)\wedge t_0}\|_{L^p(\Omega;\mathbb{R}^d)} \,\mathrm{d}s \\
&\le L_a \sqrt{d} L_\xi D \frac{1}{2}(T - t_0)\, h.
\end{aligned} \tag{IV.68}$$

Similarly to previous inequality (IV.68) and inequality (II.25), it holds

$$\begin{aligned}
\mathcal{R}_4 &\le L_a \sqrt{d} \sum_{l=0}^{D} \int_{t_0}^{T} \Big\| \int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} a(\mathcal{T}(u, X_u))\,\mathrm{d}u \Big\|_{L^p(\Omega;\mathbb{R}^d)} \mathrm{d}s \\
&\le L_a \sqrt{d} K_a (D+1)\big(1 + \|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \frac{1}{2}(T - t_0)\, h,
\end{aligned} \tag{IV.69}$$

where the linear growth condition in Assumption IV.8 *iv)* is used.

Now, we consider with term $\mathcal{R}_5$. As indicated in Section IV.2, a greater effort needs to be spend on the term $\mathcal{R}_5$ in order to estimate it properly. In the following, we consider term $\mathcal{R}_5^0$ and the terms $\mathcal{R}_5^l$ for $l \in \{1,\ldots,D\}$ separately. Starting with term $\mathcal{R}_5^0$, the triangle inequality initially implies

$$\begin{aligned}
\mathcal{R}_5^0 \le{}& \Big\| \sup_{t\in[t_0,T]} \Big\| \sum_{\substack{n=0\\ t_{n+1}\le t}}^{N-1} \int_{t_n}^{t_{n+1}} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_n, X_{t_n})) \sum_{j=1}^{m} \int_{t_n}^{s} b^{i,j}(\mathcal{T}(u, X_u))\,\mathrm{d}W_u^j \,\mathrm{d}s \Big\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&+ \Big\| \sup_{t\in[t_0,T]} \Big\| \int_{\lfloor t\rfloor}^{t} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(\lfloor t\rfloor, X_{\lfloor t\rfloor})) \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{i,j}(\mathcal{T}(u, X_u))\,\mathrm{d}W_u^j \,\mathrm{d}s \Big\| \Big\|_{L^p(\Omega;\mathbb{R})},
\end{aligned} \tag{IV.70}$$

cf. [78, Inequality (10.8.4)]. Consider the second term first. Using the triangle inequality, Assumption IV.8 *ii)*, inequality (IV.67), and the Cauchy-Schwarz inequality, we obtain, similarly to [78, Inequality (10.8.6)], that

$$\begin{aligned}
&\Big\| \sup_{t\in[t_0,T]} \Big\| \int_{\lfloor t\rfloor}^{t} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(\lfloor t\rfloor, X_{\lfloor t\rfloor})) \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{i,j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j\,\mathrm{d}s \Big\|\Big\|_{L^p(\Omega;\mathbb{R})} \\
&\le L_a\sqrt{d} \Big\| \sup_{t\in[t_0,T]} \int_{\lfloor t\rfloor}^{t} \Big\| \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j \Big\| \mathrm{d}s \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\le L_a\sqrt{d} \Big\| \sup_{t\in[t_0,T]} \Big(\int_{\lfloor t\rfloor}^{t} \Big\| \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j \Big\|^2 \mathrm{d}s \Big)^{\frac{1}{2}} \sqrt{t-\lfloor t\rfloor} \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\le L_a\sqrt{d} \Big\| \Big(\int_{t_0}^{T} \Big\| \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j \Big\|^2 \mathrm{d}s \Big)^{\frac{1}{2}} \Big\|_{L^p(\Omega;\mathbb{R})} \sqrt{h} \\
&= L_a\sqrt{d} \Big\| \int_{t_0}^{T} \Big\| \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j \Big\|^2 \mathrm{d}s \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})}^{\frac{1}{2}} \sqrt{h} \\
&\le L_a\sqrt{d} \Big(\int_{t_0}^{T} \Big\| \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j \Big\|_{L^p(\Omega;\mathbb{R}^d)}^2 \mathrm{d}s \Big)^{\frac{1}{2}} \sqrt{h}. \qquad \text{(IV.71)}
\end{aligned}$$

We consider the integrand of the integral over time, which is further estimated in the following. At first, Zakai's inequality from Theorem II.6 implies

$$\Big\| \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j \Big\|_{L^p(\Omega;\mathbb{R}^d)}^2 \le (p-1) \int_{\lfloor s\rfloor}^{s} \Big\| \sum_{j=1}^{m} \|b^{j}(\mathcal{T}(u,X_u))\|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \mathrm{d}u \qquad \text{(IV.72)}$$

for all $s \in [t_0, T]$. Taking the linear growth condition of diffusion coefficient b^j from Assumption IV.8 *iv)* into account, it holds

$$\begin{aligned}
\Big\| \sum_{j=1}^{m} \|b^{j}(\mathcal{T}(u,X_u))\|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} &\le K_b^2 m \Big\| \sup_{l\in\{0,1,\ldots,D\}} \big(1+\|X_{u-\tau_l}\|^2\big) \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \\
&\le K_b^2 m \Big\| 1 + \sup_{t\in[t_0-\tau,T]} \|X_t\|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \\
&\le K_b^2 m \Big(1 + \Big\| \sup_{t\in[t_0-\tau,T]} \|X_t\|^2 \Big\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \Big) \\
&= K_b^2 m \big(1 + \|X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}^2\big)
\end{aligned}$$

for all $u \in [t_0, T]$. Inserting this into inequality (IV.72), we obtain

$$\Big\| \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{j}(\mathcal{T}(u,X_u))\,\mathrm{d}W_u^j \Big\|_{L^p(\Omega;\mathbb{R}^d)}^2 \le (p-1) K_b^2 m \big(1 + \|X\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}^2\big)(s-\lfloor s\rfloor) \qquad \text{(IV.73)}$$

for all $s \in [t_0, T]$. Together with the inequalities (IV.71) and (IV.65), we get

$$\begin{aligned}
&\bigg\| \sup_{t\in[t_0,T]} \bigg\| \int_{\lfloor t\rfloor}^{t} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor})) \sum_{j=1}^{m} \int_{\lfloor s\rfloor}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg\| \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\leq L_a \sqrt{d}\sqrt{p-1} K_b \sqrt{m} \big(1 + \|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \bigg(\int_{t_0}^{T} s - \lfloor s\rfloor \,\mathrm{d}s \bigg)^{\frac{1}{2}} \sqrt{h} \\
&\leq L_a \sqrt{d}\sqrt{p-1} K_b \sqrt{m} \big(1 + \|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \frac{1}{\sqrt{2}} \sqrt{T-t_0}\, h,
\end{aligned} \tag{IV.74}$$

and thus, we found a desired estimate of order $\mathcal{O}(h)$ for the second term in inequality (IV.70).

Now, we continue with the first term in inequality (IV.70). At first, we have

$$\begin{aligned}
&\bigg\| \sup_{t\in[t_0,T]} \bigg\| \sum_{\substack{n=0\\ t_{n+1}\leq t}}^{N-1} \int_{t_n}^{t_{n+1}} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_n, X_{t_n})) \sum_{j=1}^{m} \int_{t_n}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg\| \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&= \bigg\| \sup_{n\in\{1,\ldots,N\}} \bigg\| \sum_{\nu=0}^{n-1} \int_{t_\nu}^{t_{\nu+1}} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_\nu, X_{t_\nu})) \sum_{j=1}^{m} \int_{t_\nu}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg\| \bigg\|_{L^p(\Omega;\mathbb{R})}.
\end{aligned} \tag{IV.75}$$

The time-discrete process

$$\bigg(\sum_{\nu=0}^{n-1} \int_{t_\nu}^{t_{\nu+1}} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_\nu, X_{t_\nu})) \sum_{j=1}^{m} \int_{t_\nu}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg)_{n\in\{1,\ldots,N\}}$$

is discrete martingale in $L^p(\Omega;\mathbb{R}^d)$ with respect to filtration $(\mathscr{F}_{t_n})_{n\in\{1,\ldots,N\}}$ as each summand is $\mathscr{F}_{t_{\nu+1}}/\mathscr{B}(\mathbb{R}^d)$-measurable and as

$$\begin{aligned}
&\mathrm{E}\bigg[\int_{t_\nu}^{t_{\nu+1}} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_\nu, X_{t_\nu})) \sum_{j=1}^{m} \int_{t_\nu}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg| \mathscr{F}_{t_\nu} \bigg] \\
&= \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_\nu, X_{t_\nu})) \mathrm{E}\bigg[\int_{t_\nu}^{t_{\nu+1}} \sum_{j=1}^{m} \int_{t_\nu}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg| \mathscr{F}_{t_\nu} \bigg] \\
&= 0
\end{aligned}$$

P-almost surely for all $\nu \in \{0, 1, \ldots, N-1\}$ by the stochastic integration by parts formula based on Itô's formula, cf. [64, Equation (3.13)], and properties of the stochastic integral. Thus, the discrete Burkholder-type inequality in Theorem II.5 applies to the right-hand side of equation (IV.75), and together with the triangle inequality, we obtain

$$\begin{aligned}
&\bigg\| \sup_{n\in\{1,\ldots,N\}} \bigg\| \sum_{\nu=0}^{n-1} \int_{t_\nu}^{t_{\nu+1}} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_\nu, X_{t_\nu})) \sum_{j=1}^{m} \int_{t_\nu}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg\| \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\leq \frac{p}{\sqrt{p-1}} \bigg(\sum_{\nu=0}^{N-1} \bigg\| \int_{t_\nu}^{t_{\nu+1}} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_\nu, X_{t_\nu})) \sum_{j=1}^{m} \int_{t_\nu}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg\|^2_{L^p(\Omega;\mathbb{R}^d)} \bigg)^{\frac{1}{2}}.
\end{aligned} \tag{IV.76}$$

Next, we consider the $L^p(\Omega;\mathbb{R}^d)$-norms of the summands from the last inequality (IV.76) only. Using the triangle inequality, the Lipschitz continuity of the drift coefficient from Assumption IV.8 *ii)*, inequality (IV.67), and the square root of inequality (IV.73), we obtain

$$\begin{aligned}
&\Big\| \int_{t_\nu}^{t_{\nu+1}} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_\nu, X_{t_\nu})) \sum_{j=1}^{m} \int_{t_\nu}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j \,\mathrm{d}s \Big\|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad\leq L_a\sqrt{d} \int_{t_\nu}^{t_{\nu+1}} \Big\| \sum_{j=1}^{m} \int_{t_\nu}^{s} b^j(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j \Big\|_{L^p(\Omega;\mathbb{R}^d)} \mathrm{d}s \\
&\quad\leq L_a\sqrt{d}\sqrt{p-1}K_b\sqrt{m}\big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \int_{t_\nu}^{t_{\nu+1}} \sqrt{s-t_\nu}\,\mathrm{d}s \\
&\quad= L_a\sqrt{d}\sqrt{p-1}K_b\sqrt{m}\big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \frac{2}{3}(t_{\nu+1}-t_\nu)^{\frac{3}{2}}. \qquad \text{(IV.77)}
\end{aligned}$$

Inserting this into inequality (IV.76) yields

$$\begin{aligned}
&\Big\| \sup_{n\in\{1,\dots,N\}} \Big\| \sum_{\nu=0}^{n-1} \int_{t_\nu}^{t_{\nu+1}} \sum_{i=1}^{d} \partial_{x_0^i} a(\mathcal{T}(t_\nu, X_{t_\nu})) \sum_{j=1}^{m} \int_{t_\nu}^{s} b^{i,j}(\mathcal{T}(u, X_u))\, \mathrm{d}W_u^j \,\mathrm{d}s \Big\|\Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \frac{p}{\sqrt{p-1}} L_a\sqrt{d}\sqrt{p-1}K_b\sqrt{m}\big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \frac{2}{3}\Big(\sum_{\nu=0}^{N-1}(t_{\nu+1}-t_\nu)^3\Big)^{\frac{1}{2}} \\
&\quad\leq \frac{p}{\sqrt{p-1}} L_a\sqrt{d}\sqrt{p-1}K_b\sqrt{m}\big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \frac{2}{3}\sqrt{T-t_0}\,h, \qquad \text{(IV.78)}
\end{aligned}$$

and together with inequality (IV.74), it follows that

$$\mathcal{R}_5^0 \leq L_a\sqrt{d}\sqrt{p-1}K_b\sqrt{m}\big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\Big(\frac{2}{3}\frac{p}{\sqrt{p-1}}+\frac{1}{\sqrt{2}}\Big)\sqrt{T-t_0}\,h. \qquad \text{(IV.79)}$$

Now, let us consider term $\mathcal{R}_5^l$ for $l\in\{1,\dots,D\}$. In order to overcome the problem of the missing discrete martingale property, the Malliavin calculus is used.

Recall that initial condition $\xi\colon [t_0-\tau,t_0]\times\Omega\to\mathbb{R}^d$ is a stochastic process, where ξ_t is $\mathscr{F}_{t_0}/\mathscr{B}(\mathbb{R}^d)$-measurable for all $t\in[t_0-\tau,t_0]$ in particular. Since the increments $W_t-W_{t_0}$, $t\in[t_0,T]$, of the Wiener process are independent of $\mathscr{F}_{t_0}$, process ξ is independent of sub-σ-algebra $\mathscr{G}\subseteq\mathscr{F}$, see equation (II.2) where $\mathscr{G}:=\mathscr{G}_T$ and cf. Chapter III. But this just means that ξ_t is not $\mathscr{G}/\mathscr{B}(\mathbb{R}^d)$-measurable for any $t\in[t_0-\tau,t_0]$ in general. Indeed, consider the preimage $A=\xi_t^{-1}(B)$ for any $B\in\mathscr{B}(\mathbb{R}^d)$ and $t\in[t_0-\tau,t_0]$. Since ξ_t in independent of $\mathscr{G}$, it holds

$$\mathrm{P}[A\cap C]=\mathrm{P}[A]\,\mathrm{P}[C] \qquad \text{(IV.80)}$$

for all $C\in\mathscr{G}$. Assume for a moment that ξ_t is $\mathscr{G}/\mathscr{B}(\mathbb{R}^d)$-measurable. Then, we would have $A\in\mathscr{G}$ and equation (IV.80) would imply

$$\mathrm{P}[A]=\big(\mathrm{P}[A]\big)^2 \qquad \text{(IV.81)}$$

for all preimages $A=\xi_t^{-1}(B)$, where $B\in\mathscr{B}(\mathbb{R}^d)$ and $t\in[t_0-\tau,t_0]$. However, equation (IV.81) only holds true if $\mathrm{P}[A]=0$ or $\mathrm{P}[A]=1$. But this would mean that ξ is just a modification of a deterministic process, which is a contradiction to our assumption on initial condition ξ, see Assumption IV.8 *vii)*.

Due to these considerations, solution X of SDDE (II.1) with initial condition ξ is also not $\mathscr{B}([t_0-\tau,T])\otimes\mathscr{G}/\mathscr{B}(\mathbb{R}^d)$-measurable in general. Thus, X_t does not belong to the space $\mathscr{D}^p(\Omega;\mathbb{R}^d)$ for any $t\in[t_0,T]$, and the Malliavin calculus cannot be used directly.

In order to make techniques from the Malliavin calculus applicable, we use [107, Lemma III.1.3], which we here represent in case of SDDEs.

Lemma IV.18 (Cf. [107, Lemma III.1.3])
Let the Borel-measurable coefficients a and b^j, $j\in\{1,\dots,m\}$, of SDDE (II.1) *fulfill Assumption IV.8 ii) and Assumption IV.8 iv). Consider the solution X^ζ of SDDE* (II.1) *with respect to initial condition ζ that belongs to $S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)$ for some $p\in[2,\infty[$. Further, let initial condition ζ be a simple random variable of the form*

$$\zeta=\sum_{k=1}^{K} z_k \mathbb{1}_{A_k},$$

where $z_k\in C([t_0-\tau,t_0];\mathbb{R}^d)$, $A_k\in\mathscr{F}_{t_0}$ with $A_k\cap A_l=\emptyset$ for $k\neq l$, $\bigcup_{k=1}^K A_k=\Omega$, and $K\in\mathbb{N}$.

Then, there exists an $\mathcal{N}\in\mathscr{F}_{t_0}$ with $\mathrm{P}[\mathcal{N}]=0$ such that

$$X_t^\zeta(\omega)=\sum_{k=1}^{K} X_t^{z_k}(\omega)\mathbb{1}_{A_k}(\omega)$$

for all $(t,\omega)\in[t_0-\tau,T]\times(\Omega\setminus\mathcal{N})$, where X^{z_k} is the unique strong solution of SDDE (II.1) *regarding initial condition z_k.*

We remark at this point that the solution of SDDE (II.1) is in general, opposed to SODEs, not linear with respect to its initial condition, cf. [108].

To be able to use Lemma IV.18, the initial condition has to be a simple random variable that takes values in the space of continuous functions. In the following, we describe how to approximate initial condition ξ by a sequence of such random variables on whose corresponding solutions we then can apply Lemma IV.18.

Consider initial condition ξ from Assumption IV.8 *vii)*. As P-almost all realizations of ξ are continuous, there exists an $\mathcal{N}\in\mathscr{F}_{t_0}$ with $\mathrm{P}[\mathcal{N}]=0$ such that $\xi(\omega)$ is continuous for all $\omega\in\Omega\setminus\mathcal{N}$. Let us define a process $\tilde{\xi}\in S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)$ by

$$\tilde{\xi}_t(\omega):=\begin{cases}\xi_t(\omega) & \text{if } (t,\omega)\in[t_0-\tau,t_0]\times(\Omega\setminus\mathcal{N}) \text{ and}\\ 0 & \text{if } (t,\omega)\in[t_0-\tau,t_0]\times\mathcal{N}.\end{cases} \tag{IV.82}$$

Then, the processes ξ and $\tilde{\xi}$ are indistinguishable, and it holds $\tilde{\xi}(\omega)\in C([t_0-\tau,t_0];\mathbb{R}^d)$ for all $\omega\in\Omega$. Moreover, the solutions X^ξ and $X^{\tilde{\xi}}$ of SDDE (II.1) with respect to their initial conditions ξ and $\tilde{\xi}$, respectively, are also indistinguishable. This follows from Markov's inequality and Lemma II.10.

Since all realizations of $\tilde{\xi}$ are continuous, there exists a sequence $(\zeta^r)_{r\in\mathbb{N}}$ of simple random variables $\zeta^r = \sum_{k=1}^{K_r} z_k^r \mathbb{1}_{A_k^r}$ where $z_k^r \in C([t_0-\tau,t_0];\mathbb{R}^d)$, $A_k^r \in \mathscr{F}_{t_0}$ with $A_k^r \cap A_l^r = \emptyset$ for $k \neq l$, $\bigcup_{k=1}^{K} A_k^r = \Omega$, and $K_r \in \mathbb{N}$ such that

$$\sup_{t\in[t_0-\tau,t_0]} \|\zeta_t^r(\omega)\| \leq \sup_{t\in[t_0-\tau,t_0]} \|\tilde{\xi}_t(\omega)\| \tag{IV.83}$$

and

$$\lim_{r\to\infty} \sup_{t\in[t_0-\tau,t_0]} \|\zeta_t^r(\omega) - \tilde{\xi}_t(\omega)\| = 0 \tag{IV.84}$$

for all $\omega \in \Omega$. This can be seen as follows.

For $t \in [t_0-\tau,t_0]$, define the projections $\pi_t\colon (\mathbb{R}^d)^{[t_0-\tau,t_0]} \to \mathbb{R}^d$ by $\pi_t(x) = x(t)$ for all maps $x\colon [t_0-\tau,t_0] \to \mathbb{R}^d$. Since $\tilde{\xi}$ is $\mathscr{B}([t_0-\tau,t_0]) \otimes \mathscr{F}_{t_0}/\mathscr{B}(\mathbb{R}^d)$-measurable in particular, random variable $\pi_t(\tilde{\xi}) = \tilde{\xi}_t$ is $\mathscr{F}_{t_0}/\mathscr{B}(\mathbb{R}^d)$-measurable for all $t \in [t_0-\tau,t_0]$. Then, [43, Satz 1.3.4] implies that $\tilde{\xi}$ can be viewed as an $\mathscr{F}_{t_0}/\sigma(\{\pi_t : t \in [t_0-\tau,t_0]\})$-measurable random variable. Further, it holds for the Borel-σ-algebra of $C([t_0-\tau,t_0];\mathbb{R}^d)$ that

$$\mathscr{B}(C([t_0-\tau,t_0];\mathbb{R}^d)) = C([t_0-\tau,t_0];\mathbb{R}^d) \cap \sigma(\{\pi_t : t \in [t_0-\tau,t_0]\}), \tag{IV.85}$$

see e. g. [43, Bemerkung 7.2.8 (e)] or [53, Beispiel 1.24]. As $\tilde{\xi}(\omega) \in C([t_0-\tau,t_0];\mathbb{R}^d)$ for all $\omega \in \Omega$, the stochastic process $\tilde{\xi}$ can be viewed as an $\mathscr{F}_{t_0}/\mathscr{B}(C([t_0-\tau,t_0];\mathbb{R}^d))$-measurable variable by equation (IV.85).

Then, the existence of such sequence $(\zeta^r)_{r\in\mathbb{N}}$ follows from [61, Corollary 1.1.7]. Since all realizations of ζ^r are continuous, ζ^r is $\mathscr{B}([t_0-\tau,t_0]) \otimes \mathscr{F}_{t_0}/\mathscr{B}(\mathbb{R}^d)$-measurable for all $r \in \mathbb{N}$, see [75, Proposition 1.1.13]. Using these considerations and Lemma IV.18, we can show the following lemma.

Lemma IV.19
Let the Borel-measurable coefficients a and b^j, $j \in \{1,\dots,m\}$, of SDDE (II.1) *fulfill Assumption IV.8 ii) and Assumption IV.8 iv), where $a(t, t-\tau_1,\dots,t-\tau_D,\cdot,\dots,\cdot) \in C^1(\mathbb{R}^{d\times(D+1)};\mathbb{R}^d)$ for all $t \in [t_0,T]$. Further, let initial condition ξ belong to $S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)$ for some $p \in [2,\infty[$ and* P*-almost surely have continuous realizations.*

Then, there exists a sequence $(\zeta^r)_{r\in\mathbb{N}}$ of simple random variables $\zeta^r = \sum_{k=1}^{K_r} z_k^r \mathbb{1}_{A_k^r}$, where $z_k^r \in C([t_0-\tau,t_0];\mathbb{R}^d)$, $A_k^r \in \mathscr{F}_{t_0}$ with $A_k^r \cap A_l^r = \emptyset$ for $k \neq l$, $\bigcup_{k=1}^{K_r} A_k^r = \Omega$, and $K_r \in \mathbb{N}$ such that formulas (IV.83) *and* (IV.84) *are fulfilled. Further, it holds*

$$\begin{aligned}\mathcal{R}_5^l = \lim_{r\to\infty} \Bigg\| \sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} \Bigg\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}^{z_k^r})) \\ \times \sum_{j=1}^{m} \int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j\,\mathrm{d}s \Bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \Bigg\|_{L^p(\Omega;\mathbb{R})}\end{aligned} \tag{IV.86}$$

for $l \in \{1,\dots,D\}$.

Proof. Consider initial condition $\tilde{\xi}$ defined in equation (IV.82). Since the stochastic processes ξ and $\tilde{\xi}$ are indistinguishable, the solutions X^{ξ} and $X^{\tilde{\xi}}$ are also indistinguishable as mentioned above. Thus, we can write

$$\mathcal{R}_5^l = \left\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{\tilde{\xi}})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j \, \mathrm{d}s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}, \tag{IV.87}$$

where we switched X^{ξ} and $X^{\tilde{\xi}}$ compared to formula (IV.61). In the following, we show

$$\mathcal{R}_5^l = \lim_{r\to\infty} \left\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{\zeta^r})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\zeta^r}))\, \mathrm{d}W_u^j \, \mathrm{d}s \right\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}, \tag{IV.88}$$

where sequence $(\zeta^r)_{r\in\mathbb{N}}$ is specified in the statement of this lemma. The existence of such a sequence $(\zeta^r)_{r\in\mathbb{N}}$ has already been discussed prior to this lemma. Considering the difference of the arguments in the $S^p([t_0,T]\times\Omega;\mathbb{R}^d)$-norms in equation (IV.87) and (IV.88), it holds

$$\begin{aligned}
&\bigg\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{\zeta^r})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\zeta^r}))\, \mathrm{d}W_u^j \\
&\qquad - \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{\tilde{\xi}})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j \, \mathrm{d}s \bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&= \bigg\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{\zeta^r})) \\
&\qquad\qquad \times \left(\sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\zeta^r}))\, \mathrm{d}W_u^j - \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j \right) \\
&\qquad + \sum_{i=1}^{d} \left(\partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{\zeta^r})) - \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{\tilde{\xi}})) \right) \\
&\qquad\qquad \times \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j \, \mathrm{d}s \bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)},
\end{aligned} \tag{IV.89}$$

where we subtracted and added the term

$$\sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{\zeta^r})) \sum_{j=1}^{m} \int_{(t_n-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j.$$

Thence, equation (IV.88) follows if the left-hand side of equation (IV.89) converges to zero as $r \to \infty$.

Applying the triangle inequality to the right-hand side of equation (IV.89), using the global

Lipschitz continuity of the drift coefficient and using inequality (IV.67), we have

$$\begin{aligned}
&\Big\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{\zeta^r})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\zeta^r}))\, \mathrm{d}W_u^j \\
&\quad - \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{\tilde{\xi}})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j \,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\leq L_a \sqrt{d} \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \Big\| \sum_{j=1}^{m} \int_{(t_n-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^j(\mathcal{T}(u, X_u^{\zeta^r})) - b^j(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j \Big\|_{L^p(\Omega;\mathbb{R}^d)} \mathrm{d}s \\
&\quad + \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \Big\| \sum_{i=1}^{d} \| \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{\zeta^r})) - \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{\tilde{\xi}})) \| \\
&\qquad \times \Big| \sum_{j=1}^{m} \int_{(t_n-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j \Big| \Big\|_{L^p(\Omega;\mathbb{R})} \mathrm{d}s.
\end{aligned} \tag{IV.90}$$

In order to show convergence to zero of the right-hand side of inequality (IV.90) above as $r \to \infty$, we need the following considerations.

According to inequality (IV.83) and the convergence in formula (IV.84), the dominated convergence theorem implies

$$\lim_{r\to\infty} \|\zeta^r - \tilde{\xi}\|_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} = 0.$$

Then, using Lemma II.10, we obtain

$$\lim_{r\to\infty} \|X^{\zeta^r} - X^{\tilde{\xi}}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = 0. \tag{IV.91}$$

Due to the convergence in equation (IV.91) and the Lipschitz continuity of b^j from Assumption IV.8 *ii)*, it follows

$$\lim_{r\to\infty} \max_{j\in\{1,\ldots,m\}} \|b^j(\mathcal{T}(\cdot, X_\cdot^{\zeta^r})) - b^j(\mathcal{T}(\cdot, X_\cdot^{\tilde{\xi}}))\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} = 0. \tag{IV.92}$$

Further, we show

$$\begin{aligned}
&\lim_{r\to\infty} \Big\| \sum_{i=1}^{d} \| \partial_{x_l^i} a(\mathcal{T}(t, X_t^{\zeta^r})) - \partial_{x_l^i} a(\mathcal{T}(t, X_t^{\tilde{\xi}})) \| \Big| \sum_{j=1}^{m} \int_{(\lfloor s \rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j \Big| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&= 0
\end{aligned} \tag{IV.93}$$

for all $s \in [t_0, T]$ in the sequel. The convergence in equation (IV.91) implies

$$\sup_{t\in[t_0-\tau,T]} \|X_t^{\zeta^r} - X_t^{\tilde{\xi}}\| \to 0$$

in probability as $r \to \infty$, that is

$$\lim_{r\to\infty} \mathrm{P}\Big[\Big\{\omega \in \Omega : \sup_{t\in[t_0-\tau,T]} \|X_t^{\zeta^r}(\omega) - X_t^{\tilde{\xi}}(\omega)\| > \varepsilon\Big\}\Big] = 0$$

for all $\varepsilon > 0$, [67, Theorem 17.2]. By the continuity of partial derivative $\partial_{x_l^i} a$, we obtain

$$\lim_{r\to\infty} \mathrm{P}\Big[\Big\{\omega \in \Omega : \|\partial_{x_l^i} a(\mathcal{T}(t, X_t^{\zeta^r}(\omega))) - \partial_{x_l^i} a(\mathcal{T}(t, X_t^{\tilde{\xi}}(\omega)))\| > \varepsilon\Big\}\Big] = 0$$

for all $\varepsilon > 0$ and all $t \in [t_0, T]$, see [67, Theorem 17.5]. Further, it even holds

$$\begin{aligned}\lim_{r\to\infty} \mathrm{P}\Big[\Big\{\omega \in \Omega : &\sum_{i=1}^{d} \|\partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}^{\zeta^r}(\omega))) - \partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}^{\tilde{\xi}}(\omega)))\| \\ &\times \Big|\sum_{j=1}^{m} \int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j\Big|(\omega) > \varepsilon\Big\}\Big] = 0\end{aligned}$$

for all $\varepsilon > 0$ and all $s \in [t_0, T]$. If we have, in addition, the uniformly integrability to the power of p of this sequence that converges to zero in probability, we obtain by Vitali's convergence theorem, see e. g. [38, p. 262] or [74, Proposition 4.12], the convergence in equation (IV.93). Thus, we show the uniformly integrability in the following. The Lipschitz continuity of drift coefficient a yields

$$\|\partial_{x_l^i} a(\mathcal{T}(t, X_t^{\zeta^r}(\omega))) - \partial_{x_l^i} a(\mathcal{T}(t, X_t^{\tilde{\xi}}(\omega)))\| \le 2L_a \tag{IV.94}$$

for all $(t, \omega) \in [t_0, T] \times \Omega$. Moreover, analogously to inequality (IV.73), Theorem II.6, the linear growth of the diffusion coefficients, see Assumption IV.8 *iv)*, and Theorem II.8 imply

$$\begin{aligned}&\Big\|\sum_{j=1}^{m} \int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j\Big\|_{L^p(\Omega;\mathbb{R})} \\ &\quad\le \sqrt{p-1} K_b \sqrt{m}\big(1 + \|X^{\tilde{\xi}}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} (T - t_0) \\ &\quad< \infty\end{aligned}$$

for all $s \in [t_0, T]$. According to this and inequality (IV.94), we obtain the uniform boundedness

$$\begin{aligned}\sup_{r\in\mathbb{N}} \sup_{s\in[t_0,T]} \Big\| &\sum_{i=1}^{d} \|\partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}^{\zeta^r})) - \partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}^{\tilde{\xi}}))\| \\ &\times \Big|\sum_{j=1}^{m} \int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j\Big|\Big\|_{L^p(\Omega;\mathbb{R})} < \infty,\end{aligned} \tag{IV.95}$$

and hence, we have the uniform integrability of

$$\begin{aligned}\Big(\Big(&\sum_{i=1}^{d} \|\partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}^{\zeta^r})) - \partial_{x_l^i} a(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}^{\tilde{\xi}}))\| \\ &\times \Big|\sum_{j=1}^{m} \int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{\tilde{\xi}}))\, \mathrm{d}W_u^j\Big|\Big)^p\Big)_{r\in\mathbb{N}}\end{aligned}$$

for all $s \in [t_0, T]$. Then, the convergence in equation (IV.93) follows by Vitali's convergence theorem.

Next, we consider the limit of the right-hand side in inequality (IV.90). Using Theorem II.6 and equation (IV.92), we infer for the limit of the first term on the right-hand side of inequality (IV.90) that

$$\begin{aligned}
&\lim_{r\to\infty} L_a\sqrt{d}\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\Big\|\sum_{j=1}^{m}\int_{(t_n-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^j(\mathcal{T}(u,X_u^{\zeta^r}))-b^j(\mathcal{T}(u,X_u^{\tilde{\xi}}))\,\mathrm{d}W_u^j\Big\|_{L^p(\Omega;\mathbb{R}^d)}\mathrm{d}s\\
&\le L_a\sqrt{d}\sqrt{p-1}\\
&\quad\sum_{n=0}^{N-1}\lim_{r\to\infty}\int_{t_n}^{t_{n+1}}\bigg(\int_{(t_n-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}\sum_{j=1}^{m}\|b^j(\mathcal{T}(u,X_u^{\zeta^r}))-b^j(\mathcal{T}(u,X_u^{\tilde{\xi}}))\|^2_{L^p(\Omega;\mathbb{R}^d)}\,\mathrm{d}u\bigg)^{\frac{1}{2}}\mathrm{d}s\\
&\le L_a\sqrt{d}\sqrt{p-1}\sqrt{m}(T-t_0)^{\frac{3}{2}}\lim_{r\to\infty}\max_{j\in\{1,\ldots,m\}}\|b^j(\mathcal{T}(\cdot,X_\cdot^{\zeta^r}))-b^j(\mathcal{T}(\cdot,X_\cdot^{\tilde{\xi}}))\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}\\
&=0. \qquad\qquad \text{(IV.96)}
\end{aligned}$$

Further, using the uniform boundedness in formula (IV.95), the dominated convergence theorem and equation (IV.93) imply for the limit of the second term in inequality (IV.90) that

$$\begin{aligned}
&\lim_{r\to\infty}\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\Big\|\sum_{i=1}^{d}\Big\|\partial_{x_l^i}a(\mathcal{T}(t_n,X_{t_n}^{\zeta^r}))-\partial_{x_l^i}a(\mathcal{T}(t_n,X_{t_n}^{\tilde{\xi}}))\Big\|\\
&\qquad\times\Big|\sum_{j=1}^{m}\int_{(t_n-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u,X_u^{\tilde{\xi}}))\,\mathrm{d}W_u^j\Big|\Big\|_{L^p(\Omega;\mathbb{R})}\mathrm{d}s \qquad \text{(IV.97)}\\
&=0.
\end{aligned}$$

According to the convergence in formulas (IV.96) and (IV.97), the right-hand side of inequality (IV.90) converges to zero as $r\to\infty$, and hence, equation (IV.88) holds true.

We now show the equivalence of formulas (IV.86) and (IV.88) by applying Lemma IV.18 to solution X^{ζ^r} in formula (IV.88). In the following, we frequently use that

$$f\bigg(\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}x_k^r\bigg)=\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}f(x_k^r) \qquad \text{(IV.98)}$$

for functions f, where x_k^r is in the domain of f, cf. [107, p. 50]. Using formula (IV.98), Lemma IV.18 implies

$$\partial_{x_l^i}a(\mathcal{T}(t_n,X_{t_n}^{\zeta^r}))=\sum_{k=1}^{K_r}\partial_{x_l^i}a(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))\,\mathbb{1}_{A_k^r} \qquad \text{(IV.99)}$$

for $n\in\{0,1,\ldots,N\}$ P-almost surely and

$$\sum_{j=1}^{m}\int_{(t_n-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u,X_u^{\zeta^r}))\,\mathrm{d}W_u^j=\sum_{k=1}^{K_r}\sum_{j=1}^{m}\int_{(t_n-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\,\mathrm{d}W_u^j\,\mathbb{1}_{A_k^r} \qquad \text{(IV.100)}$$

for all $s\in[t_n,T]$ with $n\in\{0,1,\ldots,N\}$ P-almost surely. Inserting equations (IV.99) and (IV.100) into formula (IV.88) and using that $\mathbb{1}_{A_k^r}\cdot\mathbb{1}_{A_l^r}=0$ if $k\neq l$, we further obtain for the Euclidean norm of the argument of the $S^p([t_0,T]\times\Omega;\mathbb{R}^d)$-norm in formula (IV.88) by formula (IV.98) with $f(\cdot)=\|\cdot\|$ that

$$\begin{aligned}&\Big\|\int_{t_0}^{t}\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}^{\zeta^r}))\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u^{\zeta^r}))\,\mathrm{d}W_u^j\,\mathrm{d}s\Big\|\\&=\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\Big\|\int_{t_0}^{t}\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}^{z_k^r}))\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\,\mathrm{d}W_u^j\,\mathrm{d}s\Big\|\end{aligned}\tag{IV.101}$$

for all $t\in[t_0,T]$ P-almost surely. Since $z_k^r\in C([t_0-\tau,t_0];\mathbb{R}^d)$ is deterministic, the solution $X^{z_k^r}$ and thus also the Euclidean norm on the right-hand side of equation (IV.101) above are independent of $\mathscr{F}_{t_0}$. Moreover, random variable $\mathbb{1}_{A_k^r}$ is an $\mathscr{F}_{t_0}$-measurable as $A_k^r\in\mathscr{F}_{t_0}$. We now insert equation (IV.101) into formula (IV.88). Using formula (IV.98), we infer by properties of the conditional expectation that

$$\begin{aligned}\mathcal{R}_5^l&=\lim_{r\to\infty}\Bigg(\mathrm{E}\Big[\mathrm{E}\Big[\sup_{t\in[t_0,T]}\Big\|\int_{t_0}^{t}\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}^{\zeta^r}))\\&\qquad\times\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u^{\zeta^r}))\,\mathrm{d}W_u^j\,\mathrm{d}s\Big\|^p\Big|\mathscr{F}_{t_0}\Big]\Big]\Bigg)^{\frac{1}{p}}\\&=\lim_{r\to\infty}\Bigg(\mathrm{E}\Big[\mathrm{E}\Big[\sup_{t\in[t_0,T]}\Big(\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\Big\|\int_{t_0}^{t}\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}^{z_k^r}))\\&\qquad\times\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\,\mathrm{d}W_u^j\,\mathrm{d}s\Big\|\Big)^p\Big|\mathscr{F}_{t_0}\Big]\Big]\Bigg)^{\frac{1}{p}}\\&=\lim_{r\to\infty}\Bigg(\mathrm{E}\Big[\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\mathrm{E}\Big[\sup_{t\in[t_0,T]}\Big\|\int_{t_0}^{t}\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}^{z_k^r}))\\&\qquad\times\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\,\mathrm{d}W_u^j\,\mathrm{d}s\Big\|^p\Big|\mathscr{F}_{t_0}\Big]\Big]\Bigg)^{\frac{1}{p}}\\&=\lim_{r\to\infty}\Bigg(\mathrm{E}\Big[\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\mathrm{E}\Big[\sup_{t\in[t_0,T]}\Big\|\int_{t_0}^{t}\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}^{z_k^r}))\\&\qquad\times\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\,\mathrm{d}W_u^j\,\mathrm{d}s\Big\|^p\Big]\Big]\Bigg)^{\frac{1}{p}}.\end{aligned}\tag{IV.102}$$

The inner expectation on the right-hand side of equation (IV.102) above is just the $S^p([t_0,T]\times\Omega;\mathbb{R}^d)$-norm to the power of p, and thus, we have

$$\begin{aligned}\mathcal{R}_5^l&=\lim_{r\to\infty}\Bigg(\mathrm{E}\Big[\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\Big\|\int_{t_0}^{\cdot}\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}^{z_k^r}))\\&\qquad\times\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\,\mathrm{d}W_u^j\,\mathrm{d}s\Big\|^p_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}\Big]\Bigg)^{\frac{1}{p}}.\end{aligned}\tag{IV.103}$$

Using further that

$$\Bigg(\mathrm{E}\Bigg[\sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} c_k^p\Bigg]\Bigg)^{\frac{1}{p}} = \Bigg(\mathrm{E}\Bigg[\Bigg(\sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} c_k\Bigg)^p\Bigg]\Bigg)^{\frac{1}{p}} = \Bigg\|\sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} c_k\Bigg\|_{L^p(\Omega;\mathbb{R})}$$

for any $c_k \in \mathbb{R}$ with $c_k \geq 0$ by formula (IV.98), equation (IV.86) finally follows from equation (IV.103). ☐

Considering equation (IV.86) in Lemma IV.19, we separated the terms that are $\mathscr{F}_{t_0}$-measurable from those that are independent of σ-algebra $\mathscr{F}_{t_0}$. To be more precise, the stochastic process inside the $S^p([t_0,T] \times \Omega; \mathbb{R}^d)$-norm is not only independent of σ-algebra $\mathscr{F}_{t_0}$, it is rather $\mathscr{B}([t_0,T]) \otimes \mathscr{G} / \mathscr{B}(\mathbb{R}^d)$-measurable. Thus, we have the necessary $\mathscr{G}$-measurability, which is needed in order to apply techniques from the Malliavin calculus, cf. Chapter III.

In order to prove that term $\mathcal{R}_5^l$ is of order $\mathcal{O}(h)$ for $l \in \{1, \ldots, D\}$ as $h \to 0$, we consider the term

$$\Bigg\|\int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{z_k^r})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l) \vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r})) \, \mathrm{d}W_u^j \, \mathrm{d}s\Bigg\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \tag{IV.104}$$

from equation (IV.86) in Lemma IV.19 and prove that it converges for all $z_k^r \in C([t_0 - \tau, t_0]; \mathbb{R}^d)$ to zero with order $\mathcal{O}(h)$ as $h \to 0$.

As we mentioned in Section IV.2, cf. inequality (IV.43), the stochastic process inside the norm of term (IV.104) is not a time-continuous nor, restricted to the points in time $\{t_0, t_1, \ldots, t_N\}$, a time-discrete martingale or submartingale in general. Thus, neither the Burkholder inequality nor Doob's maximal inequality is applicable. In order to handle the supremum over time without Doob's martingale inequality, the so-called factorization method is used, cf. [3, p. 246] and [4, p. 142] as well as [26] and [27, p. 128]. The method is based on the following lemma, cf. [3, p. 246].

Lemma IV.20
For all $\vartheta \in \,]0,1[$ and $s, t \in \mathbb{R}$ with $s < t$, it holds

$$\int_s^t (r-s)^{-\vartheta}(t-r)^{\vartheta-1} \, \mathrm{d}r = \frac{\pi}{\sin(\pi\vartheta)}.$$

Proof. According to [5, Theorem 1.1.4 and Theorem 1.2.1], Euler's reflection formula states

$$\int_0^1 v^{-\vartheta}(1-v)^{\vartheta-1} \, \mathrm{d}v = \frac{\pi}{\sin(\pi\vartheta)}$$

for $\vartheta \in \,]0,1[$. Then, for $s,t \in \mathbb{R}$ with $s < t$, the substitution $v = \frac{r-s}{t-s}$ yields

$$\begin{aligned}\int_0^1 v^{-\vartheta}(1-v)^{\vartheta-1} \, \mathrm{d}v &= \frac{1}{t-s}\int_s^t \Big(\frac{r-s}{t-s}\Big)^{-\vartheta}\Big(1-\frac{r-s}{t-s}\Big)^{\vartheta-1} \mathrm{d}r \\ &= \int_s^t (r-s)^{-\vartheta}(t-r)^{\vartheta-1} \, \mathrm{d}r,\end{aligned}$$

which completes the proof. ☐

Using this lemma, Fubini's theorem leads to the following result.

Lemma IV.21
Let $f \in L^1([t_0, T]; \mathbb{R}^d)$ *and* $p \in]1, \infty[$*. For all* $\vartheta \in]\frac{1}{p}, 1[$*, it holds*

$$\sup_{t \in [t_0, T]} \left\| \int_{t_0}^{t} f(s) \,\mathrm{d}s \right\|^p \leq \Big(\frac{\sin(\pi\vartheta)}{\pi} \Big)^p \Big(\frac{p-1}{p\vartheta - 1} \Big)^{p-1} (T - t_0)^{p\vartheta - 1} \int_{t_0}^{T} \left\| \int_{t_0}^{r} (r-s)^{-\vartheta} f(s) \,\mathrm{d}s \right\|^p \mathrm{d}r. \tag{IV.105}$$

Proof. The proof is inspired by [3, p. 246]. There, the Skorohod integral is considered instead of the integral over time. Using Lemma IV.20 and Fubini's theorem, we obtain

$$\begin{aligned} \sup_{t \in [t_0, T]} \left\| \int_{t_0}^{t} f(s) \,\mathrm{d}s \right\|^p &= \sup_{t \in [t_0, T]} \left\| \int_{t_0}^{t} \Big(\int_{s}^{t} (r-s)^{-\vartheta} (t-r)^{\vartheta - 1} \,\mathrm{d}r \Big) \Big(\frac{\pi}{\sin(\pi\vartheta)} \Big)^{-1} f(s) \,\mathrm{d}s \right\|^p \\ &= \Big(\frac{\sin(\pi\vartheta)}{\pi} \Big)^p \sup_{t \in [t_0, T]} \left\| \int_{t_0}^{t} \int_{t_0}^{r} (r-s)^{-\vartheta} (t-r)^{\vartheta - 1} f(s) \,\mathrm{d}s \,\mathrm{d}r \right\|^p \\ &= \Big(\frac{\sin(\pi\vartheta)}{\pi} \Big)^p \sup_{t \in [t_0, T]} \left\| \int_{t_0}^{t} \int_{t_0}^{r} (r-s)^{-\vartheta} f(s) \,\mathrm{d}s (t-r)^{\vartheta - 1} \,\mathrm{d}r \right\|^p . \end{aligned}$$

Applying the triangle inequality and Hölder's inequality, it follows

$$\begin{aligned} &\sup_{t \in [t_0, T]} \left\| \int_{t_0}^{t} f(s) \,\mathrm{d}s \right\|^p \\ &\leq \Big(\frac{\sin(\pi\vartheta)}{\pi} \Big)^p \sup_{t \in [t_0, T]} \Big(\int_{t_0}^{t} \left\| \int_{t_0}^{r} (r-s)^{-\vartheta} f(s) \,\mathrm{d}s \right\| (t-r)^{\vartheta - 1} \,\mathrm{d}r \Big)^p \\ &\leq \Big(\frac{\sin(\pi\vartheta)}{\pi} \Big)^p \sup_{t \in [t_0, T]} \int_{t_0}^{t} \left\| \int_{t_0}^{r} (r-s)^{-\vartheta} f(s) \,\mathrm{d}s \right\|^p \mathrm{d}r \Big(\int_{t_0}^{t} (t-r)^{\frac{p}{p-1}(\vartheta - 1)} \,\mathrm{d}r \Big)^{p-1} \\ &= \Big(\frac{\sin(\pi\vartheta)}{\pi} \Big)^p \Big(\frac{p-1}{p\vartheta - 1} \Big)^{p-1} \sup_{t \in [t_0, T]} \int_{t_0}^{t} \left\| \int_{t_0}^{r} (r-s)^{-\vartheta} f(s) \,\mathrm{d}s \right\|^p \mathrm{d}r \, (t - t_0)^{p\vartheta - 1} \qquad \text{(IV.106)} \\ &\leq \Big(\frac{\sin(\pi\vartheta)}{\pi} \Big)^p \Big(\frac{p-1}{p\vartheta - 1} \Big)^{p-1} (T - t_0)^{p\vartheta - 1} \int_{t_0}^{T} \left\| \int_{t_0}^{r} (r-s)^{-\vartheta} f(s) \,\mathrm{d}s \right\|^p \mathrm{d}r, \end{aligned}$$

where $\frac{p}{p-1}(\vartheta - 1) > -1$ because of $\vartheta > \frac{1}{p}$. □

Lemma IV.21 states, roughly speaking, that the supremum over time can be estimated by changing it to an integral over time and multiply the integrand by a factor. The remarkable property of estimate (IV.105) is that it does not consider the Euclidean norm of the integrand f. This is important later on in order to obtain the desired order of convergence $\alpha = 1$ of the Milstein scheme. Using the triangle inequality instead of Lemma IV.20 and Fubini's theorem in the proof of Lemma IV.21, the resulting estimate would be too rough, cf. term (IV.40) and inequality (IV.41) in Section IV.2.

The next lemma transfers the statement of Lemma IV.21 to integrands f that are stochastic processes.

Lemma IV.22
Let $f\colon [t_0,T]\times\Omega\to\mathbb{R}^d$ be a measurable stochastic process such that

$$\mathrm{E}\Big[\Big(\int_{t_0}^{T}\|f(s)\|^2\,\mathrm{d}s\Big)^{\frac{p}{2}}\Big]<\infty \tag{IV.107}$$

for some $p\in\,]2,\infty[$. Then, for all $\vartheta\in\,]\frac{1}{p},\frac{1}{2}[$, it holds

$$\begin{aligned}&\Big\|\int_{t_0}^{\cdot}f(s)\,\mathrm{d}s\Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}\\&\quad\leq\frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}(T-t_0)^{\vartheta-\frac{1}{p}}\Big(\int_{t_0}^{T}\Big\|\int_{t_0}^{t}(t-s)^{-\vartheta}f(s)\,\mathrm{d}s\Big\|_{L^p(\Omega;\mathbb{R}^d)}^p\,\mathrm{d}t\Big)^{\frac{1}{p}}.\end{aligned}$$

Proof. According to assumption (IV.107) and due to the Cauchy-Schwarz inequality, P-almost all realizations of stochastic process f lie in $L^1([t_0,T];\mathbb{R}^d)$. Then, we obtain by Lemma IV.21 that

$$\begin{aligned}&\Big\|\int_{t_0}^{\cdot}f(s)\,\mathrm{d}s\Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)}\\&\quad\leq\frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}(T-t_0)^{\vartheta-\frac{1}{p}}\mathrm{E}\Big[\int_{t_0}^{T}\Big\|\int_{t_0}^{t}(t-s)^{-\vartheta}f(s)\,\mathrm{d}s\Big\|^p\,\mathrm{d}t\Big]^{\frac{1}{p}}\end{aligned}$$

for all $\vartheta\in]\frac{1}{p},1[$ at first. Considering the expectation on the right-hand side of the inequality above and using assumption (IV.107) on the measurable stochastic process f as well as that $-2\vartheta>-1$ if $\vartheta\in]\frac{1}{p},\frac{1}{2}[$, the Cauchy-Schwarz inequality implies

$$\begin{aligned}&\int_{t_0}^{T}\mathrm{E}\Big[\Big\|\int_{t_0}^{t}(t-s)^{-\vartheta}f(s)\,\mathrm{d}s\Big\|^p\Big]\,\mathrm{d}t\\&\quad\leq\int_{t_0}^{T}\Big(\int_{t_0}^{t}(t-s)^{-2\vartheta}\,\mathrm{d}s\Big)^{\frac{p}{2}}\mathrm{E}\Big[\Big(\int_{t_0}^{t}\|f(s)\|^2\,\mathrm{d}s\Big)^{\frac{p}{2}}\Big]\,\mathrm{d}t\\&\quad\leq\int_{t_0}^{T}\Big(\int_{t_0}^{t}(t-s)^{-2\vartheta}\,\mathrm{d}s\Big)^{\frac{p}{2}}\,\mathrm{d}t\,\mathrm{E}\Big[\Big(\int_{t_0}^{T}\|f(s)\|^2\,\mathrm{d}s\Big)^{\frac{p}{2}}\Big]\\&\quad=(1-2\vartheta)^{-\frac{p}{2}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-1}(T-t_0)^{1+(1-2\vartheta)\frac{p}{2}}\mathrm{E}\Big[\Big(\int_{t_0}^{T}\|f(s)\|^2\,\mathrm{d}s\Big)^{\frac{p}{2}}\Big]\\&\quad<\infty.\end{aligned}\tag{IV.108}$$

Due to this, the assertion of this lemma follows by Fubini's theorem for all $\vartheta\in\,]\frac{1}{p},\frac{1}{2}[$. □

Next, we apply Lemma IV.22, where integrand f is chosen to be the integrand of the integral over time in term (IV.104), that is

$$f(s)=\sum_{i=1}^{d}\partial_{x_l^i}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}^{z_k^r}))\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\,\mathrm{d}W_u^j$$

for $s\in[t_0,T]$, where $l\in\{1,\dots,D\}$, $k\in\{1,\dots,K_r\}$, and $r\in\mathbb{N}$.

In the following, let $l \in \{1, \dots, D\}$ and $z_k^r \in C([t_0 - \tau, t_0]; \mathbb{R}^d)$ with $k \in \{1, \dots, K_r\}$ and $r \in \mathbb{N}$ be arbitrarily fixed.

Using Assumption IV.8 *ii)* and Assumption IV.8 *iv)*, it holds $f \in H^p([t_0, T] \times \Omega; \mathbb{R}^d)$, that is the assumptions regarding integrand f in Lemma IV.22 are fulfilled. Then, Lemma IV.22 yields

$$\begin{aligned}
&\Big\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{z_k^r})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l) \vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j \,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\leq \frac{\sin(\pi\vartheta)}{\pi} \Big(\frac{p-1}{p\vartheta - 1} \Big)^{1-\frac{1}{p}} (T - t_0)^{\vartheta - \frac{1}{p}} \Big(\int_{t_0}^{T} \Big\| \int_{t_0}^{t} (t-s)^{-\vartheta} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{z_k^r})) \\
&\qquad \times \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j \,\mathrm{d}s \Big\|_{L^p(\Omega;\mathbb{R}^d)}^p \mathrm{d}t \Big)^{\frac{1}{p}} \\
&=: \frac{\sin(\pi\vartheta)}{\pi} \Big(\frac{p-1}{p\vartheta - 1} \Big)^{1-\frac{1}{p}} (T - t_0)^{\vartheta - \frac{1}{p}} \big(\mathcal{R}_5^{l, z_k^r} \big)^{\frac{1}{p}}
\end{aligned} \tag{IV.109}$$

for all $\vartheta \in]\frac{1}{p}, \frac{1}{2}[$, where $p \in]2, \infty[$ is specified by the assumptions in Theorem IV.9. Because of the deterministic initial condition $z_k^r \in C([t_0 - \tau, t_0]; \mathbb{R}^d)$, solution $X^{z_k^r}$ is in particular $\mathscr{B}([t_0 - \tau, T]) \otimes \mathscr{G} / \mathscr{B}(\mathbb{R}^d)$-measurable. Thus, the argument of the $L^p(\Omega; \mathbb{R}^d)$-norm in inequality (IV.109) is $\mathscr{G}/\mathscr{B}(\mathbb{R}^d)$-measurable and belongs actually to subspace $L_{\mathscr{G}}^p(\Omega; \mathbb{R}^d) \subset L^p(\Omega; \mathbb{R}^d)$.

Let $\vartheta \in]\frac{1}{p}, \frac{1}{2}[$ be arbitrarily fixed in the following. Here, condition $\vartheta < \frac{1}{2}$ is first of all needed in order to ensure the boundedness in inequality (IV.108) and later on to obtain the desired order of convergence $\alpha = 1$. Thus, together with condition $\vartheta > \frac{1}{p}$, which is needed in order to derive equation (IV.106), we have to assume in fact that $p > 2$.

Next, we look more closely at term $\mathcal{R}_5^{l, z_k^r}$.

The stochastic integral in inequality (IV.109) above equals zero as long as $s - \tau_l \leq t_0$. So, without loss of generality, we can assume

$$T \geq t_0 + \tau_l \tag{IV.110}$$

in the following, because we otherwise have $\mathcal{R}_5^{l, z_k^r} = 0$.

Still considering the stochastic integral in inequality (IV.109), the following holds. If the point in time $t_0 + \tau_l$ is not a point of the discretization, then $\lfloor t_0 + \tau_l \rfloor - \tau_l < t_0$, and we have

$$\sum_{j=1}^{m} \int_{(\lfloor t_0 + \tau_l \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j = \sum_{j=1}^{m} \int_{t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j. \tag{IV.111}$$

In consideration of equation (IV.111) and assumption (IV.110), we rewrite term $\mathcal{R}_5^{l, z_k^r}$ in inequality (IV.109) to

$$
\begin{aligned}
\mathcal{R}_5^{l,z_k^r} &= \int_{t_0+\tau_l}^{T} \bigg\| \int_{t_0+\tau_l}^{t} (t-s)^{-\vartheta} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{z_k^r})) \\
&\qquad \times \sum_{j=1}^{m} \int_{(\lfloor s \rfloor-\tau_l)\vee t_0}^{s-\tau_l} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^p \mathrm{d}t \\
&= \int_{t_0+\tau_l}^{T} \bigg\| \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor t_0+\tau_l \rfloor, X_{\lfloor t_0+\tau_l \rfloor}^{z_k^r})) \\
&\qquad \times \int_{t_0+\tau_l}^{\lceil t_0+\tau_l \rceil \wedge t} (t-s)^{-\vartheta} \sum_{j=1}^{m} \int_{t_0}^{s-\tau_l} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j\, \mathrm{d}s \\
&\quad + \sum_{\substack{n=1 \\ t_n \geq \lceil t_0+\tau_l \rceil}}^{N-1} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) \\
&\qquad \times \int_{t_n\wedge t}^{t_{n+1}\wedge t} (t-s)^{-\vartheta} \sum_{j=1}^{m} \int_{(t_n\wedge t)-\tau_l}^{s-\tau_l} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j\, \mathrm{d}s \bigg\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^p \mathrm{d}t \\
&= \int_{t_0+\tau_l}^{T} \bigg\| \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor t_0+\tau_l \rfloor, X_{\lfloor t_0+\tau_l \rfloor}^{z_k^r})) \\
&\qquad \times \int_{t_0}^{(\lceil t_0+\tau_l \rceil \wedge t)-\tau_l} (t-v-\tau_l)^{-\vartheta} \sum_{j=1}^{m} \int_{t_0}^{v} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j\, \mathrm{d}v \\
&\quad + \sum_{\substack{n=1 \\ t_n \geq \lceil t_0+\tau_l \rceil}}^{N-1} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) \int_{(t_n\wedge t)-\tau_l}^{(t_{n+1}\wedge t)-\tau_l} (t-v-\tau_l)^{-\vartheta} \\
&\qquad \times \sum_{j=1}^{m} \int_{(t_n\wedge t)-\tau_l}^{v} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j\, \mathrm{d}v \bigg\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^p \mathrm{d}t,
\end{aligned} \tag{IV.112}
$$

where we used the substitution $s = v + \tau_l$ in the last step.

Recall the stochastic integration by parts formula based on Itô's formula, see e. g. [64] or [75, p. 155], for a martingale $(M_u)_{u\in[s,t]}$ and a continuous function of bounded variation $(C_u)_{u\in[s,t]}$, where $t \in \mathbb{R}$ with $t \geq s$. The covariation process vanishes, and it holds

$$
\int_s^t M_u\, \mathrm{d}C_u = C_t M_t - C_s M_s - \int_s^t C_u\, \mathrm{d}M_u
$$

for all $t \geq s$ P-almost surely. Applying this to the integrals in formula (IV.112), we obtain

$$\begin{aligned}
&\int_{t_0}^{(\lceil t_0+\tau_l\rceil \wedge t)-\tau_l} (t-v-\tau_l)^{-\vartheta} \sum_{j=1}^{m} \int_{t_0}^{v} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j\,\mathrm{d}v \\
&= \int_{t_0}^{(\lceil t_0+\tau_l\rceil \wedge t)-\tau_l} (t-v-\tau_l)^{-\vartheta}\,\mathrm{d}v \sum_{j=1}^{m} \int_{t_0}^{(\lceil t_0+\tau_l\rceil \wedge t)-\tau_l} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j \\
&\quad - \sum_{j=1}^{m} \int_{t_0}^{(\lceil t_0+\tau_l\rceil \wedge t)-\tau_l} \int_{t_0}^{u} (t-v-\tau_l)^{-\vartheta}\,\mathrm{d}v\, b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j \\
&= \sum_{j=1}^{m} \int_{t_0}^{(\lceil t_0+\tau_l\rceil \wedge t)-\tau_l} \int_{u}^{(\lceil t_0+\tau_l\rceil \wedge t)-\tau_l} (t-v-\tau_l)^{-\vartheta}\,\mathrm{d}v\, b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j
\end{aligned} \tag{IV.113}$$

for all $t \in [t_0+\tau_l, T]$ P-almost surely and

$$\begin{aligned}
&\int_{(t_n\wedge t)-\tau_l}^{(t_{n+1}\wedge t)-\tau_l} (t-v-\tau_l)^{-\vartheta} \sum_{j=1}^{m} \int_{(t_n\wedge t)-\tau_l}^{v} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j\,\mathrm{d}v \\
&= \int_{(t_n\wedge t)-\tau_l}^{(t_{n+1}\wedge t)-\tau_l} (t-v-\tau_l)^{-\vartheta}\,\mathrm{d}v \sum_{j=1}^{m} \int_{(t_n\wedge t)-\tau_l}^{(t_{n+1}\wedge t)-\tau_l} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j \\
&\quad - \sum_{j=1}^{m} \int_{(t_n\wedge t)-\tau_l}^{(t_{n+1}\wedge t)-\tau_l} \int_{(t_n\wedge t)-\tau_l}^{u} (t-v-\tau_l)^{-\vartheta}\,\mathrm{d}v\, b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j \\
&= \sum_{j=1}^{m} \int_{(t_n\wedge t)-\tau_l}^{(t_{n+1}\wedge t)-\tau_l} \int_{u}^{(t_{n+1}\wedge t)-\tau_l} (t-v-\tau_l)^{-\vartheta}\,\mathrm{d}v\, b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j
\end{aligned} \tag{IV.114}$$

for all $t \in [t_0+\tau_l, T]$ and $t_n \geq \lceil t_0+\tau_l\rceil$, where $n \in \{1,\dots,N\}$, P-almost surely. Both equations (IV.113) and (IV.114) can also be understood as the application of a stochastic version of Fubini's theorem, see e. g. [15, 27, 120, 134]. Inserting equations (IV.113) and (IV.114) into formula (IV.112) and using the substitution $v = s - \tau_l$, we obtain

$$\begin{aligned}
\mathcal{R}_5^{l,z_k^r} = \int_{t_0+\tau_l}^{T} \Bigg\| &\sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor t_0+\tau_l\rfloor, X_{\lfloor t_0+\tau_l\rfloor}^{z_k^r})) \\
&\times \sum_{j=1}^{m} \int_{t_0}^{(\lceil t_0+\tau_l\rceil \wedge t)-\tau_l} \int_{u+\tau_l}^{\lceil t_0+\tau_l\rceil \wedge t} (t-s)^{-\vartheta}\,\mathrm{d}s\, b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j \\
+ &\sum_{\substack{n=1 \\ t_n \geq \lceil t_0+\tau_l\rceil}}^{N-1} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) \\
&\times \sum_{j=1}^{m} \int_{(t_n\wedge t)-\tau_l}^{(t_{n+1}\wedge t)-\tau_l} \int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\,\mathrm{d}s\, b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j \Bigg\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^{p} \mathrm{d}t.
\end{aligned}$$

We rewrite this to

$$\mathcal{R}_5^{l,z_k^r} = \int_{t_0+\tau_l}^{T} \Bigg\| \sum_{\substack{n=0\\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) \times \sum_{j=1}^{m} \int_{((t_n\wedge t)-\tau_l)\vee t_0}^{(t_{n+1}\wedge t)-\tau_l} \int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\,\mathrm{d}s\, b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}W_u^j \Bigg\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^p \mathrm{d}t.$$

In the following, we use techniques from the Malliavin calculus in order to estimate term $\mathcal{R}_5^{l,z_k^r}$.

Because of the deterministic initial condition $z_k^r \in C([t_0-\tau,t_0];\mathbb{R}^d)$, solution $(X_u^{z_k^r})_{u\in[t_0-\tau,T]}$ is $(\mathscr{G}_{u\vee t_0})_{u\in[t_0-\tau,T]}$-progressively measurable, and thus, the integrand of the Itô integral in the equation above is adapted to filtration $(\mathscr{G}_u)_{u\in[((t_n\wedge t)-\tau_l)\vee t_0,(t_{n+1}\wedge t)-\tau_l]}$. Then, using the property of Skorohod integrals in Proposition III.22, we obtain

$$\mathcal{R}_5^{l,z_k^r} = \int_{t_0+\tau_l}^{T} \Bigg\| \sum_{\substack{n=0\\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) \sum_{j=1}^{m} \int_{t_0}^{T} \mathbb{1}_{]((t_n\wedge t)-\tau_l)\vee t_0,(t_{n+1}\wedge t)-\tau_l]}(u) \times \int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\,\mathrm{d}s\, b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\delta W_u^j \Bigg\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^p \mathrm{d}t.$$

According to Theorem III.26, it holds $X_t^{z_k^r} \in \mathscr{D}^q(\Omega;\mathbb{R}^d)$ for all $t\in[t_0-\tau,T]$ and $q\in[2,\infty[$. Then, we have by Theorem III.9 and Assumption IV.8 *v)* that $\partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) \in \mathscr{D}^q(\Omega;\mathbb{R}^d)$ for all $q\in[2,\infty[$. Further, since $\|\partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))\| \leq L_a$ by Assumption IV.8 *ii)*, Proposition III.21 applies, and by linearity of the integrals, we obtain

$$\begin{aligned}\mathcal{R}_5^{l,z_k^r} = \int_{t_0+\tau_l}^{T} \Bigg\| &\sum_{j=1}^{m} \int_{t_0}^{T} \sum_{\substack{n=0\\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \mathbb{1}_{]((t_n\wedge t)-\tau_l)\vee t_0,(t_{n+1}\wedge t)-\tau_l]}(u) \\ &\times \int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\,\mathrm{d}s \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\delta W_u^j \\ &+ \int_{t_0}^{T} \sum_{\substack{n=0\\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \mathbb{1}_{]((t_n\wedge t)-\tau_l)\vee t_0,(t_{n+1}\wedge t)-\tau_l]}(u) \int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\,\mathrm{d}s \\ &\times \sum_{j=1}^{m} \sum_{i=1}^{d} \big(\mathrm{D}_u^j \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))\big) b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{d}u \Bigg\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^p \mathrm{d}t.\end{aligned}$$

Applying the triangle inequality twice yields

$$\big(\mathcal{R}_5^{l,z_k^r}\big)^{\frac{1}{p}} \leq \mathcal{R}_5' + \mathcal{R}_5'', \tag{IV.115}$$

where

$$\mathcal{R}_5' := \Big(\int_{t_0+\tau_l}^{T} \Big\| \sum_{j=1}^{m} \int_{t_0}^{T} \sum_{\substack{n=0 \\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \mathbb{1}_{]((t_n \wedge t)-\tau_l)\vee t_0,((t_{n+1}\wedge t)-\tau_l]}(u) \int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\, \mathrm{d}s \\ \times \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \delta W_u^j \Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^p \mathrm{d}t \Big)^{\frac{1}{p}} \tag{IV.116}$$

and

$$\mathcal{R}_5'' := \Big(\int_{t_0+\tau_l}^{T} \Big\| \int_{t_0}^{T} \sum_{\substack{n=0 \\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \mathbb{1}_{]((t_n \wedge t)-\tau_l)\vee t_0,((t_{n+1}\wedge t)-\tau_l]}(u) \int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\, \mathrm{d}s \\ \times \sum_{j=1}^{m} \sum_{i=1}^{d} \big(\mathrm{D}_u^j \partial_{x_l^i} a(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))\big) b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}u \Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}^p \mathrm{d}t \Big)^{\frac{1}{p}}. \tag{IV.117}$$

The terms $\mathcal{R}_5'$ and $\mathcal{R}_5''$ are separately estimated in the following. In order to show that term $\mathcal{R}_5'$ is of order $\mathcal{O}(h)$ as $h \to 0$, the continuity of Skorohod integral operator δ is used, cf. Proposition III.25.

Using inequality (III.20), for term $\mathcal{R}_5'$ defined in formula (IV.116), we obtain

$$\mathcal{R}_5' \leq c_{\delta,p} \Bigg(\int_{t_0+\tau_l}^{T} \Bigg(\sum_{\iota=1}^{d} \Big\| \sum_{j=1}^{m} \int_{t_0}^{T} \Big| \sum_{\substack{n=0 \\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \mathbb{1}_{]((t_n \wedge t)-\tau_l)\vee t_0,((t_{n+1}\wedge t)-\tau_l]}(u) \\ \times \int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\, \mathrm{d}s \sum_{i=1}^{d} \partial_{x_l^i} a^{\iota}(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) b^{i,j}(\mathcal{T}(u, X_u^{z_k^r})) \Big|^2 \mathrm{d}u \Big\|_{L^{\frac{p}{2}}_{\mathscr{G}}(\Omega;\mathbb{R})} \\ + \Big\| \sum_{j_1,j_2=1}^{m} \int_{t_0}^{T} \int_{t_0}^{T} \Big| \sum_{\substack{n=0 \\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \mathbb{1}_{]((t_n \wedge t)-\tau_l)\vee t_0,((t_{n+1}\wedge t)-\tau_l]}(u) \int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\, \mathrm{d}s \\ \times \sum_{i=1}^{d} \mathrm{D}_v^{j_2} \Big(\partial_{x_l^i} a^{\iota}(\mathcal{T}(t_n, X_{t_n}^{z_k^r})) b^{i,j_1}(\mathcal{T}(u, X_u^{z_k^r})) \Big) \Big|^2 \mathrm{d}v\, \mathrm{d}u \Big\|_{L^{\frac{p}{2}}_{\mathscr{G}}(\Omega;\mathbb{R})} \Bigg)^{\frac{p}{2}} \mathrm{d}t \Bigg)^{\frac{1}{p}}. \tag{IV.118}$$

Similarly to identity (IV.98), it holds

$$\Big| \sum_{\substack{n=0 \\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \mathbb{1}_{]((t_n \wedge t)-\tau_l)\vee t_0,((t_{n+1}\wedge t)-\tau_l]}(u)\, f_n(u,\omega) \Big|^2 \\ = \sum_{\substack{n=0 \\ t_n \geq \lfloor t_0+\tau_l \rfloor}}^{N-1} \mathbb{1}_{]((t_n \wedge t)-\tau_l)\vee t_0,((t_{n+1}\wedge t)-\tau_l]}(u)\, |f_n(u,\omega)|^2$$

for all $(u,\omega) \in [t_0,T]\times\Omega$ and processes $f_n\colon [t_0,T]\times\Omega\to\mathbb{R}$, where $n\in\{0,1,\dots,N-1\}$. Using this, we rewrite the right-hand side of inequality (IV.118) and have

$$\begin{aligned}\mathcal{R}_5' \leq c_{\delta,p}\Bigg(\int_{t_0+\tau_l}^{T}\Bigg(\sum_{\iota=1}^{d}\Bigg\|\int_{t_0}^{T}\sum_{\substack{n=0\\ t_n\geq\lfloor t_0+\tau_l\rfloor}}^{N-1}\mathbb{1}_{]((t_n\wedge t)-\tau_l)\vee t_0,(t_{n+1}\wedge t)-\tau_l]}(u)\Bigg(\int_{u+\tau_l}^{t_{n+1}\wedge t}(t-s)^{-\vartheta}\,\mathrm{d}s\Bigg)^2\\ \times\sum_{j=1}^{m}\Bigg|\sum_{i=1}^{d}\partial_{x_l^i}a^{\iota}(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\Bigg|^2\,\mathrm{d}u\Bigg\|_{L_{\mathscr{G}}^{\frac{p}{2}}(\Omega;\mathbb{R})}\\ +\Bigg\|\int_{t_0}^{T}\sum_{\substack{n=0\\ t_n\geq\lfloor t_0+\tau_l\rfloor}}^{N-1}\mathbb{1}_{]((t_n\wedge t)-\tau_l)\vee t_0,(t_{n+1}\wedge t)-\tau_l]}(u)\Bigg(\int_{u+\tau_l}^{t_{n+1}\wedge t}(t-s)^{-\vartheta}\,\mathrm{d}s\Bigg)^2\\ \times\sum_{j_1,j_2=1}^{m}\int_{t_0}^{T}\Bigg|\sum_{i=1}^{d}\mathrm{D}_v^{j_2}\Big(\partial_{x_l^i}a^{\iota}(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))b^{i,j_1}(\mathcal{T}(u,X_u^{z_k^r}))\Big)\Bigg|^2\,\mathrm{d}v\,\mathrm{d}u\Bigg\|_{L_{\mathscr{G}}^{\frac{p}{2}}(\Omega;\mathbb{R})}\Bigg)^{\frac{p}{2}}\mathrm{d}t\Bigg)^{\frac{1}{p}}.\end{aligned}$$

Further, the triangle inequality implies

$$\begin{aligned}\mathcal{R}_5' \leq c_{\delta,p}\Bigg(\int_{t_0+\tau_l}^{T}\Bigg(\int_{t_0}^{T}\sum_{\substack{n=0\\ t_n\geq\lfloor t_0+\tau_l\rfloor}}^{N-1}\mathbb{1}_{]((t_n\wedge t)-\tau_l)\vee t_0,(t_{n+1}\wedge t)-\tau_l]}(u)\Bigg(\int_{u+\tau_l}^{t_{n+1}\wedge t}(t-s)^{-\vartheta}\,\mathrm{d}s\Bigg)^2\\ \times\Bigg(\sum_{\iota=1}^{d}\sum_{j=1}^{m}\Bigg\|\sum_{i=1}^{d}\partial_{x_l^i}a^{\iota}(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\Bigg\|^2_{L_{\mathscr{G}}^{p}(\Omega;\mathbb{R})}\\ +\sum_{\iota=1}^{d}\sum_{j_1,j_2=1}^{m}\int_{t_0}^{T}\Bigg\|\sum_{i=1}^{d}\mathrm{D}_v^{j_2}\Big(\partial_{x_l^i}a^{\iota}(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))b^{i,j_1}(\mathcal{T}(u,X_u^{z_k^r}))\Big)\Bigg\|^2_{L_{\mathscr{G}}^{p}(\Omega;\mathbb{R})}\mathrm{d}v\Bigg)\\ \times\,\mathrm{d}u\Bigg)^{\frac{p}{2}}\mathrm{d}t\Bigg)^{\frac{1}{p}}.\end{aligned}\tag{IV.119}$$

Next, we consider the $L_{\mathscr{G}}^{p}(\Omega;\mathbb{R})$-norms inside the integrals of inequality (IV.119). We start with the first one. Using Assumption IV.8 *ii)* and Assumption IV.8 *iv)*, the triangle inequality and inequality (IV.67) imply

$$\begin{aligned}&\sum_{\iota=1}^{d}\sum_{j=1}^{m}\Bigg\|\sum_{i=1}^{d}\partial_{x_l^i}a^{\iota}(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\Bigg\|^2_{L_{\mathscr{G}}^{p}(\Omega;\mathbb{R})}\\ &\leq L_a^2 d^2\sum_{j=1}^{m}\big\|b^{j}(\mathcal{T}(u,X_u^{z_k^r}))\big\|^2_{L_{\mathscr{G}}^{p}(\Omega;\mathbb{R})}\\ &\leq L_a^2 d^2 K_b^2 m\big(1+\|X^{z_k^r}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big).\end{aligned}\tag{IV.120}$$

The second $L_{\mathscr{G}}^{p}(\Omega;\mathbb{R})$-norm in inequality (IV.119) involves the Malliavin derivative. As mentioned above, we have $X_t^{z_k^r}\in\mathscr{D}^q(\Omega;\mathbb{R}^d)$ for all $t\in[t_0-\tau,T]$ and $q\in[2,\infty[$ by Theorem III.26. Then, according to Assumption IV.8 *v)* and Assumption IV.8 *iv)*, we have by Theorem III.9 and Remark III.19 that $\partial_{x_l^i}a^{\iota}(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))\in\mathscr{D}^q(\Omega;\mathbb{R})$ and $b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\in\mathscr{D}^q(\Omega;\mathbb{R})$ for all

$q \in [2,\infty[$ as well. Using the product rule of the Malliavin derivative, cf. equation (III.8), and the chain rule from Theorem III.9, it $\mathrm{P}|_{\mathscr{G}}$-almost surely holds

$$\begin{aligned}
&\mathrm{D}_v^{j_2}\Big(\partial_{x_l^i} a^\iota(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))b^{i,j_1}(\mathcal{T}(u, X_u^{z_k^r}))\Big)\\
&\quad= \big(\mathrm{D}_v^{j_2}\partial_{x_l^i} a^\iota(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))\big)b^{i,j_1}(\mathcal{T}(u, X_u^{z_k^r})) + \partial_{x_l^i} a^\iota(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))\,\mathrm{D}_v^{j_2} b^{i,j_1}(\mathcal{T}(u, X_u^{z_k^r}))\\
&\quad= \sum_{l_2=0}^{D}\sum_{i_2=0}^{d} \partial_{x_{l_2}^{i_2}}\partial_{x_l^i} a^\iota(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))\big(\mathrm{D}_v^{j_2} X_{t_n-\tau_{l_2}}^{z_k^r,i_2}\big)b^{i,j_1}(\mathcal{T}(u, X_u^{z_k^r}))\\
&\qquad+ \partial_{x_l^i} a^\iota(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))\sum_{l_2=0}^{D}\sum_{i_2=0}^{d} \partial_{x_{l_2}^{i_2}} b^{i,j_1}(\mathcal{T}(u, X_u^{z_k^r}))\,\mathrm{D}_v^{j_2} X_{t_n-\tau_{l_2}}^{z_k^r,i_2}
\end{aligned} \tag{IV.121}$$

for $\lambda|_{[t_0,T]}$-almost all $v \in [t_0,T]$. Under Assumption IV.8 *v)*, Assumption IV.8 *iv)*, and Assumption IV.8 *ii)*, the triangle inequality and inequality (IV.67) imply for the argument of the second $L^p_{\mathscr{G}}(\Omega;\mathbb{R})$-norm in inequality (IV.119) that

$$\begin{aligned}
&\Big|\sum_{i=1}^{d}\mathrm{D}_v^{j_2}\Big(\partial_{x_l^i} a^\iota(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))b^{i,j_1}(\mathcal{T}(u, X_u^{z_k^r}))\Big)\Big|\\
&\quad\le K_{\partial^2 a} d\Big(\sup_{t\in[t_0-\tau,T]}\big(1+\|X_t^{z_k^r}\|^2\big)^{\frac{\varrho_a}{2}}\Big)\sum_{l_2=0}^{D}\big\|\mathrm{D}_v^{j_2} X_{t_n-\tau_{l_2}}^{z_k^r}\big\|\big\|b^{j_1}(\mathcal{T}(u, X_u^{z_k^r}))\big\|\\
&\qquad+ L_a\sqrt{d}\sum_{l_2=0}^{D}\sum_{i_2=0}^{d}\big\|\partial_{x_{l_2}^{i_2}} b^{j_1}(\mathcal{T}(u, X_u^{z_k^r}))\big\|\big|\mathrm{D}_v^{j_2} X_{t_n-\tau_{l_2}}^{z_k^r,i_2}\big|\\
&\quad\le K_{\partial^2 a} d K_b\Big(\sup_{t\in[t_0-\tau,T]}\big(1+\|X_t^{z_k^r}\|^2\big)^{\frac{\varrho_a+1}{2}}\Big)\sum_{l_2=0}^{D}\big\|\mathrm{D}_v^{j_2} X_{t_n-\tau_{l_2}}^{z_k^r}\big\| + L_a d L_b\sum_{l_2=0}^{D}\big\|\mathrm{D}_v^{j_2} X_{t_n-\tau_{l_2}}^{z_k^r}\big\|
\end{aligned} \tag{IV.122}$$

for $\lambda|_{[t_0,T]}$-almost all $v \in [t_0,T]$ holds $\mathrm{P}|_{\mathscr{G}}$-almost surely. Hence, by triangle inequality and Hölder's inequality with $\frac{\varrho_a+1}{\varrho_a+2}+\frac{1}{\varrho_a+2}=1$ for the second $L^p_{\mathscr{G}}(\Omega;\mathbb{R})$-norm in inequality (IV.119), we obtain

$$\begin{aligned}
&\sum_{\iota=1}^{d}\sum_{j_1,j_2=1}^{m}\int_{t_0}^{T}\Big\|\sum_{i=1}^{d}\mathrm{D}_v^{j_2}\Big(\partial_{x_l^i} a^\iota(\mathcal{T}(t_n, X_{t_n}^{z_k^r}))b^{i,j_1}(\mathcal{T}(u, X_u^{z_k^r}))\Big)\Big\|^2_{L^p_{\mathscr{G}}(\Omega;\mathbb{R})}\,\mathrm{d}v\\
&\quad\le dm(D+1)^2\sum_{j_2=1}^{m}\int_{t_0}^{T}\Big(K_{\partial^2 a} d K_b\Big\|\Big(\sup_{t\in[t_0-\tau,T]}\big(1+\|X_t^{z_k^r}\|^2\big)^{\frac{\varrho_a+1}{2}}\Big)\\
&\qquad\times\Big(\sup_{t\in[t_0-\tau,T]}\big\|\mathrm{D}_v^{j_2} X_t^{z_k^r}\big\|\Big)\Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R})} + L_a d L_b\big\|\mathrm{D}_v^{j_2} X^{z_k^r}\big\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\Big)^2\,\mathrm{d}v\\
&\quad\le dm(D+1)^2\sum_{j_2=1}^{m}\int_{t_0}^{T}\Big(K_{\partial^2 a} d K_b\big(1+\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+1}{2}}\\
&\qquad\times\big\|\mathrm{D}_v^{j_2} X^{z_k^r}\big\|_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} + L_a d L_b\big\|\mathrm{D}_v^{j_2} X^{z_k^r}\big\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\Big)^2\,\mathrm{d}v.
\end{aligned} \tag{IV.123}$$

Using further inequality (III.22) from Theorem III.26, we have

$$\begin{aligned}
&\sum_{\iota=1}^{d}\sum_{j_1,j_2=1}^{m}\int_{t_0}^{T}\Big\|\sum_{i=1}^{d}\mathrm{D}_v^{j_2}\Big(\partial_{x_l^i}a^\iota(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))b^{i,j_1}(\mathcal{T}(u,X_u^{z_k^r}))\Big)\Big\|_{L^p(\Omega;\mathbb{R})}^2\,\mathrm{d}v\\
&\quad\le dm^2(D+1)^2(T-t_0)\Big(K_{\partial^2 a}dK_bC_{\mathrm{D},(\varrho_a+2)p}\big(1+\|X^{z_k^r}\|_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}^2\big)^{\frac{\varrho_a+2}{2}}\\
&\qquad+L_adL_bC_{\mathrm{D},p}\big(1+\|X^{z_k^r}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}^2\big)^{\frac12}\Big)^2.
\end{aligned}\tag{IV.124}$$

Inserting the results from inequalities (IV.120) and (IV.124) into inequality (IV.119), it follows for term $\mathcal{R}_5'$, using that the right-hand sides of the inequalities (IV.120) and (IV.124) are independent of $n\in\{0,1,\dots,N-1\}$, $u\in[t_0,T]$, and $t\in[t_0+\tau_l,T]$, that

$$\begin{aligned}
\mathcal{R}_5'\le c_{\delta,p}\Bigg(&L_a^2d^2K_b^2m\big(1+\|X^{z_k^r}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}^2\big)+dm^2(D+1)^2(T-t_0)\\
&\times\Big(K_{\partial^2 a}dK_bC_{\mathrm{D},(\varrho_a+2)p}\big(1+\|X^{z_k^r}\|_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}^2\big)^{\frac{\varrho_a+2}{2}}\\
&\quad+L_adL_bC_{\mathrm{D},p}\big(1+\|X^{z_k^r}\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}^2\big)^{\frac12}\Big)^2\Bigg)^{\frac12}\\
&\times\Bigg(\int_{t_0+\tau_l}^{T}\Bigg(\int_{t_0}^{T}\sum_{\substack{n=0\\t_n\ge\lfloor t_0+\tau_l\rfloor}}^{N-1}\mathbb{1}_{]((t_n\wedge t)-\tau_l)\vee t_0,(t_{n+1}\wedge t)-\tau_l]}(u)\\
&\quad\times\Bigg(\int_{u+\tau_l}^{t_{n+1}\wedge t}(t-s)^{-\vartheta}\,\mathrm{d}s\Bigg)^2\mathrm{d}u\Bigg)^{\frac p2}\mathrm{d}t\Bigg)^{\frac1p}.
\end{aligned}\tag{IV.125}$$

Next, we estimate the integrals over time

$$\begin{aligned}
I&:=\Bigg(\int_{t_0+\tau_l}^{T}\Bigg(\int_{t_0}^{T}\sum_{\substack{n=0\\t_n\ge\lfloor t_0+\tau_l\rfloor}}^{N-1}\mathbb{1}_{]((t_n\wedge t)-\tau_l)\vee t_0,(t_{n+1}\wedge t)-\tau_l]}(u)\Bigg(\int_{u+\tau_l}^{t_{n+1}\wedge t}(t-s)^{-\vartheta}\,\mathrm{d}s\Bigg)^2\mathrm{d}u\Bigg)^{\frac p2}\mathrm{d}t\Bigg)^{\frac1p}\\
&=\Bigg(\int_{t_0+\tau_l}^{T}\Bigg(\sum_{\substack{n=0\\t_n\ge\lfloor t_0+\tau_l\rfloor}}^{N-1}\int_{((t_n\wedge t)-\tau_l)\vee t_0}^{(t_{n+1}\wedge t)-\tau_l}\Bigg(\int_{u+\tau_l}^{t_{n+1}\wedge t}(t-s)^{-\vartheta}\,\mathrm{d}s\Bigg)^2\mathrm{d}u\Bigg)^{\frac p2}\mathrm{d}t\Bigg)^{\frac1p}
\end{aligned}\tag{IV.126}$$

from inequality (IV.125) above. We make use of the following lemma.

Lemma IV.23
Let $a,b,c\in\mathbb{R}$ such that $a<b\le c$, and let $\vartheta\in\,]0,1[$. It holds

$$\int_a^b(c-s)^{-\vartheta}\,\mathrm{d}s\le\frac{1}{1-\vartheta}(c-a)^{-\vartheta}(b-a).$$

Proof. At first, we have

$$\int_a^b(c-s)^{-\vartheta}\,\mathrm{d}s=\frac{1}{1-\vartheta}\big((c-a)^{1-\vartheta}-(c-b)^{1-\vartheta}\big).\tag{IV.127}$$

In the case of $b = c$, we obtain

$$(c-a)^{1-\vartheta} - (c-b)^{1-\vartheta} = (c-a)^{-\vartheta}(c-a) = (c-a)^{-\vartheta}(b-a). \tag{IV.128}$$

Now, let $b < c$. Since $c - a > c - b$ implies $(c-a)^{-\vartheta} < (c-b)^{-\vartheta}$, it follows

$$\begin{aligned}(c-a)^{1-\vartheta} - (c-b)^{1-\vartheta} &= (c-a)^{-\vartheta}(c-a) - (c-b)^{-\vartheta}(c-b) \\ &\leq (c-a)^{-\vartheta}\big((c-a) - (c-b)\big) \\ &= (c-a)^{-\vartheta}(b-a). \end{aligned} \tag{IV.129}$$

Inserting the results from formulas (IV.128) and (IV.129) into equation (IV.127) proves the assertion. □

According to Lemma IV.23, for the inner integral in formula (IV.126), we have

$$\int_{u+\tau_l}^{t_{n+1}\wedge t} (t-s)^{-\vartheta}\,\mathrm{d}s \leq \frac{1}{1-\vartheta}(t-u-\tau_l)^{-\vartheta}((t_{n+1}\wedge t) - u - \tau_l),$$

where

$$(t_{n+1}\wedge t) - u - \tau_l \leq h$$

for all $u \in \big[((t_n \wedge t) - \tau_l) \vee t_0, (t_{n+1}\wedge t) - \tau_l\big]$. Thus, it holds

$$\begin{aligned} I &\leq \frac{1}{1-\vartheta}\Bigg(\int_{t_0+\tau_l}^{T}\Bigg(\sum_{\substack{n=0\\ t_n \geq \lfloor t_0+\tau_l\rfloor}}^{N-1}\int_{((t_n\wedge t)-\tau_l)\vee t_0}^{(t_{n+1}\wedge t)-\tau_l}(t-u-\tau_l)^{-2\vartheta}\,\mathrm{d}u\Bigg)^{\frac{p}{2}}\mathrm{d}t\Bigg)^{\frac{1}{p}} h \\ &= \frac{1}{1-\vartheta}\Bigg(\int_{t_0+\tau_l}^{T}\Bigg(\int_{t_0}^{t-\tau_l}(t-u-\tau_l)^{-2\vartheta}\,\mathrm{d}u\Bigg)^{\frac{p}{2}}\mathrm{d}t\Bigg)^{\frac{1}{p}} h, \end{aligned} \tag{IV.130}$$

where we simplified the term in the second step by summing up the inner integrals. Taking further into account that $\vartheta \in]\frac{1}{p}, \frac{1}{2}[$, where $p \in]2, \infty[$, it holds $-2\vartheta > -1$, and by simple integration of the right-hand side of formula (IV.130), we obtain

$$\begin{aligned} I &\leq \frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Bigg(\int_{t_0+\tau_l}^{T}(t-t_0-\tau_l)^{(1-2\vartheta)\frac{p}{2}}\,\mathrm{d}t\Bigg)^{\frac{1}{p}} h \\ &= \frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}\Big((T-t_0-\tau_l)^{1+(1-2\vartheta)\frac{p}{2}}\Big)^{\frac{1}{p}} h. \end{aligned}$$

Since

$$\Big((T-t_0-\tau_l)^{1+(1-2\vartheta)\frac{p}{2}}\Big)^{\frac{1}{p}} = (T-t_0-\tau_l)^{\frac{1}{2}-\vartheta+\frac{1}{p}} \leq (T-t_0)^{\frac{1}{2}-\vartheta+\frac{1}{p}},$$

it holds in summary that

$$I \leq \frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}(T-t_0)^{\frac{1}{2}-\vartheta+\frac{1}{p}} h. \tag{IV.131}$$

Inserting above inequality (IV.131) regarding term I defined in equation (IV.126) into inequality (IV.125), we obtain

$$\begin{aligned}\mathcal{R}_5' \le c_{\delta,p}\Bigg(& L_a^2 d^2 K_b^2 m\big(1+\|X^{z_k^r}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big) + dm^2(D+1)^2(T-t_0)\\ & \times\Big(K_{\partial^2 a} d K_b C_{\mathrm{D},(\varrho_a+2)p}\big(1+\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}\\ & \quad + L_a d L_b C_{\mathrm{D},p}\big(1+\|X^{z_k^r}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\Big)^2\Bigg)^{\frac{1}{2}}\\ & \times \frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}(T-t_0)^{\frac{1}{2}-\vartheta+\frac{1}{p}} h.\end{aligned}$$

Further, applying inequality $\sqrt{c_1+c_2}\le\sqrt{c_1}+\sqrt{c_2}$ that holds for $c_1,c_2\in\mathbb{R}$ with $c_1,c_2\ge 0$, to the right-hand side of the previous inequality, we have

$$\begin{aligned}\mathcal{R}_5' \le c_{\delta,p}\Big(& L_a d\sqrt{m}\big(K_b+\sqrt{dm(T-t_0)}(D+1)L_b C_{\mathrm{D},p}\big)\big(1+\|X^{z_k^r}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\\ & + d^{\frac{3}{2}} m(D+1)\sqrt{T-t_0}K_{\partial^2 a}K_b C_{\mathrm{D},(\varrho_a+2)p}\big(1+\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}\Big)\\ & \times \frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}(T-t_0)^{\frac{1}{2}-\vartheta+\frac{1}{p}} h.\end{aligned} \tag{IV.132}$$

That is, term $\mathcal{R}_5'$ is of order $\mathcal{O}(h)$ as $h\to 0$.

In the following, we consider term $\mathcal{R}_5''$ of inequality (IV.115). First, we move the $L^p(\Omega;\mathbb{R}^d)$-norm into the integral and estimate

$$\begin{aligned}\mathcal{R}_5'' \le \Bigg(\int_{t_0+\tau_l}^{T}\Bigg(& \sum_{\substack{n=0\\ t_n\ge\lfloor t_0+\tau_l\rfloor}}^{N-1}\int_{((t_n\wedge t)-\tau_l)\vee t_0}^{(t_{n+1}\wedge t)-\tau_l}\int_{u+\tau_l}^{t_{n+1}\wedge t}(t-s)^{-\vartheta}\,\mathrm{d}s\\ & \times\sum_{j=1}^{m}\Big\|\sum_{i=1}^{d}\big(\mathrm{D}_u^j\partial_{x_l^i}a(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))\big)b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}\mathrm{d}u\Bigg)^p\mathrm{d}t\Bigg)^{\frac{1}{p}},\end{aligned} \tag{IV.133}$$

where we used the linearity of the integral over time as in equation (IV.126).

According to equation (IV.121) and inequality (IV.122), it holds, using Assumption IV.8 *v)* and Assumption IV.8 *iv)* as well as Theorem III.26 and Theorem III.9, for the $L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)$-norm in inequality (IV.133) above that

$$\begin{aligned}&\sum_{j=1}^{m}\Big\|\sum_{i=1}^{d}\big(\mathrm{D}_u^j\partial_{x_l^i}a(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))\big)b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R}^d)}\\ &\quad\le K_{\partial^2 a}dK_b\sum_{j=1}^{m}\Big\|\Big(\sup_{t\in[t_0-\tau,T]}\big(1+\|X_t^{z_k^r}\|^2\big)^{\frac{\varrho_a+1}{2}}\Big)\sum_{l_2=0}^{D}\|\mathrm{D}_u^j X^{z_k^r}_{t_n-\tau_{l_2}}\|\Big\|_{L^p_{\mathscr{G}}(\Omega;\mathbb{R})}.\end{aligned}$$

Further, similarly to inequalities (IV.123) and (IV.124), Hölder's inequality with $\frac{\varrho_a+1}{\varrho_a+2}+\frac{1}{\varrho_a+2}=1$ and inequality (III.22) from Theorem III.26 imply

$$\begin{aligned}
&\sum_{j=1}^{m}\Big\|\sum_{i=1}^{d}\big(\mathrm{D}_u^j\partial_{x_l^i}a(\mathcal{T}(t_n,X_{t_n}^{z_k^r}))\big)b^{i,j}(\mathcal{T}(u,X_u^{z_k^r}))\Big\|_{L^p(\Omega;\mathbb{R}^d)}\\
&\quad\le K_{\partial^2 a}dK_b(D+1)\big(1+\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+1}{2}}\\
&\qquad\times\sum_{j=1}^{m}\big\|\mathrm{D}_u^jX^{z_k^r}\big\|_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\\
&\quad\le K_{\partial^2 a}dK_b(D+1)C_{\mathrm{D},(\varrho_a+2)p}m\big(1+\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}.
\end{aligned}$$

Thus, by inserting this into inequality (IV.133), we obtain

$$\begin{aligned}
\mathcal{R}_5''&\le K_{\partial^2 a}dK_b(D+1)C_{\mathrm{D},(\varrho_a+2)p}m\big(1+\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}\\
&\quad\times\bigg(\int_{t_0+\tau_l}^{T}\bigg(\sum_{\substack{n=0\\ t_n\ge\lfloor t_0+\tau_l\rfloor}}^{N-1}\int_{((t_n\wedge t)-\tau_l)\vee t_0}^{(t_{n+1}\wedge t)-\tau_l}\int_{u+\tau_l}^{t_{n+1}\wedge t}(t-s)^{-\vartheta}\,\mathrm{d}s\,\mathrm{d}u\bigg)^p\mathrm{d}t\bigg)^{\frac{1}{p}}.
\end{aligned}\tag{IV.134}$$

Similarly to inequality (IV.131), it holds by Lemma IV.23 and integration that

$$\begin{aligned}
&\bigg(\int_{t_0+\tau_l}^{T}\bigg(\sum_{\substack{n=0\\ t_n\ge\lfloor t_0+\tau_l\rfloor}}^{N-1}\int_{((t_n\wedge t)-\tau_l)\vee t_0}^{(t_{n+1}\wedge t)-\tau_l}\int_{u+\tau_l}^{t_{n+1}\wedge t}(t-s)^{-\vartheta}\,\mathrm{d}s\,\mathrm{d}u\bigg)^p\mathrm{d}t\bigg)^{\frac{1}{p}}\\
&\quad\le\frac{1}{1-\vartheta}\bigg(\int_{t_0+\tau_l}^{T}\bigg(\sum_{\substack{n=0\\ t_n\ge\lfloor t_0+\tau_l\rfloor}}^{N-1}\int_{((t_n\wedge t)-\tau_l)\vee t_0}^{(t_{n+1}\wedge t)-\tau_l}(t-u-\tau_l)^{-\vartheta}\,\mathrm{d}u\bigg)^p\mathrm{d}t\bigg)^{\frac{1}{p}}h\\
&\quad=\frac{1}{1-\vartheta}\bigg(\int_{t_0+\tau_l}^{T}\bigg(\int_{t_0}^{t-\tau_l}(t-u-\tau_l)^{-\vartheta}\,\mathrm{d}u\bigg)^p\mathrm{d}t\bigg)^{\frac{1}{p}}h\\
&\quad=\frac{1}{(1-\vartheta)^2}\bigg(\int_{t_0+\tau_l}^{T}(t-t_0-\tau_l)^{(1-\vartheta)p}\,\mathrm{d}t\bigg)^{\frac{1}{p}}h\\
&\quad=\frac{1}{(1-\vartheta)^2}\big(1+(1-\vartheta)p\big)^{-\frac{1}{p}}(T-t_0-\tau_l)^{1-\vartheta+\frac{1}{p}}h\\
&\quad\le\frac{1}{(1-\vartheta)^2}\big(1+(1-\vartheta)p\big)^{-\frac{1}{p}}(T-t_0)^{1-\vartheta+\frac{1}{p}}h.
\end{aligned}$$

Then, inserting this into inequality (IV.134) results in

$$\begin{aligned}
\mathcal{R}_5''&\le K_{\partial^2 a}dK_b(D+1)C_{\mathrm{D},(\varrho_a+2)p}m\big(1+\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}\\
&\quad\times\frac{1}{(1-\vartheta)^2}\big(1+(1-\vartheta)p\big)^{-\frac{1}{p}}(T-t_0)^{1-\vartheta+\frac{1}{p}}h
\end{aligned},\tag{IV.135}$$

and term $\mathcal{R}_5''$ is of order $\mathcal{O}(h)$ as $h\to 0$, too.

Now, we combine inequalities (IV.109), (IV.115), (IV.132), and (IV.135), and for term (IV.104) from equation (IV.86) in Lemma IV.19, we obtain

$$
\begin{aligned}
&\Big\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{z_k^r})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j \,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\leq \frac{\sin(\pi\vartheta)}{\pi} \Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}} (T-t_0)^{\vartheta-\frac{1}{p}} \\
&\quad\times \Bigg(c_{\delta,p} \Big(L_a d\sqrt{m} \big(K_b + \sqrt{dm(T-t_0)}(D+1) L_b C_{\mathrm{D},p}\big) \big(1 + \|X^{z_k^r}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \\
&\qquad + d^{\frac{3}{2}} m(D+1)\sqrt{T-t_0} K_{\partial^2 a} K_b C_{\mathrm{D},(\varrho_a+2)p} \big(1 + \|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}} \Big) \\
&\qquad \times \frac{1}{1-\vartheta} \frac{1}{\sqrt{1-2\vartheta}} \Big(1 + (1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}} (T-t_0)^{\frac{1}{2}-\vartheta+\frac{1}{p}} h \\
&\quad + K_{\partial^2 a} d K_b (D+1) C_{\mathrm{D},(\varrho_a+2)p} m \big(1 + \|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}} \\
&\qquad \times \frac{1}{(1-\vartheta)^2} \big(1 + (1-\vartheta)p\big)^{-\frac{1}{p}} (T-t_0)^{1-\vartheta+\frac{1}{p}} h \Bigg).
\end{aligned}
$$

By rearranging the right-hand side of the above inequality, we finally have

$$
\begin{aligned}
&\Big\| \int_{t_0}^{\cdot} \sum_{i=1}^{d} \partial_{x_l^i} a(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor}^{z_k^r})) \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^{i,j}(\mathcal{T}(u, X_u^{z_k^r}))\, \mathrm{d}W_u^j \,\mathrm{d}s \Big\|_{S^p([t_0,T]\times\Omega;\mathbb{R}^d)} \\
&\leq \frac{\sin(\pi\vartheta)}{\pi} \Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}} \frac{1}{1-\vartheta} \frac{1}{\sqrt{1-2\vartheta}} \Big(1 + (1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}} c_{\delta,p} L_a d\sqrt{m} \\
&\qquad \times \Big(K_b + \sqrt{dm(T-t_0)}(D+1) L_b C_{\mathrm{D},p}\Big) \\
&\qquad \times \big(1 + \|X^{z_k^r}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \sqrt{T-t_0}\, h \\
&\quad + \frac{\sin(\pi\vartheta)}{\pi} \Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}} \frac{1}{1-\vartheta} K_{\partial^2 a} d K_b (D+1) C_{\mathrm{D},(\varrho_a+2)p} m \\
&\qquad \times \Big(\frac{c_{\delta,p}\sqrt{d}}{\sqrt{1-2\vartheta}} \Big(1 + (1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}} + \frac{1}{1-\vartheta} \big(1 + (1-\vartheta)p\big)^{-\frac{1}{p}} \Big) \\
&\qquad \times \big(1 + \|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}} (T-t_0)\, h
\end{aligned}
\tag{IV.136}
$$

for $l \in \{1, \ldots, D\}$ and $z_k^r \in C([t_0-\tau, T]; \mathbb{R}^d)$, where $\vartheta \in]\frac{1}{p}, \frac{1}{2}[$.

Next, we insert estimate (IV.136), which is of order $\mathcal{O}(h)$ as $h \to 0$, into equation (IV.86) from

Lemma IV.19. Using the triangle inequality, we obtain

$$
\begin{aligned}
\mathcal{R}_5^l \leq{}& \frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}\frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}c_{\delta,p}L_a d\sqrt{m} \\
&\times\Big(K_b+\sqrt{dm(T-t_0)}(D+1)L_b C_{\mathrm{D},p}\Big) \\
&\times\lim_{r\to\infty}\Big\|\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\big(1+\|X^{z_k^r}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\Big\|_{L^p(\Omega;\mathbb{R})}\sqrt{T-t_0}h \\
&+\frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}\frac{1}{1-\vartheta}K_{\partial^2 a}dK_b(D+1)C_{\mathrm{D},(\varrho_a+2)p}m \\
&\times\Big(\frac{c_{\delta,p}\sqrt{d}}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}+\frac{1}{1-\vartheta}\big(1+(1-\vartheta)p\big)^{-\frac{1}{p}}\Big) \\
&\times\lim_{r\to\infty}\Big\|\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\big(1+\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}\Big\|_{L^p(\Omega;\mathbb{R})}(T-t_0)h.
\end{aligned}
\tag{IV.137}
$$

In the following, we calculate and estimate the limits in inequality (IV.137) above. We only show the calculations for the second limit because the first one follows from the same considerations with exponent one instead of ϱ_a+2.

To begin with, we only consider the $L^p_{\mathscr{G}}(\Omega;\mathbb{R})$-norm for arbitrary $r\in\mathbb{N}$. Using property (IV.98) of step functions, rewriting the norm and applying the triangle, we obtain

$$
\begin{aligned}
&\Big\|\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\big(1+\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}\Big\|_{L^p(\Omega;\mathbb{R})} \\
&=\Big\|\Big(1+\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\Big)^{\frac{\varrho_a+2}{2}}\Big\|_{L^p(\Omega;\mathbb{R})} \\
&=\Big\|1+\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\Big\|^{\frac{\varrho_a+2}{2}}_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})} \\
&\leq\Big(1+\Big\|\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\Big\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})}\Big)^{\frac{\varrho_a+2}{2}}.
\end{aligned}
\tag{IV.138}
$$

Next, we consider the $L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})$ from the last line of previous calculations only. By rewriting the $S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$-norm, it holds

$$
\begin{aligned}
&\Big\|\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\Big\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})} \\
&=\Big\|\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\Big(\mathrm{E}\Big[\sup_{t\in[t_0-\tau,T]}\|X_t^{z_k^r}\|^{(\varrho_a+2)p}\Big]\Big)^{\frac{2}{(\varrho_a+2)p}}\Big\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})}.
\end{aligned}
\tag{IV.139}
$$

Recall that, for arbitrary $k\in\{1,\dots,K_r\}$ and $r\in\mathbb{N}$, the random variable $\mathbb{1}_{A_k^r}$ is $\mathscr{F}_{t_0}$-measurable, and solution $X^{z_k^r}$ is $(\mathscr{G}_{t\vee t_0})_{t\in[t_0-\tau,T]}$-progressively measurable. Thus, the random variable in the expectation on the right-hand side of equation (IV.139) is $\mathscr{G}/\mathscr{B}(\mathbb{R})$-measurable and independent of σ-algebra $\mathscr{F}_{t_0}$. We infer from the properties of conditional expectations and property (IV.98) of step functions that

$$\begin{aligned}
&\Bigg\| \sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} \Bigg(\mathrm{E}\Bigg[\sup_{t\in[t_0-\tau,T]} \|X_t^{z_k^r}\|^{(\varrho_a+2)p} \Bigg] \Bigg)^{\frac{2}{(\varrho_a+2)p}} \Bigg\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})} \\
&\quad = \Bigg\| \sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} \Bigg(\mathrm{E}\Bigg[\sup_{t\in[t_0-\tau,T]} \|X_t^{z_k^r}\|^{(\varrho_a+2)p} \Bigg| \mathscr{F}_{t_0} \Bigg] \Bigg)^{\frac{2}{(\varrho_a+2)p}} \Bigg\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})} \\
&\quad = \Bigg\| \Bigg(\mathrm{E}\Bigg[\sup_{t\in[t_0-\tau,T]} \sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} \|X_t^{z_k^r}\|^{(\varrho_a+2)p} \Bigg| \mathscr{F}_{t_0} \Bigg] \Bigg)^{\frac{2}{(\varrho_a+2)p}} \Bigg\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})} \\
&\quad = \Bigg\| \Bigg(\mathrm{E}\Bigg[\sup_{t\in[t_0-\tau,T]} \Bigg\| \sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} X_t^{z_k^r} \Bigg\|^{(\varrho_a+2)p} \Bigg| \mathscr{F}_{t_0} \Bigg] \Bigg)^{\frac{2}{(\varrho_a+2)p}} \Bigg\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})},
\end{aligned} \tag{IV.140}$$

also cf. the proof of Lemma IV.19, where the previous calculations are used vice-versa. Then, using Lemma IV.18 with $\zeta^r = \sum_{k=1}^{K_r} z_k^r \mathbb{1}_{A_k^r}$, we obtain

$$\begin{aligned}
&\Bigg\| \Bigg(\mathrm{E}\Bigg[\sup_{t\in[t_0-\tau,T]} \Bigg\| \sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} X_t^{z_k^r} \Bigg\|^{(\varrho_a+2)p} \Bigg| \mathscr{F}_{t_0} \Bigg] \Bigg)^{\frac{2}{(\varrho_a+2)p}} \Bigg\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})} \\
&\quad = \Bigg\| \Bigg(\mathrm{E}\Bigg[\sup_{t\in[t_0-\tau,T]} \|X_t^{\zeta^r}\|^{(\varrho_a+2)p} \Bigg| \mathscr{F}_{t_0} \Bigg] \Bigg)^{\frac{2}{(\varrho_a+2)p}} \Bigg\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})} \\
&\quad = \Bigg(\mathrm{E}\Bigg[\mathrm{E}\Bigg[\sup_{t\in[t_0-\tau,T]} \|X_t^{\zeta^r}\|^{(\varrho_a+2)p} \Bigg| \mathscr{F}_{t_0} \Bigg] \Bigg] \Bigg)^{\frac{2}{(\varrho_a+2)p}} \\
&\quad = \Bigg(\mathrm{E}\Bigg[\sup_{t\in[t_0-\tau,T]} \|X_t^{\zeta^r}\|^{(\varrho_a+2)p} \Bigg] \Bigg)^{\frac{2}{(\varrho_a+2)p}} \\
&\quad = \|X^{\zeta^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}.
\end{aligned} \tag{IV.141}$$

Summarizing the results from formulas (IV.138), (IV.139), (IV.140), and (IV.141) gives

$$\begin{aligned}
&\Bigg\| \sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} \big(1 + \|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}} \Bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq \big(1 + \|X^{\zeta^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}.
\end{aligned} \tag{IV.142}$$

Taking Lemma II.10 and equation (IV.84) into account, the dominated convergence theorem, according to inequality (IV.83), implies

$$\lim_{r\to\infty} \|X^{\zeta^r}\|_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = \|X^{\tilde{\xi}}\|_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} = \|X^{\xi}\|_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}$$

as solutions $X^{\tilde{\xi}}$ and X^{ξ} are indistinguishable, cf. equation (IV.82). Thus, taking the limit $r \to \infty$ in inequality (IV.142), we obtain

$$\begin{aligned}
&\lim_{r\to\infty} \Bigg\| \sum_{k=1}^{K_r} \mathbb{1}_{A_k^r} \big(1 + \|X^{z_k^r}\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}} \Bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq \big(1 + \|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}
\end{aligned} \tag{IV.143}$$

with the notation $X = X^{\xi}$. Completely analogous to this, it also holds

$$\lim_{r\to\infty}\Big\|\sum_{k=1}^{K_r}\mathbb{1}_{A_k^r}\big(1+\|X^{z_k^r}\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\Big\|_{L^p(\Omega;\mathbb{R})} \leq \big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}. \tag{IV.144}$$

Inserting inequalities (IV.143) and (IV.144) into inequality (IV.137), we finally obtain with

$$\begin{aligned}\mathcal{R}_5^l \leq{}& \frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}\frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}c_{\delta,p}L_a d\sqrt{m}\\ &\times\Big(K_b+\sqrt{dm(T-t_0)}(D+1)L_bC_{\mathrm{D},p}\Big)\big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\sqrt{T-t_0}\,h\\ &+\frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}\frac{1}{1-\vartheta}K_{\partial^2 a}dK_b(D+1)C_{\mathrm{D},(\varrho_a+2)p}m\\ &\times\Big(\frac{c_{\delta,p}\sqrt{d}}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}+\frac{1}{1-\vartheta}\big(1+(1-\vartheta)p\big)^{-\frac{1}{p}}\Big)\\ &\times\big(1+\|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}(T-t_0)\,h\end{aligned} \tag{IV.145}$$

the desired estimate of order $\mathcal{O}(h)$ as $h\to 0$ for term $\mathcal{R}_5^l$, where $l\in\{1,\dots,D\}$ and $\vartheta\in\,]\frac{1}{p},\frac{1}{2}[$ arbitrarily.

Thus, in view of inequality (IV.79), we in total have

$$\begin{aligned}\mathcal{R}_5 ={}& \sum_{l=0}^{D}\mathcal{R}_5^l\\ \leq{}& \Big(L_a\sqrt{d}\sqrt{p-1}K_b\sqrt{m}\Big(\frac{2}{3}\frac{p}{\sqrt{p-1}}+\frac{1}{\sqrt{2}}\Big)\\ &+D\frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}\frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}c_{\delta,p}L_a d\sqrt{m}\\ &\times\Big(K_b+\sqrt{dm(T-t_0)}(D+1)L_bC_{\mathrm{D},p}\Big)\Big)\big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\sqrt{T-t_0}\,h\\ &+D\frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}\frac{1}{1-\vartheta}K_{\partial^2 a}dK_b(D+1)C_{\mathrm{D},(\varrho_a+2)p}m\\ &\times\Big(\frac{c_{\delta,p}\sqrt{d}}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}}+\frac{1}{1-\vartheta}\big(1+(1-\vartheta)p\big)^{-\frac{1}{p}}\Big)\\ &\times\big(1+\|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a+2}{2}}(T-t_0)\,h,\end{aligned} \tag{IV.146}$$

where $\vartheta\in\,]\frac{1}{p},\frac{1}{2}[$ can be chosen arbitrarily. That is, term $\mathcal{R}_5$ is of order $\mathcal{O}(h)$ as $h\to 0$.

In the following, we give two remarks on the terms that depend on $\vartheta\in\,]\frac{1}{p},\frac{1}{2}[$ in the upper bound of term $\mathcal{R}_5$ in estimate (IV.146).

Remark IV.24
Numerical simulations on the ϑ-depending terms in inequality (IV.146) *indicate that*

$$\lim_{p\to\infty} \min_{\vartheta\in]\frac{1}{p},\frac{1}{2}[} \frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}} \frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}} = 1$$

and

$$\lim_{p\to\infty} \min_{\vartheta\in]\frac{1}{p},\frac{1}{2}[} \frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}} \frac{1}{(1-\vartheta)^2}\big(1+(1-\vartheta)p\big)^{-\frac{1}{p}} = 1.$$

These terms occur through application of Lemma IV.22. Lemma IV.22 is used in order to estimate the supremum over time in term $\mathcal{R}_5^l$, where $l \in \{1,\dots,D\}$. As previously mentioned in Section IV.2, Doob's maximal inequality cannot be applied because the processes under consideration are no martingales nor submartingales. Doob's maximal inequality holds true with constant $\frac{p}{p-1}$, see e. g. [35, Theorem 3.4 on p. 317] or [67, Theorem 26.3]. Thus, our constants are consistent in the sense that also

$$\lim_{p\to\infty}\frac{p}{p-1} = 1.$$

Remark IV.25
The ϑ-depending terms in inequality (IV.146) *can be bounded from above as follows. Taking the monotonicity of the single factors with respect to $\vartheta \in]\frac{1}{p},\frac{1}{2}[$ into account, it holds*

$$\begin{aligned}&\min_{\vartheta\in]\frac{1}{p},\frac{1}{2}[} \frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}} \frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}} \\ &\quad\le \min_{\vartheta\in]\frac{1}{p},\frac{1}{2}[} \frac{2}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}\frac{1}{\sqrt{1-2\vartheta}}\end{aligned} \tag{IV.147}$$

for all $p \in]2,\infty[$. Simple calculations show that the minimum occurs at $\vartheta = \frac{p}{3p-2} \in]\frac{1}{p},\frac{1}{2}[$ for all $p \in]2,\infty[$. Inserting this position, we obtain

$$\min_{\vartheta\in]\frac{1}{p},\frac{1}{2}[} \frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}} \frac{1}{1-\vartheta}\frac{1}{\sqrt{1-2\vartheta}}\Big(1+(1-2\vartheta)\frac{p}{2}\Big)^{-\frac{1}{p}} \le \frac{2}{\pi}\Big(3+\frac{4}{p-2}\Big)^{\frac{3}{2}-\frac{1}{p}}$$

for all $p \in]2,\infty[$. Similarly to inequality (IV.147), *it holds by monotonicity that*

$$\begin{aligned}&\min_{\vartheta\in]\frac{1}{p},\frac{1}{2}[} \frac{\sin(\pi\vartheta)}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}} \frac{1}{(1-\vartheta)^2}\big(1+(1-\vartheta)p\big)^{-\frac{1}{p}} \\ &\quad\le \min_{\vartheta\in]\frac{1}{p},\frac{1}{2}[} \frac{4}{\pi}\Big(\frac{p-1}{p\vartheta-1}\Big)^{1-\frac{1}{p}}\Big(1+\frac{p}{2}\Big)^{-\frac{1}{p}} \\ &\quad\le \frac{4}{\pi}\Big(\frac{p-1}{\frac{p}{2}-1}\Big)^{1-\frac{1}{p}}\Big(1+\frac{p}{2}\Big)^{-\frac{1}{p}} \\ &\quad= \frac{8}{\pi}\Big(\frac{p-1}{p-2}\Big)^{1-\frac{1}{p}}(p+2)^{-\frac{1}{p}}\end{aligned}$$

for all $p \in]2,\infty[$. However, these upper bounds are not optimal in view of Remark IV.24 because

$$\lim_{p\to\infty}\frac{2}{\pi}\Big(3+\frac{4}{p-2}\Big)^{\frac{3}{2}-\frac{1}{p}} = \frac{2}{\pi}\sqrt{27} > 1$$

and

$$\lim_{p\to\infty}\frac{8}{\pi}\Big(\frac{p-1}{p-2}\Big)^{1-\frac{1}{p}}(p+2)^{-\frac{1}{p}}=\frac{8}{\pi}>1.$$

Now, we continue with the next term $\mathcal{R}_6$. Applying the triangle inequality and using the growth condition from Assumption IV.8 *v)*, we obtain at first that

$$\begin{aligned}\mathcal{R}_6 &\le \sum_{l_1,l_2=0}^{D}\int_{t_0}^{T}\Big\|\sum_{i_1,i_2=1}^{d}\int_0^1\big\|\partial_{x^{i_1}_{l_1}}\partial_{x^{i_2}_{l_2}}a(\mathcal{T}(\lfloor s\rfloor,X_{\lfloor s\rfloor}+\theta(X_s-X_{\lfloor s\rfloor})))\big\|(1-\theta)\,\mathrm{d}\theta \\ &\qquad\times|X^{i_1}_{s-\tau_{l_1}}-X^{i_1}_{\lfloor s\rfloor-\tau_{l_1}}||X^{i_2}_{s-\tau_{l_2}}-X^{i_2}_{\lfloor s\rfloor-\tau_{l_2}}|\Big\|_{L^p(\Omega;\mathbb{R})}\mathrm{d}s \\ &\le K_{\partial^2 a}\sum_{l_1,l_2=0}^{D}\int_{t_0}^{T}\Big\|\int_0^1\sup_{l\in\{0,1,\ldots,D\}}\big(1+\|X_{\lfloor s\rfloor-\tau_l}+\theta(X_{s-\tau_l}-X_{\lfloor s\rfloor-\tau_l})\|^2\big)^{\frac{\varrho_a}{2}}(1-\theta)\,\mathrm{d}\theta \\ &\qquad\times\sum_{i_1=1}^{d}|X^{i_1}_{s-\tau_{l_1}}-X^{i_1}_{\lfloor s\rfloor-\tau_{l_1}}|\sum_{i_2=1}^{d}|X^{i_2}_{s-\tau_{l_2}}-X^{i_2}_{\lfloor s\rfloor-\tau_{l_2}}|\Big\|_{L^p(\Omega;\mathbb{R})}\mathrm{d}s.\end{aligned}\tag{IV.148}$$

Since

$$\begin{aligned}\|X_{\lfloor s\rfloor-\tau_l}+\theta(X_{s-\tau_l}-X_{\lfloor s\rfloor-\tau_l})\| &= \|(1-\theta)X_{\lfloor s\rfloor-\tau_l}+\theta X_{s-\tau_l}\| \\ &\le (1-\theta)\|X_{\lfloor s\rfloor-\tau_l}\|+\theta\|X_{s-\tau_l}\| \\ &\le \sup_{t\in[t_0-\tau,T]}\|X_t\|\end{aligned}$$

for all $\theta\in[0,1]$ and all $s\in[t_0,T]$, and since $\int_0^1(1-\theta)\,\mathrm{d}\theta=\frac{1}{2}$, we further have

$$\begin{aligned}\mathcal{R}_6 &\le \frac{1}{2}K_{\partial^2 a}\sum_{l_1,l_2=0}^{D}\int_{t_0}^{T}\Big\|\Big(1+\sup_{t\in[t_0-\tau,T]}\|X_t\|^2\Big)^{\frac{\varrho_a}{2}} \\ &\qquad\times\sum_{i_1=1}^{d}|X^{i_1}_{s-\tau_{l_1}}-X^{i_1}_{\lfloor s\rfloor-\tau_{l_1}}|\sum_{i_2=1}^{d}|X^{i_2}_{s-\tau_{l_2}}-X^{i_2}_{\lfloor s\rfloor-\tau_{l_2}}|\Big\|_{L^p(\Omega;\mathbb{R})}\mathrm{d}s,\end{aligned}$$

and inequality (IV.67) yields

$$\begin{aligned}\mathcal{R}_6 &\le \frac{1}{2}K_{\partial^2 a}d\sum_{l_1,l_2=0}^{D}\int_{t_0}^{T}\Big\|\Big(1+\sup_{t\in[t_0-\tau,T]}\|X_t\|^2\Big)^{\frac{\varrho_a}{2}} \\ &\qquad\times\|X_{s-\tau_{l_1}}-X_{\lfloor s\rfloor-\tau_{l_1}}\|\|X_{s-\tau_{l_2}}-X_{\lfloor s\rfloor-\tau_{l_2}}\|\Big\|_{L^p(\Omega;\mathbb{R})}\mathrm{d}s.\end{aligned}$$

Next, using first Hölder's inequality with $\frac{\varrho_a}{\varrho_a+2}+\frac{2}{\varrho_a+2}=1$ and afterwards the Cauchy-Schwarz inequality, it holds

$$\begin{aligned}\mathcal{R}_6 &\leq \frac{1}{2}K_{\partial^2 a}d\big(1+\|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a}{2}}\\ &\quad\times\sum_{l_1,l_2=0}^{D}\int_{t_0}^{T}\Big\|\|X_{s-\tau_{l_1}}-X_{\lfloor s\rfloor-\tau_{l_1}}\|\|X_{s-\tau_{l_2}}-X_{\lfloor s\rfloor-\tau_{l_2}}\|\Big\|_{L^{\frac{\varrho_a+2}{2}p}(\Omega;\mathbb{R})}\,\mathrm{d}s\\ &\leq \frac{1}{2}K_{\partial^2 a}d\big(1+\|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a}{2}}\\ &\quad\times\int_{t_0}^{T}\bigg(\sum_{l=0}^{D}\|X_{s-\tau_l}-X_{\lfloor s\rfloor-\tau_l}\|_{L^{(\varrho_a+2)p}(\Omega;\mathbb{R}^d)}\bigg)^2\mathrm{d}s.\end{aligned}\tag{IV.149}$$

Note that the inequalities above are satisfied in case of $\varrho_a=0$ as well. We now estimate the term $\sum_{l=0}^{D}\|X_{s-\tau_l}-X_{\lfloor s\rfloor-\tau_l}\|_{L^{(\varrho_a+2)p}(\Omega;\mathbb{R}^d)}$ of the previous inequality. Similarly to inequality (IV.26), it holds

$$\begin{aligned}&\sum_{l=0}^{D}\|X_{s-\tau_l}-X_{\lfloor s\rfloor-\tau_l}\|_{L^{(\varrho_a+2)p}(\Omega;\mathbb{R}^d)}\\ &\quad\leq\sum_{l=1}^{D}\|\xi_{(s-\tau_l)\wedge t_0}-\xi_{(\lfloor s\rfloor-\tau_l)\wedge t_0}\|_{L^{(\varrho_a+2)p}(\Omega;\mathbb{R}^d)}+\sum_{l=0}^{D}\|X_{(s-\tau_l)\vee t_0}-X_{(\lfloor s\rfloor-\tau_l)\vee t_0}\|_{L^{(\varrho_a+2)p}(\Omega;\mathbb{R}^d)}\end{aligned}\tag{IV.150}$$

for all $s\in[t_0,T]$ in view of equation (IV.25). Further, analogously to inequalities (IV.27) and (IV.28), Assumption IV.8 *vii)* and Lemma II.9 imply

$$\sum_{l=1}^{D}\|\xi_{(s-\tau_l)\wedge t_0}-\xi_{(\lfloor s\rfloor-\tau_l)\wedge t_0}\|_{L^{(\varrho_a+2)p}(\Omega;\mathbb{R}^d)}\leq DL_\xi\sqrt{T-t_0}\sqrt{s-\lfloor s\rfloor}\tag{IV.151}$$

and

$$\begin{aligned}&\sum_{l=0}^{D}\|X_{(s-\tau_l)\vee t_0}-X_{(\lfloor s\rfloor-\tau_l)\vee t_0}\|_{L^{(\varrho_a+2)p}(\Omega;\mathbb{R}^d)}\\ &\quad\leq(D+1)\big(K_a\sqrt{T-t_0}+\sqrt{(\varrho_a+2)p-1}K_b\sqrt{m}\big)\\ &\qquad\times\big(1+\|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\sqrt{s-\lfloor s\rfloor}\end{aligned}\tag{IV.152}$$

for all $s\in[t_0,T]$. Due to this and inequality (IV.65), we obtain

$$\begin{aligned}\mathcal{R}_6&\leq\frac{1}{4}K_{\partial^2 a}d\big(1+\|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_a}{2}}\\ &\quad\times\Big(DL_\xi\sqrt{T-t_0}+(D+1)\big(K_a\sqrt{T-t_0}+\sqrt{(\varrho_a+2)p-1}K_b\sqrt{m}\big)\\ &\qquad\times\big(1+\|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\Big)^2(T-t_0)\,h.\end{aligned}\tag{IV.153}$$

That is, term $\mathcal{R}_6$ is of order $\mathcal{O}(h)$ as $h\to 0$.

We continue with the next terms $\mathcal{R}_7$ and $\mathcal{R}_8$. In inequality (IV.56), we already proved

$$\mathcal{R}_7\leq\frac{p}{\sqrt{p-1}}L_b\sqrt{m}\bigg(\int_{t_0}^{T}\|X-Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}\,\mathrm{d}s\bigg)^{\frac{1}{2}},\tag{IV.154}$$

and further, similarly to inequality (IV.54), we have

$$\mathcal{R}_8 \leq \frac{p}{\sqrt{p-1}} L_{\mathrm{t},b} \sqrt{m} \big(1 + \|X\|^2_{S^{(\gamma_b \vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_b}{2}} \frac{1}{\sqrt{3}} \sqrt{T-t_0}\, h \tag{IV.155}$$

by taking Assumption IV.8 *vi)* and

$$\begin{aligned}\bigg(\int_{t_0}^{T} (s - \lfloor s \rfloor)^2 \,\mathrm{d}s\bigg)^{\frac{1}{2}} &= \bigg(\sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} (s-t_n)^2 \,\mathrm{d}s\bigg)^{\frac{1}{2}} = \bigg(\sum_{n=0}^{N-1} \frac{1}{3}(t_{n+1}-t_n)^3\bigg)^{\frac{1}{2}} \\ &\leq \frac{1}{\sqrt{3}} \sqrt{T-t_0}\, h \end{aligned}\tag{IV.156}$$

into account.

Terms $\mathcal{R}_9$ and $\mathcal{R}_{10}$ are estimated similarly to terms $\mathcal{R}_3$ and $\mathcal{R}_4$ in inequalities (IV.68) and (IV.69). We infer by Theorem II.6 and inequality (IV.156) that

$$\begin{aligned}\mathcal{R}_9 &\leq \frac{p}{\sqrt{p-1}} L_b \sqrt{m}\sqrt{d} \sum_{l=1}^{D} \bigg(\int_{t_0}^{T} \|\xi_{(s-\tau_l)\wedge t_0} - \xi_{(\lfloor s \rfloor-\tau_l)\wedge t_0}\|^2_{L^p(\Omega;\mathbb{R}^d)} \,\mathrm{d}s\bigg)^{\frac{1}{2}} \\ &\leq \frac{p}{\sqrt{p-1}} L_b \sqrt{m}\sqrt{d} L_\xi D \frac{1}{\sqrt{3}} \sqrt{T-t_0}\, h \end{aligned}\tag{IV.157}$$

and

$$\begin{aligned}\mathcal{R}_{10} &\leq \frac{p}{\sqrt{p-1}} L_b \sqrt{m}\sqrt{d} \sum_{l=0}^{D} \bigg(\int_{t_0}^{T} \bigg\| \int_{(\lfloor s \rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} a(\mathcal{T}(u, X_u)) \,\mathrm{d}u \bigg\|^2_{L^p(\Omega;\mathbb{R}^d)} \mathrm{d}s\bigg)^{\frac{1}{2}} \\ &\leq \frac{p}{\sqrt{p-1}} L_b \sqrt{m}\sqrt{d} K_a (D+1) \big(1 + \|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}} \frac{1}{\sqrt{3}} \sqrt{T-t_0}\, h. \end{aligned}\tag{IV.158}$$

Let us continue with term $\mathcal{R}_{11}$. At first, Zakai's inequality from Theorem II.6 and the triangle inequality imply

$$\begin{aligned}\mathcal{R}_{11} \leq \frac{p}{\sqrt{p-1}} \sum_{l=0}^{D} \bigg(\int_{t_0}^{T} \sum_{j_1=1}^{m} \bigg\| \sum_{j_2=1}^{m} \bigg| \int_{(\lfloor s \rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \bigg| \\ \times \bigg\| \sum_{i=1}^{d} \Big(\partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_{(\lfloor s \rfloor-\tau_l)\vee t_0})) \\ - \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor})) b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, Y_{(\lfloor s \rfloor-\tau_l)\vee t_0}))\Big) \bigg\| \bigg\|^2_{L^p(\Omega;\mathbb{R})} \mathrm{d}s\bigg)^{\frac{1}{2}}. \end{aligned}\tag{IV.159}$$

According to the global Lipschitz condition in Assumption IV.8 *iii)*, it holds for the inner Euclidean norm on the right-hand side of inequality (IV.159) above that

$$\begin{aligned}\max_{\substack{j_1,j_2\in\{1,\dots,m\}\\ l\in\{0,1,\dots,D\}}} \bigg\| \sum_{i=1}^{d} \Big(\partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor})) b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_{(\lfloor s \rfloor-\tau_l)\vee t_0})) \\ - \partial_{x_l^i} b^{j_1}(\mathcal{T}(\lfloor s \rfloor, Y_{\lfloor s \rfloor})) b^{i,j_2}(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, Y_{(\lfloor s \rfloor-\tau_l)\vee t_0}))\Big) \bigg\| \end{aligned}\tag{IV.160}$$

$$\leq L_{\partial b} \Big(\sup_{t\in[t_0-\tau,T]} \big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{\beta}{2}} \Big) \Big(\sup_{t\in[t_0-\tau,s]} \|X_t - Y_t\| \Big)$$

for all $s \in [t_0, T]$. Considering the last factor on the right-hand side of inequality (IV.160) above and using inequality (II.7), it holds

$$\|X_t - Y_t\| \leq \big(\|X_t\| + \|Y_t\|\big)^{2\cdot\frac{1}{2}} \leq \sqrt{2}\big(\|X_t\|^2 + \|Y_t\|^2\big)^{\frac{1}{2}} \leq \sqrt{2}\big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{1}{2}}$$

for all $t \in [t_0 - \tau, T]$, and thus, we obtain

$$\sup_{t\in[t_0-\tau,s]} \|X_t - Y_t\| \leq 2^{\frac{1}{4}} \bigg(\sup_{t\in[t_0-\tau,T]} \big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{1}{4}}\bigg)\bigg(\sup_{t\in[t_0-\tau,s]} \|X_t - Y_t\|\bigg)^{\frac{1}{2}} \quad \text{(IV.161)}$$

for all $s \in [t_0, T]$. Inserting inequalities (IV.160) and (IV.161) into inequality (IV.159), we obtain

$$\begin{aligned}\mathcal{R}_{11} \leq \frac{p}{\sqrt{p-1}} 2^{\frac{1}{4}} L_{\partial b} \sqrt{m} \sum_{l=0}^{D} \bigg(\int_{t_0}^{T} \bigg\| \sum_{j_2=1}^{m} \bigg| \int_{(\lfloor s\rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \bigg| \\ \times \bigg(\sup_{t\in[t_0-\tau,T]} \big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{2\beta+1}{4}} \bigg) \bigg(\sup_{t\in[t_0-\tau,s]} \|X_t - Y_t\| \bigg)^{\frac{1}{2}} \bigg\|_{L^p(\Omega;\mathbb{R})}^2 \mathrm{d}s \bigg)^{\frac{1}{2}}. \end{aligned} \quad \text{(IV.162)}$$

In the following, we apply inequality (II.7) again in order to separate the term

$$\bigg(\int_{t_0}^{T} \|X - Y\|_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R}^d)}^2 \,\mathrm{d}s \bigg)^{\frac{1}{2}}$$

that contributes to constant C_2 in inequality (IV.62). According to Gronwall's Lemma II.7, constant C_2 has an exponential influence on the estimate. Due to this, we apply the inequality

$$c_1 c_2 = c_1 \gamma \gamma^{-1} c_2 \leq \frac{1}{2}\gamma^2 c_1^2 + \frac{1}{2}\gamma^{-2} c_2^2 \quad \text{(IV.163)}$$

with $\gamma = \big(\frac{p}{\sqrt{p-1}} 2^{\frac{1}{4}} L_{\partial b}\sqrt{m}(D+1)\big)^{\frac{1}{2}}$ instead of inequality (II.7). Applying inequality (IV.163) to the argument of the $L^p(\Omega;\mathbb{R})$-norm in inequality (IV.162), where

$$c_2 = \bigg(\sup_{t\in[t_0-\tau,s]} \|X_t - Y_t\| \bigg)^{\frac{1}{2}},$$

we obtain

$$\begin{aligned}\mathcal{R}_{11} \leq \frac{p}{\sqrt{p-1}} 2^{\frac{1}{4}} L_{\partial b} \sqrt{m} \sum_{l=0}^{D} \bigg(\int_{t_0}^{T} \bigg\| \frac{1}{2}\gamma^2 \bigg(\sum_{j_2=1}^{m} \bigg| \int_{(\lfloor s\rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \bigg| \\ \times \bigg(\sup_{t\in[t_0-\tau,T]} \big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{2\beta+1}{4}} \bigg)\bigg)^2 + \frac{1}{2}\gamma^{-2} \bigg(\sup_{t\in[t_0-\tau,s]} \|X_t - Y_t\| \bigg) \bigg\|_{L^p(\Omega;\mathbb{R})}^2 \mathrm{d}s \bigg)^{\frac{1}{2}}. \end{aligned}$$

Using the triangle inequality and rearranging the terms, it further follows

$$
\begin{aligned}
\mathcal{R}_{11} \leq{} & \frac{p}{\sqrt{p-1}} 2^{\frac{1}{4}} L_{\partial b} \sqrt{m} \sum_{l=0}^{D} \bigg(\int_{t_0}^{T} \bigg\| \frac{1}{2}\gamma^2 \Big(\sum_{j_2=1}^{m} \Big| \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \Big| \\
& \times \Big(\sup_{t\in[t_0-\tau,T]} \big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{2\beta+1}{4}} \Big) \Big)^2 \bigg\|_{L^p(\Omega;\mathbb{R})}^2 \mathrm{d}s \bigg)^{\frac{1}{2}} \\
& + \frac{p}{\sqrt{p-1}} 2^{\frac{1}{4}} L_{\partial b} \sqrt{m} \sum_{l=0}^{D} \bigg(\int_{t_0}^{T} \bigg\| \frac{1}{2}\gamma^{-2} \Big(\sup_{t\in[t_0-\tau,s]} \|X_t - Y_t\| \Big) \bigg\|_{L^p(\Omega;\mathbb{R})}^2 \mathrm{d}s \bigg)^{\frac{1}{2}} \\
={} & \frac{1}{2}\gamma^2 \frac{p}{\sqrt{p-1}} 2^{\frac{1}{4}} L_{\partial b} \sqrt{m} \sum_{l=0}^{D} \bigg(\int_{t_0}^{T} \bigg\| \sum_{j_2=1}^{m} \Big| \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \Big| \\
& \times \Big(\sup_{t\in[t_0-\tau,T]} \big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{2\beta+1}{4}} \Big) \bigg\|_{L^{2p}(\Omega;\mathbb{R})}^4 \mathrm{d}s \bigg)^{\frac{1}{2}} \\
& + \frac{1}{2}\gamma^{-2} \frac{p}{\sqrt{p-1}} 2^{\frac{1}{4}} L_{\partial b} \sqrt{m}(D+1) \bigg(\int_{t_0}^{T} \|X - Y\|_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R})}^2 \mathrm{d}s \bigg)^{\frac{1}{2}} \\
={} & \frac{p^2}{\sqrt{2}(p-1)} L_{\partial b}^2 m(D+1) \sum_{l=0}^{D} \bigg(\int_{t_0}^{T} \bigg\| \sum_{j_2=1}^{m} \Big| \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \Big| \\
& \times \Big(\sup_{t\in[t_0-\tau,T]} \big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{2\beta+1}{4}} \Big) \bigg\|_{L^{2p}(\Omega;\mathbb{R})}^4 \mathrm{d}s \bigg)^{\frac{1}{2}} \\
& + \frac{1}{2} \bigg(\int_{t_0}^{T} \|X - Y\|_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R})}^2 \,\mathrm{d}s \bigg)^{\frac{1}{2}}.
\end{aligned}
\tag{IV.164}
$$

Next, we only consider the integrand of the first term on the right-hand side of inequality (IV.164) above and show that it is of order $\mathcal{O}(h)$ as $h \to 0$. Using Hölder's inequality with $\frac{1}{2(\beta+1)} + \frac{2\beta+1}{2(\beta+1)} = 1$ and the triangle inequality, we have

$$
\begin{aligned}
& \bigg\| \sum_{j_2=1}^{m} \Big| \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \Big| \Big(\sup_{t\in[t_0-\tau,T]} \big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{2\beta+1}{4}} \Big) \bigg\|_{L^{2p}(\Omega;\mathbb{R})}^4 \\
& \leq \bigg\| \sum_{j_2=1}^{m} \Big| \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \Big| \bigg\|_{L^{4(\beta+1)p}(\Omega;\mathbb{R})}^4 \bigg\| \sup_{t\in[t_0-\tau,T]} \big(1 + \|X_t\|^2 + \|Y_t\|^2\big)^{\frac{2\beta+1}{4}} \bigg\|_{L^{\frac{4(\beta+1)}{2\beta+1}p}(\Omega;\mathbb{R})}^4 \\
& = \bigg\| \sum_{j_2=1}^{m} \Big| \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \Big| \bigg\|_{L^{4(\beta+1)p}(\Omega;\mathbb{R})}^4 \big\|1 + \|X\|^2 + \|Y\|^2\big\|_{S^{(\beta+1)p}([t_0-\tau,T]\times\Omega;\mathbb{R})}^{2\beta+1} \\
& \leq \bigg(\sum_{j_2=1}^{m} \Big\| \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \Big\|_{L^{4(\beta+1)p}(\Omega;\mathbb{R})} \bigg)^4 \\
& \quad \times \big(1 + \|X\|_{S^{2(\beta+1)p}([t_0-\tau,T]\times\Omega;\mathbb{R})}^2 + \|Y\|_{S^{2(\beta+1)p}([t_0-\tau,T]\times\Omega;\mathbb{R})}^2\big)^{2\beta+1}.
\end{aligned}
\tag{IV.165}
$$

Since

$$
\bigg(\sum_{j_2=1}^{m} \Big\| \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \mathrm{d}W_u^{j_2} \Big\|_{L^{4(\beta+1)p}(\Omega;\mathbb{R})} \bigg)^4 \leq m^4 \big(4(\beta+1)p - 1\big)^2 (s - \lfloor s \rfloor)^2
$$

for all $s \in [t_0, T]$ by Theorem II.6, inserting inequality (IV.165) into inequality (IV.164) and using inequality (IV.156) finally yield

$$
\begin{aligned}
\mathcal{R}_{11} &\leq \frac{p^2}{\sqrt{2}(p-1)} L_{\partial b}^2 m^3 (D+1)^2 \big(4(\beta+1)p-1\big) \bigg(\int_{t_0}^{T} (s - \lfloor s \rfloor)^2 \, \mathrm{d}s \bigg)^{\frac{1}{2}} \\
&\qquad \times \big(1 + \|X\|^2_{S^{2(\beta+1)p}([t_0-\tau,T]\times\Omega;\mathbb{R})} + \|Y\|^2_{S^{2(\beta+1)p}([t_0-\tau,T]\times\Omega;\mathbb{R})}\big)^{\frac{2\beta+1}{2}} \\
&\quad + \frac{1}{2} \bigg(\int_{t_0}^{T} \|X - Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R})} \, \mathrm{d}s \bigg)^{\frac{1}{2}} \\
&\leq \frac{p^2}{\sqrt{6}(p-1)} L_{\partial b}^2 m^3 (D+1)^2 \big(4(\beta+1)p-1\big) \sqrt{T - t_0}\, h \\
&\qquad \times \big(1 + \|X\|^2_{S^{2(\beta+1)p}([t_0-\tau,T]\times\Omega;\mathbb{R})} + \|Y\|^2_{S^{2(\beta+1)p}([t_0-\tau,T]\times\Omega;\mathbb{R})}\big)^{\frac{2\beta+1}{2}} \\
&\quad + \frac{1}{2} \bigg(\int_{t_0}^{T} \|X - Y\|^2_{S^p([t_0-\tau,s]\times\Omega;\mathbb{R})} \, \mathrm{d}s \bigg)^{\frac{1}{2}}.
\end{aligned} \tag{IV.166}
$$

That is, term $\mathcal{R}_{11}$ contributes to both constants C_1 and C_2 in inequality (IV.62).

We now consider term $\mathcal{R}_{12}$. Similarly to estimates (IV.157) and (IV.158) of terms $\mathcal{R}_9$ and $\mathcal{R}_{10}$, Zakai's inequality from Theorem II.6, the triangle inequality, Assumption IV.8 *ii)*, and inequality (IV.67) imply

$$
\begin{aligned}
\mathcal{R}_{12} &\leq \frac{p}{\sqrt{p-1}} L_b \sqrt{m} \sqrt{d} \sum_{l=0}^{D} \\
&\quad \times \bigg(\int_{t_0}^{T} \bigg\| \sum_{j=1}^{m} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} b^j(\mathcal{T}(u, X_u)) - b^j(\mathcal{T}((\lfloor s \rfloor - \tau_l)\vee t_0, X_u)) \, \mathrm{d}W_u^j \bigg\|^2_{L^p(\Omega;\mathbb{R}^d)} \mathrm{d}s \bigg)^{\frac{1}{2}}.
\end{aligned}
$$

Using again Theorem II.6 and the triangle inequality, we obtain by Assumption IV.8 *vi)* that

$$
\begin{aligned}
\mathcal{R}_{12} &\leq p L_b \sqrt{m} \sqrt{d} \sum_{l=0}^{D} \\
&\quad \times \bigg(\int_{t_0}^{T} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \sum_{j=1}^{m} \big\| b^j(\mathcal{T}(u, X_u)) - b^j(\mathcal{T}((\lfloor s \rfloor - \tau_l) \vee t_0, X_u)) \big\|^2_{L^p(\Omega;\mathbb{R}^d)} \mathrm{d}u \, \mathrm{d}s \bigg)^{\frac{1}{2}} \\
&\leq p L_b m \sqrt{d} L_{\mathrm{t},b} \sum_{l=0}^{D} \bigg(\int_{t_0}^{T} \int_{(\lfloor s \rfloor - \tau_l)\vee t_0}^{(s-\tau_l)\vee t_0} \bigg\| \sup_{k\in\{0,1,\dots,D\}} \big(1 + \|X_{u-\tau_k}\|^2\big)^{\frac{\gamma_b}{2}} \bigg\|^2_{L^p(\Omega;\mathbb{R})} \\
&\qquad \times \big(u - ((\lfloor s \rfloor - \tau_l) \vee t_0)\big)^2 \, \mathrm{d}u \, \mathrm{d}s \bigg)^{\frac{1}{2}}.
\end{aligned} \tag{IV.167}
$$

Considering the $L^p(\Omega; \mathbb{R})$-norm on the right-hand side of inequality (IV.167) above, it analogously to inequality (IV.52) holds

$$
\bigg\| \sup_{k\in\{0,1,\dots,D\}} \big(1 + \|X_{u-\tau_k}\|^2\big)^{\frac{\gamma_b}{2}} \bigg\|_{L^p(\Omega;\mathbb{R})} \leq \big(1 + \|X\|^2_{S^{(\gamma_b\vee 1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_b}{2}}.
$$

Inserting this into inequality (IV.167) and using

$$\begin{aligned}
&\Big(\int_{t_0}^{T}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}\big(u-(\lfloor s\rfloor-\tau_l)\vee t_0\big)^2\,\mathrm{d}u\,\mathrm{d}s\Big)^{\frac{1}{2}}\\
&\quad=\Big(\int_{t_0}^{T}\frac{1}{3}\big(((\lfloor s\rfloor-\tau_l)\vee t_0)-((\lfloor s\rfloor-\tau_l)\vee t_0)\big)^3\,\mathrm{d}s\Big)^{\frac{1}{2}}\\
&\quad\le\Big(\int_{t_0}^{T}\frac{1}{3}(s-\lfloor s\rfloor)^3\,\mathrm{d}s\Big)^{\frac{1}{2}}\\
&\quad=\frac{1}{\sqrt{3}}\Big(\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}(s-t_n)^3\,\mathrm{d}s\Big)^{\frac{1}{2}}\\
&\quad=\frac{1}{\sqrt{12}}\Big(\sum_{n=0}^{N-1}(t_{n+1}-t_n)^4\Big)^{\frac{1}{2}}\\
&\quad\le\frac{1}{\sqrt{12}}\sqrt{T-t_0}\,h^{\frac{3}{2}}\\
&\quad\le\frac{1}{\sqrt{12}}(T-t_0)\,h, \qquad\qquad \text{(IV.168)}
\end{aligned}$$

we finally obtain

$$\mathcal{R}_{12}\le pL_bm\sqrt{d}L_{\mathrm{t},b}(D+1)\big(1+\|X\|^2_{S^{(\gamma_b\vee1)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\gamma_b}{2}}\frac{1}{\sqrt{12}}(T-t_0)\,h. \qquad \text{(IV.169)}$$

Similarly to the considerations on previous term $\mathcal{R}_{12}$, it holds, using Theorem II.6, the triangle inequality, Assumption IV.8 *ii)*, and inequality (IV.67), that

$$\begin{aligned}
\mathcal{R}_{13}\le\frac{p}{\sqrt{p-1}}L_b\sqrt{m}\sqrt{d}\sum_{l=0}^{D}\Big(\int_{t_0}^{T}\Big\|\sum_{j=1}^{m}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}&\Big(b^j(\mathcal{T}((\lfloor s\rfloor-\tau_l)\vee t_0,X_u))\\
&-b^j(\mathcal{T}((\lfloor s\rfloor-\tau_l)\vee t_0,X_{(\lfloor s\rfloor-\tau_l)\vee t_0}))\Big)\,\mathrm{d}W_u^j\Big\|^2_{L^p(\Omega;\mathbb{R}^d)}\mathrm{d}s\Big)^{\frac{1}{2}}.
\end{aligned}$$

Using Theorem II.6 and Assumption IV.8 *ii)* again, similarly to inequalities (IV.24) and (IV.167), we have

$$\mathcal{R}_{13}\le pL_b^2m\sqrt{d}\sum_{l=0}^{D}\Big(\int_{t_0}^{T}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}\Big(\sum_{k=0}^{D}\|X_{u-\tau_k}-X_{((\lfloor s\rfloor-\tau_l)\vee t_0)-\tau_k}\|_{L^p(\Omega;\mathbb{R}^d)}\Big)^2\mathrm{d}u\,\mathrm{d}s\Big)^{\frac{1}{2}}.$$

Then, similar considerations to inequalities (IV.29), (IV.150), (IV.151), and (IV.152) imply

$$\begin{aligned}
\mathcal{R}_{13}\le pL_b^2m\sqrt{d}\Big(&L_\xi D\sqrt{T-t_0}+(D+1)\big(K_a\sqrt{T-t_0}+\sqrt{p-1}K_b\sqrt{m}\big)\\
&\times\big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\Big)\sum_{l=0}^{D}\Big(\int_{t_0}^{T}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}\big(u-((\lfloor s\rfloor-\tau_l)\vee t_0)\big)\,\mathrm{d}u\,\mathrm{d}s\Big)^{\frac{1}{2}}.
\end{aligned}$$

Finally, since

$$\Big(\int_{t_0}^{T}\int_{(\lfloor s\rfloor-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}\big(u-((\lfloor s\rfloor-\tau_l)\vee t_0)\big)\,\mathrm{d}u\,\mathrm{d}s\Big)^{\frac{1}{2}}\le\frac{1}{\sqrt{6}}\sqrt{T-t_0}\,h,$$

cf. inequality (IV.168), we obtain

$$\begin{aligned}\mathcal{R}_{13} \le pL_b^2 m\sqrt{d}\Big(L_\xi D\sqrt{T-t_0} + \big(K_a\sqrt{T-t_0} + \sqrt{p-1}K_b\sqrt{m}\big)\\ \times \big(1+\|X\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}(D+1)\Big)(D+1)\frac{1}{\sqrt{6}}\sqrt{T-t_0}\,h.\end{aligned} \tag{IV.170}$$

Now, we consider the last term $\mathcal{R}_{14}$ and show that it is of order $\mathcal{O}(h)$ as $h \to 0$, too. First, Zakai's inequality from Theorem II.6 and the triangle inequality imply

$$\begin{aligned}\mathcal{R}_{14} \le \frac{p}{\sqrt{p-1}}\Bigg(\int_{t_0}^T \sum_{j=1}^m \Bigg\| \sum_{l_1,l_2=0}^D \sum_{i_1,i_2=1}^d \int_0^1 \big\|\partial_{x^{i_1}_{l_1}}\partial_{x^{i_2}_{l_2}} b^j(\mathcal{T}(\lfloor s\rfloor, X_{\lfloor s\rfloor}+\theta(X_s-X_{\lfloor s\rfloor})))\big\|(1-\theta)\,\mathrm{d}\theta\\ \times |X^{i_1}_{s-\tau_{l_1}} - X^{i_1}_{\lfloor s\rfloor-\tau_{l_1}}||X^{i_2}_{s-\tau_{l_2}} - X^{i_2}_{\lfloor s\rfloor-\tau_{l_2}}|\Bigg\|^2_{L^p(\Omega;\mathbb{R})}\mathrm{d}s\Bigg)^{\frac{1}{2}}.\end{aligned}$$

The $L^p(\Omega;\mathbb{R})$-norm on the right-hand side of the inequality above can be treated analogously to the $L^p(\Omega;\mathbb{R})$-norm in inequality (IV.148), which occurs in the estimates of term $\mathcal{R}_6$. Then, similarly to inequality (IV.149), we have

$$\begin{aligned}\mathcal{R}_{14} &\le \frac{1}{2}\frac{p}{\sqrt{p-1}}K_{\partial^2 b}\sqrt{m}d\big(1+\|X\|^2_{S^{(\varrho_b+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_b}{2}}\\ &\quad\times\Bigg(\int_{t_0}^T\Bigg(\sum_{l_1,l_2=0}^D \big\|\|X_{s-\tau_{l_1}} - X_{\lfloor s\rfloor-\tau_{l_1}}\|\|X_{s-\tau_{l_2}} - X_{\lfloor s\rfloor-\tau_{l_2}}\|\big\|_{L^{\frac{\varrho_b+2}{2}p}(\Omega;\mathbb{R})}\Bigg)^2\mathrm{d}s\Bigg)^{\frac{1}{2}}\\ &\le \frac{1}{2}\frac{p}{\sqrt{p-1}}K_{\partial^2 b}\sqrt{m}d\big(1+\|X\|^2_{S^{(\varrho_b+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_b}{2}}\\ &\quad\times\Bigg(\int_{t_0}^T\Bigg(\sum_{l=0}^D\|X_{s-\tau_l} - X_{\lfloor s\rfloor-\tau_l}\|_{L^{(\varrho_b+2)p}(\Omega;\mathbb{R}^d)}\Bigg)^4\mathrm{d}s\Bigg)^{\frac{1}{2}}.\end{aligned}$$

Using further inequalities (IV.150), (IV.151), (IV.152), and (IV.156), we obtain

$$\begin{aligned}\mathcal{R}_{14} &\le \frac{1}{2}\frac{p}{\sqrt{p-1}}K_{\partial^2 b}\sqrt{m}d\big(1+\|X\|^2_{S^{(\varrho_b+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_b}{2}}\\ &\quad\times\Big(DL_\xi\sqrt{T-t_0} + (D+1)\big(K_a\sqrt{T-t_0} + \sqrt{(\varrho_a+2)p-1}K_b\sqrt{m}\big)\\ &\qquad\times\big(1+\|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\Big)^2\Bigg(\int_{t_0}^T(s-\lfloor s\rfloor)^2\,\mathrm{d}s\Bigg)^{\frac{1}{2}}\\ &\le \frac{1}{2}\frac{p}{\sqrt{p-1}}K_{\partial^2 b}\sqrt{m}d\big(1+\|X\|^2_{S^{(\varrho_b+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{\varrho_b}{2}}\\ &\quad\times\Big(DL_\xi\sqrt{T-t_0} + (D+1)\big(K_a\sqrt{T-t_0} + \sqrt{(\varrho_a+2)p-1}K_b\sqrt{m}\big)\\ &\qquad\times\big(1+\|X\|^2_{S^{(\varrho_a+2)p}([t_0-\tau,T]\times\Omega;\mathbb{R}^d)}\big)^{\frac{1}{2}}\Big)^2\frac{1}{\sqrt{3}}\sqrt{T-t_0}h,\end{aligned} \tag{IV.171}$$

cf. inequality (IV.153).

Now, we have estimated all terms $\mathcal{R}_r$, $r \in \{1,\dots,14\}$. We refer to Table IV.17 for an overview. According to these estimates, there exist constants $C_1, C_2 > 0$, independent of h, such that inequality (IV.62) holds true. Using inequality (IV.63), Gronwall's Lemma II.7 implies

$$\|X-Y\|^2_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \le 2C_1^2h^2\mathrm{e}^{2C_2^2(T-t_0)}.$$

Thus, it holds

$$\|X - Y\|_{S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)} \leq \sqrt{2} C_1 \mathrm{e}^{C_2^2(T-t_0)} h$$

for all $h \in \,]0, T - t_0]$, which proves the assertion of Theorem IV.9. □

V

Efficient Approximation of Iterated Stochastic Integrals

Higher order approximations of solutions of SDEs, as the Milstein scheme, involve iterated stochastic integrals [23, 78, 105]. However, these approximations can only be simulated directly in special cases [78, 105].

Let us consider Milstein scheme (IV.33) regarding SDDE (II.1). In the case of *additive* noise, that is, the diffusion coefficients do not depend on solution X, the derivatives of the diffusion coefficients vanish, and the Milstein scheme equals the Euler-Maruyama scheme, cf. Corollary IV.13 and Corollary IV.14. Thus, the iterated stochastic integrals do not appear in the scheme.

SDEs with *commutative* noise are another important class of SDEs, where the Milstein scheme can be simulated directly. If the diffusion coefficients do not depend on the past history of solution X, and SDDE (II.1) satisfies the *commutativity condition*

$$\begin{aligned}&\partial_{x_0^i} b^{j_1}(t, t-\tau_1, \dots, t-\tau_D, X_t)\, b^{i,j_2}(t, t-\tau_1, \dots, t-\tau_D, X_t)\\ &\quad = \partial_{x_0^i} b^{j_2}(t, t-\tau_1, \dots, t-\tau_D, X_t)\, b^{i,j_1}(t, t-\tau_1, \dots, t-\tau_D, X_t)\end{aligned} \tag{V.1}$$

for $j_1, j_2 \in \{1, \dots, m\}$, $i \in \{1, \dots, d\}$, and $t \in [t_0, T]$, the iterated stochastic integrals in the Milstein scheme simplify to

$$\int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}W_u^j \, \mathrm{d}W_s^j = \frac{1}{2}\left(\left(\int_{t_n}^{t_{n+1}} \mathrm{d}W_u^j\right)^2 - (t_{n+1} - t_n)\right) \tag{V.2}$$

P-almost surely for $j \in \{1, \dots, m\}$ and

$$\int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}W_u^{j_2} \, \mathrm{d}W_s^{j_1} + \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}W_u^{j_1} \, \mathrm{d}W_s^{j_2} = \int_{t_n}^{t_{n+1}} \mathrm{d}W_s^{j_1} \int_{t_n}^{t_{n+1}} \mathrm{d}W_u^{j_2} \tag{V.3}$$

P-almost surely for $j_1, j_2 \in \{1, \dots, m\}$ with $j_1 \neq j_2$. These equations (V.2) and (V.3) follow from the stochastic integration by parts formula resulting from Itô's Lemma, cf. [78]. Thus, the Milstein scheme can be implemented by only simulating the normally distributed increments of the underlying Wiener process. Similar conditions to commutativity condition (V.1), in the case of present delay in the diffusion coefficients, do not simplify the delayed-iterated stochastic integrals because the stochastic integration by parts formula is not applicable, and because the delayed-iterated stochastic integrals do not commute. This even holds in case of one-dimensional noise ($m = 1$). As a consequence, modeling (delayed-)iterated stochastic integrals is an important task in order to make numerical schemes of higher order applicable.

The modeling of iterated stochastic integrals is closely related to the approximation of Lévy's area [91]. This problem has been studied by different authors, see, e. g., [42, 60, 78, 79, 96, 105, 127, 136, 137]. Since we are interested in modeling iterated stochastic integrals in case of multidimensional noise in general, that is, $m \in \mathbb{N}$ is arbitrary, the results in [42, 96, 127] are not further discussed below, because the Wiener process is only considered to be two-dimensional ($m = 2$) there. Considering iterated stochastic integrals where the Wiener process is multidimensional, approximation schemes are developed based on a series expansion of the Brownian bridge process for SODEs in [78, 79, 105] and for SDDEs in [60, 137]. The results in [78, 79, 105] were improved by Wiktorsson in [136] and generalized by Leonhard and Rößler in [90] to Q-Wiener processes driving SPDEs. In all these papers, the approximation of Levy's area is considered in the $L^2(\Omega;\mathbb{R})$-norm.

In Section V.1, the results from [60, 78, 79, 105] are extended to convergence in $L^p(\Omega;\mathbb{R})$ for arbitrary $p \in [2,\infty[$. Further, we show in case of SDDEs that the computational cost of the Milstein scheme is significantly reduced compared to [60, p. 311], see Theorem V.18 in Section V.4.

In Section V.2, a new algorithm is proposed that significantly reduces the number of normally distributed random variables, that need to be generated, compared to Wiktorsson's algorithm in [136]. Whereas Wiktorsson only analyzed the convergence of his algorithm in $L^2(\Omega;\mathbb{R})$, we show that our new algorithm is convergent in $L^p(\Omega;\mathbb{R})$ for all $p \in [2,\infty[$. The computational costs of this algorithm are compared to the algorithm from Section V.1 and to the one of Wiktorsson [136] in Section V.3. The convergence of the Milstein scheme based on these iterated stochastic integral approximations is stated in Theorem V.19 in Section V.4.

The convergence in $L^p(\Omega;\mathbb{R})$ for all $p \in [2,\infty[$ is especially relevant for pathwise approximations of SDEs that are of higher order, cf. Corollary IV.12 and [77], and may also be of interest for multilevel Monte-Carlo approximations with irregular functionals, cf. [8].

In the following, we first consider some problems of dependencies occurring in the simulation of delayed-iterated stochastic integrals when the discretization is arbitrary. Afterwards, the Fourier series expansion of the Brownian bridge is used to derive expansions of the iterated stochastic integrals.

Let $\{\tilde{t}_0, \tilde{t}_1, \ldots, \tilde{t}_{\tilde{N}-1}, \tilde{t}_{\tilde{N}}\}$ be an arbitrary discretization of the interval $[t_0, T]$ where $t_0 =: \tilde{t}_0 < \tilde{t}_1 < \ldots < \tilde{t}_{\tilde{N}-1} < \tilde{t}_{\tilde{N}} := T$. Consider points in time $\tilde{t}_n, \tilde{t}_{n+1}$, and let $\tilde{t}_n - \tau_l < \tilde{t}_k < \tilde{t}_{n+1} - \tau_l$ where $\tilde{t}_n - \tau_l \geq t_0$. We are interested in simulating the stochastic integral $\int_{\tilde{t}_n}^{\tilde{t}_{n+1}} \int_{\tilde{t}_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j$. Considering point in time $\tilde{t}_k$, we can rewrite the delayed-iterated stochastic integral to

$$\begin{aligned} &\int_{\tilde{t}_n}^{\tilde{t}_{n+1}} \int_{\tilde{t}_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j \\ &\quad = \int_{\tilde{t}_n}^{\tilde{t}_k+\tau_l} \int_{\tilde{t}_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j + \int_{\tilde{t}_k+\tau_l}^{\tilde{t}_{n+1}} \mathrm{d}W_s^j \int_{\tilde{t}_n-\tau_l}^{\tilde{t}_k} \mathrm{d}W_u^i + \int_{\tilde{t}_k+\tau_l}^{\tilde{t}_{n+1}} \int_{\tilde{t}_k}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j \end{aligned} \tag{V.4}$$

P-almost surely. Thus, the problem of simulating the delayed-iterated stochastic integral splits up to simulating the increments $\int_{\tilde{t}_n-\tau_l}^{\tilde{t}_k} \mathrm{d}W_u^i$ and $\int_{\tilde{t}_k+\tau_l}^{\tilde{t}_{n+1}} \mathrm{d}W_s^j$ as well as the delayed-iterated stochastic integrals $\int_{\tilde{t}_n}^{\tilde{t}_k+\tau_l} \int_{\tilde{t}_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j$ and $\int_{\tilde{t}_k+\tau_l}^{\tilde{t}_{n+1}} \int_{\tilde{t}_k}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j$. These random variables are taken into account automatically by adding point in time $\tilde{t}_k + \tau_l \in \,]\tilde{t}_n, \tilde{t}_{n+1}[$ to the discretization.

Hu, Mohammed, and Yan hide this problem of the dependencies in equation (V.4), cf. [60, Appendix B]. Thus, their algorithm in [60, (B.7) and (B.8)] is not applicable and implementable straightforwardly.

A similar problem of dependencies occurs if we already have simulated the iterated stochastic integral $\int_{\tilde{t}_n}^{\tilde{t}_{n+1}} \int_{\tilde{t}_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j$, and we like to add a point in time $t_{n+1} \in \,]\tilde{t}_n, \tilde{t}_{n+1}[$ to the discretization a posteriori. Then, similarly to equation (V.4), it P-almost surely holds

$$\begin{aligned} &\int_{\tilde{t}_n}^{\tilde{t}_{n+1}} \int_{\tilde{t}_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j \\ &= \int_{\tilde{t}_n}^{t_{n+1}} \int_{\tilde{t}_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j + \int_{t_{n+1}}^{\tilde{t}_{n+1}} \mathrm{d}W_s^j \int_{\tilde{t}_n-\tau_l}^{t_{n+1}-\tau_l} \mathrm{d}W_u^i + \int_{t_{n+1}}^{\tilde{t}_{n+1}} \int_{t_{n+1}-\tau_l}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j. \end{aligned} \tag{V.5}$$

By adding the point in time $t_{n+1} - \tau_l$ to the discretization a priori, the problem of dependencies of the random variables can be circumvented again. Note that, in case of SDDEs, explicit schemes, like the Milstein method, need to compute the approximation at the point in time $t_{n+1} - \tau_l$, and thus, the random variables on the right-hand side of equations (V.4) and (V.5) are needed anyhow.

Taking this into account, we refine the given discretization $\{\tilde{t}_0, \tilde{t}_1, \ldots, \tilde{t}_{\tilde{N}-1}, \tilde{t}_{\tilde{N}}\}$ to

$$\{t_0, t_1, \ldots, t_N\} := \bigcup_{n=0}^{\tilde{N}} \bigcup_{z \in \mathbb{Z}^D} \left\{ \tilde{t}_n + \sum_{l=1}^{D} z_l \tau_l \right\} \cap [t_0, T] \tag{V.6}$$

whenever $D > 0$. If $D = 0$, we emphasize that

$$\{t_0, t_1, \ldots, t_N\} := \{\tilde{t}_0, \tilde{t}_1, \ldots, \tilde{t}_{\tilde{N}-1}, \tilde{t}_{\tilde{N}}\},$$

and the discretization is still arbitrary.

Using discretization (V.6), we have, on the one hand, the opportunity to compute the Milstein scheme directly on this discretization. On the other hand, we can calculate the Milstein scheme on the prior discretization $\{\tilde{t}_0, \tilde{t}_1, \ldots, \tilde{t}_{\tilde{N}-1}, \tilde{t}_{\tilde{N}}\}$. Afterwards, we use the random variables on the right-hand side of equations (V.4) and (V.5), that are simulated using the discretization (V.6), in order to compute for example $Y_{\tilde{t}_n-\tau_l}$ via the continuity – interpolation – of the Milstein scheme (IV.33), where the point in time $\tilde{t}_n - \tau_l$ does belong to that discretization (V.6).

The dependencies occurring in equations (V.4) and (V.5) make it much more complicated to add a point to the discretization a posteriori than it is in the case of the Euler-Maruyama scheme, cf. [2, p. 24].

Therefore, throughout this chapter, it is assumed that the discretization is of form (V.6) whenever $D > 0$.

An example of a discretization that satisfies equation (V.6) in case of $D = 1$ is as follows. Set $h = \tau_1/M$ for some $M \in \mathbb{N}$, and if $T = Nh$, let $t_n = t_0 + nh$ for $n \in \{0, 1, \ldots, N\}$ be the points of that equidistant discretization.

A discretization of the form (V.6) does not have to be equidistant necessarily. In case of $D = 1$ with $\tau_1 = 3$, a discretization with $h_{2n} = 1$ and $h_{2n+1} = 2$ for $n \in \{0, 1, \ldots, N\}$ provides an example that satisfies (V.6) and is not equidistant.

For the sake of simplicity, let us introduce the notations

$$\Delta W^j_{n,\tau_l} := \begin{cases} \int_{t_n-\tau_l}^{t_{n+1}-\tau_l} \mathrm{d}W^j_s & \text{if } t_n - \tau_l \geq t_0 \text{ and} \\ 0 & \text{if } t_n - \tau_l < t_0 \end{cases}$$

as well as

$$I_{(i,j),n,\tau_l} := \begin{cases} \int_{t_n}^{t_{n+1}} \int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W^i_u \, \mathrm{d}W^j_s & \text{if } t_n - \tau_l \geq t_0 \text{ and} \\ 0 & \text{if } t_n - \tau_l < t_0 \end{cases}$$

for $i, j \in \{1, \ldots, m\}$, $l \in \{0, 1, \ldots, D\}$, and $n \in \{0, 1, \ldots, N-1\}$. Further, in case of $l = 0$, we write

$$\Delta W^j_n := \Delta W^j_{n,\tau_0} = \int_{t_n}^{t_{n+1}} \mathrm{d}W^j_s$$

and

$$I_{(i,j),n} := I_{(i,j),n,\tau_0} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}W^i_u \, \mathrm{d}W^j_s.$$

Moreover, we define $h_n := t_{n+1} - t_n$.

Since the discretization $\{t_0, t_1, \ldots, t_N\}$ is of form (V.6), there exists a unique point in time t_q in this discretization with $t_q = t_n - \tau_l$ whenever $t_n - \tau_l \geq t_0$. Moreover, it holds

$$h_n = t_{n+1} - t_n = t_{n+1} - \tau_l - (t_n - \tau_l) = t_{q+1} - t_q = h_q. \tag{V.7}$$

The algorithms, that approximate the iterated stochastic integrals and are presented below, are based on the Fourier series expansion of the Brownian bridge process, cf. [60, 78, 79, 105, 137]. Consider the Brownian bridge process

$$\left(\int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W^j_u - \frac{s - t_n}{h_n} \int_{t_n-\tau_l}^{t_{n+1}-\tau_l} \mathrm{d}W^j_u \right)_{s \in [t_n, t_{n+1}]}$$

for $j \in \{1, \ldots, m\}$, $l \in \{0, 1, \ldots, D\}$, and $n \in \{0, 1, \ldots, N-1\}$ whenever $t_n - \tau_l \geq t_0$. In the following, let $n \in \{0, 1, \ldots, N-1\}$ be arbitrarily fixed if not stated otherwise. Since P-almost all realizations of the Wiener process are continuous, the Brownian bridge process P-almost surely belongs to $L^2([t_n, t_{n+1}]; \mathbb{R})$. Hence, its Fourier series expansion with respect to the trigonometric, complete orthonormal basis

$$\left(\sqrt{\frac{1}{h_n}}\right)_{s\in[t_n,t_{n+1}]} \cup \left\{\left(\sqrt{\frac{2}{h_n}}\cos\left(\frac{2\pi}{h_n}k(s-t_n)\right)\right)_{s\in[t_n,t_{n+1}]}, k\in\mathbb{N}\right\}$$
$$\cup\left\{\left(\sqrt{\frac{2}{h_n}}\sin\left(\frac{2\pi}{h_n}k(s-t_n)\right)\right)_{s\in[t_n,t_{n+1}]}, k\in\mathbb{N}\right\}$$

of $L^2([t_n, t_{n+1}]; \mathbb{R})$ is given by

$$\begin{aligned}&\int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^j - \frac{s-t_n}{h_n}\Delta W_{n,\tau_l}^j \\ &\quad = \frac{a_{0,n,\tau_l}^j}{2} + \sum_{k=1}^{\infty} a_{k,n,\tau_l}^j \cos\left(\frac{2\pi}{h_n}k(s-t_n)\right) + b_{k,n,\tau_l}^j \sin\left(\frac{2\pi}{h_n}k(s-t_n)\right),\end{aligned} \tag{V.8}$$

where the series is P-almost surely convergent in $L^2([t_n, t_{n+1}]; \mathbb{R})$ at first. The Fourier coefficients are defined as

$$a_{k,n,\tau_l}^j := \frac{2}{h_n}\int_{t_n}^{t_{n+1}} \left(\int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^j - \frac{s-t_n}{h_n}\Delta W_{n,\tau_l}^j\right)\cos\left(\frac{2\pi}{h_n}k(s-t_n)\right)\mathrm{d}s$$

for $k \in \mathbb{N}_0$ and

$$b_{k,n,\tau_l}^j := \frac{2}{h_n}\int_{t_n}^{t_{n+1}} \left(\int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^j - \frac{s-t_n}{h_n}\Delta W_{n,\tau_l}^j\right)\sin\left(\frac{2\pi}{h_n}k(s-t_n)\right)\mathrm{d}s$$

for $k \in \mathbb{N}$. In case of $l = 0$, we just write $a_{k,n}^j := a_{k,n,\tau_0}^j$ and $b_{k,n}^j := b_{k,n,\tau_0}^j$.

The random Fourier coefficients a_{k,n,τ_l}^j and b_{k,n,τ_l}^j, $k \in \mathbb{N}$ and $j \in \{1, \dots, m\}$, are independent and $\mathrm{N}(0, \frac{h_n}{2\pi^2 k^2})$-distributed, cf. [105]. Since covariance

$$\mathrm{E}\left[\Delta W_{n,\tau_l}^i\left(\int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^j - \frac{s-t_n}{h_n}\Delta W_{n,\tau_l}^j\right)\right] = 0$$

for all $s \in [t_n, t_{n+1}]$, the increment $\Delta W_{n,\tau_l}^i$ is further on independent of a_{k,n,τ_l}^j and b_{k,n,τ_l}^j for all $k \in \mathbb{N}$, $i, j \in \{1, \dots, m\}$.

According to [66] and [135], the series in equation (V.8) also converges, uniformly for all $s \in [t_n, t_{n+1}]$, P-almost surely and in $L^p(\Omega; \mathbb{R})$ for all $p \in [1, \infty[$. Thus, the evaluation of that series at a point in time $s \in [t_n, t_{n+1}]$ is well-defined. Setting $s = t_n$ or $s = t_{n+1}$ in equation (V.8), we obtain the relation

$$\frac{a_{0,n,\tau_l}^j}{2} = -\sum_{k=1}^{\infty} a_{k,n,\tau_l}^j \tag{V.9}$$

P-almost surely, where the series converges in $L^p(\Omega; \mathbb{R})$ for all $p \in [1, \infty[$.

Let us give a remark on the convergence of random series.

Remark V.1
Let E be a separable Banach space. Series that converge in $L^p(\Omega; E)$ for some $p \in [1, \infty[$ are convergent in probability, too. If the summands are further independent, the series also converge P*-almost surely by Lévy's theorem, see [66, Theorem 3.1] and, e. g., [12, Satz 14.2] in case of $E = \mathbb{R}$.*

According to Remark V.1, series may converge P-almost surely as well, but this will not always be mentioned below. If not stated otherwise, let $p \in [2, \infty[$ be arbitrarily fixed in this chapter.

In the following, we present the series expansions of iterated stochastic integrals, cf. [60, Section 4] and [137, Subsection 3.7.2]. For this, we consider some stochastic integrals first. Using Itô's formula, cf. [84, Theorem 8.1.1], it P-almost surely holds

$$\int_{t_n}^{t_{n+1}} \frac{s - t_n}{h_n} \, \mathrm{d}W_s^j = \frac{1}{2}\Delta W_n^j - \frac{a_{0,n}^j}{2}, \tag{V.10}$$

$$\int_{t_n}^{t_{n+1}} \cos\Big(\frac{2\pi}{h_n} k(s - t_n)\Big) \, \mathrm{d}W_s^j = \pi k b_{k,n}^j, \tag{V.11}$$

and

$$\int_{t_n}^{t_{n+1}} \sin\Big(\frac{2\pi}{h_n} k(s - t_n)\Big) \, \mathrm{d}W_s^j = -\pi k a_{k,n}^j, \tag{V.12}$$

cf. [105, Lemma 7.4]. Further, for $i \in \{1, \ldots, m\}$, the increment $\Delta W_{n,\tau_l}^i$ is $\mathscr{F}_{t_{n+1}-\tau_l}/\mathscr{B}(\mathbb{R})$-measurable and independent of the Wiener processes W^j for $j \in \{1, \ldots, m\} \setminus i$. In case of $l \in \{1, \ldots, D\}$, it holds $t_{n+1} - \tau_l \leq t_n$ since the discretization is assumed to be of form (V.6).

Due to these measurability and independence properties, we can substitute the inner integral of $\int_{t_n}^{t_{n+1}} \int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W_u^i \, \mathrm{d}W_s^j$, if $i \neq j$ in case of $l = 0$, by expansion

$$\begin{aligned}\int_{t_n-\tau_l}^{s-\tau_l} &\mathrm{d}W_u^i \\ &= \frac{s - t_n}{h_n}\Delta W_{n,\tau_l}^i + \frac{a_{0,n,\tau_l}^i}{2} + \sum_{k=1}^{\infty} a_{k,n,\tau_l}^i \cos\Big(\frac{2\pi}{h_n} k(s - t_n)\Big) + b_{k,n,\tau_l}^i \sin\Big(\frac{2\pi}{h_n} k(s - t_n)\Big)\end{aligned}$$

that converges uniformly for all $s \in [t_0, T]$ P-almost surely, see equation (V.8). In the excluded case, however, we directly have $I_{(j,j),n} = \frac{1}{2}((\Delta W_n^j)^2 - h_n)$ P-almost surely, see equation (V.2). Using formulas (V.10), (V.11), and (V.12), we obtain

$$I_{(i,j),n} = \frac{1}{2}\Delta W_n^i \Delta W_n^j + \frac{a_{0,n}^i}{2}\Delta W_n^j - \frac{a_{0,n}^j}{2}\Delta W_n^i + \pi \sum_{k=1}^{\infty} k(a_{k,n}^i b_{k,n}^j - b_{k,n}^i a_{k,n}^j) \tag{V.13}$$

for $i, j \in \{1, \ldots, m\}$ with $i \neq j$, where the series converges in $L^p(\Omega; \mathbb{R})$, cf. [78]. Similarly, we have

$$I_{(i,j),n,\tau_l} = \frac{1}{2}\Delta W_{n,\tau_l}^i \Delta W_n^j + \frac{a_{0,n,\tau_l}^i}{2}\Delta W_n^j - \frac{a_{0,n}^j}{2}\Delta W_{n,\tau_l}^i + \pi \sum_{k=1}^{\infty} k(a_{k,n,\tau_l}^i b_{k,n}^j - b_{k,n,\tau_l}^i a_{k,n}^j) \tag{V.14}$$

for all $l \in \{1,\dots,D\}$ and $i,j \in \{1,\dots,m\}$, where the series converges in $L^p(\Omega;\mathbb{R})$, cf. [60, Lemma 4.1] and [137, Lemma 7.2]. We refer to the proof of Theorem V.2 below for further details on the convergence.

Thus, in case of $l = 0$, we P-almost surely have

$$I_{(i,j),n} = \frac{\Delta W_n^i \Delta W_n^j - h_n \mathbb{1}_{\{i=j\}}}{2} + A_{(i,j),n}$$

for $i,j \in \{1,\dots,m\}$ where

$$\begin{aligned} A_{(i,j),n} &:= A_{(i,j),n,\tau_0} := \frac{I_{(i,j),n} - I_{(j,i),n}}{2} \\ &= \frac{a_{0,n}^i}{2}\Delta W_n^j - \frac{a_{0,n}^j}{2}\Delta W_n^i + \pi \sum_{k=1}^{\infty} k(a_{k,n}^i b_{k,n}^j - b_{k,n}^i a_{k,n}^j) \end{aligned} \tag{V.15}$$

is the Lévy stochastic area. Here, it holds

$$A_{(i,j),n} = -A_{(j,i),n} \tag{V.16}$$

for $i,j \in \{1,\dots,m\}$ with $i \neq j$ and $A_{(j,j),n} = 0$ for $j \in \{1,\dots,m\}$, cf. [136]. Due to this relation, we only need to simulate the Lévy areas $A_{(i,j),n}$ for $i,j \in \{1,\dots,m\}$ with $i < j$.

Similarly, we P-almost surely have

$$I_{(i,j),n,\tau_l} = \frac{1}{2}\Delta W_{n,\tau_l}^i \Delta W_n^j + A_{(i,j),n,\tau_l}$$

for all $i,j \in \{1,\dots,m\}$ and $l \in \{1,\dots,D\}$ where

$$A_{(i,j),n,\tau_l} := \frac{a_{0,n,\tau_l}^i}{2}\Delta W_n^j - \frac{a_{0,n}^j}{2}\Delta W_{n,\tau_l}^i + \pi \sum_{k=1}^{\infty} k(a_{k,n,\tau_l}^i b_{k,n}^j - b_{k,n,\tau_l}^i a_{k,n}^j), \tag{V.17}$$

whenever $t_n - \tau_l \geq t_0$. Here, random variable $A_{(i,j),n,\tau_l}$ can be seen as a delayed Lévy stochastic area. If $t_n - \tau_l < t_0$, we set $A_{(i,j),n,\tau_l} := 0$. Note that expansion (V.17) above does not commute in contrast to the expansion in equation (V.15). Therefore, the delayed-iterated stochastic integrals $I_{(i,j),n,\tau_l}$ have to be simulated for each pair (i,j), $i,j \in \{1,\dots,m\}$, even if $m = 1$.

Approximations of expansions (V.15) and (V.17) are considered and analyzed in the following sections. There, we need the notation of Gamma function Γ, defined by

$$\Gamma(z) := \int_0^{\infty} x^{z-1} \mathrm{e}^{-x} \, \mathrm{d}x$$

for $z \in \mathbb{R}$ with $z > 0$, cf. [5, pp. 35–36]. The Gamma function Γ generalizes the factorial in the sense that $\Gamma(n+1) = n!$ for $n \in \mathbb{N}_0$, see [5, Theorem 1.9.4] for example.

V.1. Algorithm I: The General Case

According to the introduction of this chapter, the random variables $A_{(i,j),n,\tau_l}$ have to be simulated in order to model the iterated stochastic integrals $I_{(i,j),n,\tau_l}$ for $i,j \in \{1,\dots,m\}$ and $l \in \{0,1,\dots,D\}$, where $n \in \{0,1,\dots,N-1\}$. Only with the exception of $A_{(j,j),n} = 0$ for $j \in \{1,\dots,m\}$ in case of $l=0$, there are so far no methods available that generate these Lévy areas exactly.

In this section, we present a simple method for the approximation of iterated stochastic integral $I_{(i,j),n,\tau_l}$, which was first introduced by Milstein in case of $l=0$, cf. [105, pp. 94–100], and was then extended to iterated stochastic integrals with delay by Yan in [137, Subsection 3.7.2]. Both show that this so-called Fourier method is convergent in $L^2(\Omega;\mathbb{R})$, also see [60, 78, 79]. In this chapter, we prove the convergence in a stronger sense, namely in $L^p(\Omega;\mathbb{R})$ for arbitrary $p \in [2,\infty[$.

This simple approximation of Lévy areas is obtained by truncating series $A_{(i,j),n}$ and $A_{(i,j),n,\tau_l}$, see equations (V.15) and (V.17), after K terms, that is

$$A^K_{(i,j),n} := A^K_{(i,j),n,\tau_0} := \frac{a^i_{0,n}}{2}\Delta W^j_n - \frac{a^j_{0,n}}{2}\Delta W^i_n + \pi\sum_{k=1}^{K} k\big(a^i_{k,n}b^j_{k,n} - b^i_{k,n}a^j_{k,n}\big) \tag{V.18}$$

and

$$A^K_{(i,j),n,\tau_l} := \begin{cases} \dfrac{a^i_{0,n,\tau_l}}{2}\Delta W^j_n - \dfrac{a^j_{0,n}}{2}\Delta W^i_{n,\tau_l} + \pi\displaystyle\sum_{k=1}^{K} k\big(a^i_{k,n,\tau_l}b^j_{k,n} - b^i_{k,n,\tau_l}a^j_{k,n}\big) & \text{if } t_n-\tau_l \geq t_0, \\ 0 & \text{if } t_n-\tau_l < t_0 \end{cases} \tag{V.19}$$

for some $K \in \mathbb{N}$. Note that $A^K_{(j,j),n} = 0$ for all $j \in \{1,\dots,m\}$. Then, the iterated stochastic integrals $I_{(i,j),n,\tau_l}$ for $i,j \in \{1,\dots,m\}$, $l \in \{0,1,\dots,D\}$, and $n \in \{0,1,\dots,N-1\}$ are approximated by

$$I^K_{(i,j),n} := I^K_{(i,j),n,\tau_0} := \frac{1}{2}\big(\Delta W^i_n \Delta W^j_n - h_n \mathbb{1}_{\{i=j\}}\big) + A^K_{(i,j),n} \tag{V.20}$$

and

$$I^K_{(i,j),n,\tau_l} := \begin{cases} \dfrac{1}{2}\Delta W^i_{n,\tau_l}\Delta W^j_n + A^K_{(i,j),n,\tau_l} & \text{if } t_n-\tau_l \geq t_0, \\ 0 & \text{if } t_n-\tau_l < t_0 \end{cases} \tag{V.21}$$

where $K \in \mathbb{N}$. These approximations converge with order $\mathcal{O}(K^{-\frac{1}{2}})$ in $L^p(\Omega;\mathbb{R})$ to $I_{(i,j),n}$ and $I_{(i,j),n,\tau_l}$, respectively, as $K \to \infty$. The precise error estimates are stated in the next theorem. This theorem extends the results from [60, 78, 79, 105, 137], where the convergence in $L^2(\Omega;\mathbb{R})$ is considered.

Theorem V.2
Let $p \in [2,\infty[$ and $n \in \{0,1,\dots,N-1\}$. Consider approximation $I^K_{(i,j),n,\tau_l}$ defined by equations (V.20) *and* (V.21)*, where $K \in \mathbb{N}$, $i,j \in \{1,\dots,m\}$, and $l \in \{0,1,\dots,D\}$. It holds*

$$\max_{\substack{i,j\in\{1,\dots,m\}\\ l\in\{0,1,\dots,D\}}} \|I_{(i,j),n,\tau_l} - I^K_{(i,j),n,\tau_l}\|_{L^p(\Omega;\mathbb{R})} \leq \frac{(p-1)\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}} h_n}{\pi^{\frac{2p+1}{2p}}\sqrt{K}},$$

where in particular $\|I_{(j,j),n} - I^K_{(j,j),n}\|_{L^p(\Omega;\mathbb{R})} = 0$.

Proof. The proof is stated in Section V.5, see p. 160. □

Especially for $p = 2$, the inequality simplifies to

$$\max_{\substack{i,j\in\{1,\dots,m\}\\ l\in\{0,1,\dots,D\}}} \|I_{(i,j),n,\tau_l} - I^K_{(i,j),n,\tau_l}\|_{L^2(\Omega;\mathbb{R})} \le \frac{h_n}{\sqrt{2}\pi\sqrt{K}} \tag{V.22}$$

since $\Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$. Moreover, if $p = 2$, the error can be precisely stated, cf. [105, Lemma 7.3] for the case of $l = 0$.

Corollary V.3
Let $n \in \{0, 1, \dots, N-1\}$, *and consider approximation* $I^K_{(i,j),n,\tau_l}$ *defined by equations* (V.20) *and* (V.21), *where* $K \in \mathbb{N}$, $i, j \in \{1, \dots, m\}$, *and* $l \in \{1, \dots, D\}$. *It holds* $\|I_{(j,j),n} - I^K_{(j,j),n}\|_{L^2(\Omega;\mathbb{R})} = 0$ *for all* $j \in \{1, \dots, m\}$ *and*

$$\|I_{(i,j),n,\tau_l} - I^K_{(i,j),n,\tau_l}\|_{L^2(\Omega;\mathbb{R})} = \frac{h}{\sqrt{2}\pi}\left(\frac{\pi^2}{6} - \sum_{k=1}^{K}\frac{1}{k^2}\right)^{\frac{1}{2}}$$

for all $i, j \in \{1, \dots, m\}$ *and* $l \in \{0, 1, \dots, D\}$, *where* $i \neq j$ *if* $l = 0$.

Proof. The proof is stated in Section V.5, see p. 162. □

In the following, we provide an algorithm for the simulation of ΔW^j_n, $I^K_{(i,j),n}$, and $I^K_{(i,j),n,\tau_l}$, if $t_n - \tau_l \ge t_0$, for all $i, j \in \{1, \dots, m\}$, $l \in \{1, \dots, D\}$, and $n \in \{0, 1, \dots, N-1\}$.

Define the matrix $A^K_{n,\tau_l} := (A^K_{(i,j),n,\tau_l})_{1\le i,j\le m} \in \mathbb{R}^{m\times m}$, where $A^K_n := A^K_{n,\tau_0}$, and define the vector operator

$$\operatorname{vec}\big[(A^K_{n,\tau_l})^{\mathrm{T}}\big] := \big(A^K_{(1,1),n,\tau_l}, \dots, A^K_{(1,m),n,\tau_l}, \dots, A^K_{(m,1),n,\tau_l}, \dots, A^K_{(m,m),n,\tau_l}\big)^{\mathrm{T}} \in \mathbb{R}^{m^2}. \tag{V.23}$$

Using the Kronecker product $\otimes$, we P-almost surely have

$$\operatorname{vec}\big[(A^K_n)^{\mathrm{T}}\big] = \frac{a_{0,n}}{2} \otimes \Delta W_n - \Delta W_n \otimes \frac{a_{0,n}}{2} + \pi \sum_{k=1}^{K} k(a_{k,n} \otimes b_{k,n} - b_{k,n} \otimes a_{k,n}) \tag{V.24}$$

and

$$\operatorname{vec}\big[(A^K_{n,\tau_l})^{\mathrm{T}}\big] = \frac{a_{0,n,\tau_l}}{2} \otimes \Delta W_n - \Delta W_{n,\tau_l} \otimes \frac{a_{0,n}}{2} + \pi \sum_{k=1}^{K} k(a_{k,n,\tau_l} \otimes b_{k,n} - b_{k,n,\tau_l} \otimes a_{k,n}),$$

where $\Delta W_{n,\tau_l} := (\Delta W^1_{n,\tau_l}, \dots, \Delta W^m_{n,\tau_l})^{\mathrm{T}}$ as well as $a_{k,n,\tau_l} := (a^1_{k,n,\tau_l}, \dots, a^m_{k,n,\tau_l})^{\mathrm{T}}$ for $k \in \mathbb{N}_0$ and $b_{k,n,\tau_l} := (b^1_{k,n,\tau_l}, \dots, b^m_{k,n,\tau_l})^{\mathrm{T}}$ for $k \in \mathbb{N}$.

So far, we only considered the approximations $I^K_{(i,j),n}$ and $A^K_{(i,j),n}$ as well as $I^K_{(i,j),n,\tau_l}$ and $A^K_{(i,j),n,\tau_l}$ for fixed $i, j \in \{1, \dots, m\}$, $n \in \{0, 1, \dots, N-1\}$, $l \in \{1, \dots, D\}$, and $K \in \mathbb{N}$. In order to simulate these approximations correctly, we have to take their dependencies into account by generating the random variables. Therefore, we make the following considerations.

For some point in time $t_q = t_n - \tau_l \geq t_0$ of the discretization under consideration, we have the identities $a^j_{k,n,\tau_l} = a^j_{k,q}$ for $k \in \mathbb{N}_0$ and $b^j_{k,n,\tau_l} = b^j_{k,q}$ for $k \in \mathbb{N}$. In fact, considering Fourier coefficient a^j_{k,n,τ_l} exemplarily and using the substitution $s = r + \tau_l$, it holds

$$\begin{aligned} a^j_{k,n,\tau_l} &= \frac{2}{h_n} \int_{t_n}^{t_{n+1}} \left(\int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W^j_u - \frac{s-t_n}{h_n} \int_{t_n-\tau_l}^{t_{n+1}-\tau_l} \mathrm{d}W^j_u \right) \cos\left(\frac{2\pi}{h_n} k(s-t_n) \right) \mathrm{d}s \\ &= \frac{2}{h_n} \int_{t_n-\tau_l}^{t_{n+1}-\tau_l} \left(\int_{t_n-\tau_l}^{r} \mathrm{d}W^j_u - \frac{r+\tau_l-t_n}{h_n} \int_{t_n-\tau_l}^{t_{n+1}-\tau_l} \mathrm{d}W^j_u \right) \cos\left(\frac{2\pi}{h_n} k(r+\tau_l-t_n) \right) \mathrm{d}r \\ &= \frac{2}{h_q} \int_{t_q}^{t_{q+1}} \left(\int_{t_q}^{r} \mathrm{d}W^j_u - \frac{r-t_q}{h_q} \int_{t_q}^{t_{q+1}} \mathrm{d}W^j_u \right) \cos\left(\frac{2\pi}{h_q} k(r-t_q) \right) \mathrm{d}r \\ &= a^j_{k,q}, \end{aligned}$$

where $t_{q+1} = t_{n+1} - \tau_l$, equation (V.7), and $\Delta W^j_q = \Delta W^j_{n,\tau_l}$ are used.

If we want to ensure, cf. Theorem V.2, that

$$\|I_{(i,j),n,\tau_l} - I^{K_n}_{(i,j),n,\tau_l}\|_{L^p(\Omega;\mathbb{R})} \leq \frac{(p-1)\left(\Gamma(\frac{p+1}{2})\right)^{\frac{1}{p}} h_n}{\pi^{\frac{2p+1}{2p}} \sqrt{K_n}} \leq \varepsilon$$

and

$$\|I_{(i,j),q} - I^{K_q}_{(i,j),q}\|_{L^p(\Omega;\mathbb{R})} \leq \frac{(p-1)\left(\Gamma(\frac{p+1}{2})\right)^{\frac{1}{p}} h_q}{\pi^{\frac{2p+1}{2p}} \sqrt{K_q}} \leq \varepsilon$$

for some error $\varepsilon > 0$, we can choose $K = K_n = K_q$ since $h_n = h_q$, see equation (V.7). Then, the sums in approximations (V.18) and (V.19) have the same number of random variables, where $\Delta W_{n,\tau_l} = \Delta W_q$, $a_{0,n,\tau_l} = a_{0,q}$, $a_{k,n,\tau_l} = a_{k,q}$, and $b_{k,n,\tau_l} = b_{k,q}$ for $k \in \{1, \ldots, K\}$. These and only these random variables were already generated in the step where iterated stochastic integrals $I^K_{(i,j),q}$ for $i, j \in \{1, \ldots, m\}$ have been simulated. Thus, in order to simulate $I^K_{(i,j),n,\tau_l}$, we only have to generate the random variables ΔW_n, $a_{0,n}$, $a_{k,n}$, and $b_{k,n}$ for $k \in \{1, \ldots, K\}$ in addition.

In simulating random variable $a_{0,n}$, we further have to take into account that $a_{0,n}$ is not independent of the random variables $a_{k,n}$ for $k \in \{1, \ldots, K\}$. Since, for $k \in \mathbb{N}$, Fourier coefficients $a^j_{k,n}$, $j \in \{1, \ldots, m\}$, are independent and $\mathrm{N}\big(0, \frac{h_n}{2\pi^2 k^2}\big)$-distributed, we obtain for $j \in \{1, \ldots, m\}$ by identity (V.9) that

$$-\frac{a^j_{0,n}}{2} = \sum_{k=1}^{\infty} a^j_{k,n} = \sum_{k=1}^{K} a^j_{k,n} + \sum_{k=K+1}^{\infty} a^j_{k,n}$$

P-almost surely, where random variable

$$\sum_{k=K+1}^{\infty} a^j_{k,n} \sim \mathrm{N}\left(0, \sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^2}\right) \tag{V.25}$$

is independent of $a_{k,n}^{j}$ for all $k \in \{1, \ldots, K\}$.

Let us introduce some additional notations. We denote by $0_{i\times j}$ the matrix of zeros of size $i \times j$ and by I_m the identity matrix of size $m \times m$. For emphasis, we also write $0_{m\times 1}$ for $0 \in \mathbb{R}^m$.

According to [74, Corollary 6.11], for $n \in \{0, 1, \ldots, N-1\}$ and $k \in \{1, \ldots, K\}$, there exist $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variables B_n, $G_{0,n}$, $U_{k,n}$, and $V_{k,n}$ such that

$$\Delta W_n = \sqrt{h_n} B_n,$$

$$\sum_{k=K+1}^{\infty} a_{k,n} = \frac{\sqrt{h_n}}{\sqrt{2}\pi} \left(\frac{\pi^2}{6} - \sum_{k=1}^{K} \frac{1}{k^2} \right)^{\frac{1}{2}} G_{0,n},$$

$$a_{k,n} = \frac{\sqrt{h_n}}{\sqrt{2}\pi k} U_{k,n},$$

and

$$b_{k,n} = \frac{\sqrt{h_n}}{\sqrt{2}\pi k} V_{k,n}$$

P-almost surely. Then, we can rewrite equation (V.24) to

$$\begin{aligned} \operatorname{vec}\big[(A_n^K)^{\mathrm{T}}\big] = {} & \frac{h_n}{\sqrt{2}\pi} \sqrt{\frac{\pi^2}{6} - \sum_{k=1}^{K} \frac{1}{k^2}} \big(B_n \otimes G_{0,n} - G_{0,n} \otimes B_n \big) \\ & + \frac{h_n}{2\pi} \sum_{k=1}^{K} \frac{1}{k} \big(U_{k,n} \otimes (V_{k,n} - \sqrt{2} B_n) - (V_{k,n} - \sqrt{2} B_n) \otimes U_{k,n} \big) \end{aligned}$$

P-almost surely and approximate $I_n = (I_{(i,j),n})_{1 \le i,j \le m}$ by

$$\operatorname{vec}\big[(I_n^K)^{\mathrm{T}}\big] = \frac{1}{2} \big(\Delta W_n \otimes \Delta W_n - \operatorname{vec}[h_n I_m] \big) + \operatorname{vec}\big[(A_n^K)^{\mathrm{T}}\big]$$

P-almost surely. Similarly, we P-almost surely have

$$\operatorname{vec}\big[(I_{n,\tau_l}^K)^{\mathrm{T}}\big] = \frac{1}{2} \big(\Delta W_q \otimes \Delta W_n \big) + \operatorname{vec}\big[(A_{n,\tau_l}^K)^{\mathrm{T}}\big],$$

where $t_q = t_n - \tau_l \ge t_0$ and

$$\begin{aligned} \operatorname{vec}\big[(A_{n,\tau_l}^K)^{\mathrm{T}}\big] = {} & \frac{h_n}{\sqrt{2}\pi} \sqrt{\frac{\pi^2}{6} - \sum_{k=1}^{K} \frac{1}{k^2}} \big(B_q \otimes G_{0,n} - G_{0,q} \otimes B_n \big) \\ & + \frac{h_n}{2\pi} \sum_{k=1}^{K} \frac{1}{k} \big(U_{k,q} \otimes (V_{k,n} - \sqrt{2} B_n) - (V_{k,q} - \sqrt{2} B_q) \otimes U_{k,n} \big) \end{aligned}$$

P-almost surely. Using this, we provide the following algorithm for the approximation of iterated stochastic integrals.

Algorithm V.4

Let discretization $\{t_0, t_1, \dots, t_N\}$ *of* $[t_0, T]$ *be of form* (V.6), *and let* $p \in [2, \infty[$. *In order to simulate* ΔW_n^j *and* $I_{(i,j),n,\tau_l}$ *for* $i, j \in \{1, \dots, m\}$, $l \in \{0, 1, \dots, D\}$, *and* $n \in \{0, 1, \dots, N-1\}$ *such that*

$$\max_{\substack{i,j \in \{1,\dots,m\} \\ l \in \{0,1,\dots,D\}}} \|I_{(i,j),n,\tau_l} - I^{K_n}_{(i,j),n,\tau_l}\|_{L^p(\Omega;\mathbb{R})} \leq \varepsilon$$

for some error bound $\epsilon > 0$, *proceed as follows. For* $n = 0, 1, \dots, N-1$,

i) set

$$K_n = \left\lceil \frac{(p-1)^2 \big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}} h_n^2}{\pi^{\frac{2p+1}{p}} \varepsilon^2} \right\rceil,$$

where $\lceil \cdot \rceil$ *is the ceiling function. In case of* $p = 2$, *this especially means that* $K_n = \lceil \frac{h_n^2}{2\pi^2 \varepsilon^2} \rceil$.

ii) Generate and store independently $\mathrm{N}(0_{m\times 1}, I_m)$*-distributed random variables* B_n, $G_{0,n}$, $U_{k,n}$, *and* $V_{k,n}$ *for* $k \in \{1, \dots, K_n\}$.

iii) Set $\Delta W_n = \sqrt{h_n} B_n$, *and approximate random variable* $\mathrm{vec}\big[(A_n)^{\mathrm{T}}\big]$ *by*

$$\begin{aligned}\mathrm{vec}\big[(A_n^{K_n})^{\mathrm{T}}\big] = \frac{h_n}{\sqrt{2}\pi} \sqrt{\frac{\pi^2}{6} - \sum_{k=1}^{K_n} \frac{1}{k^2}} \big(B_n \otimes G_{0,n} - G_{0,n} \otimes B_n\big) \\ + \frac{h_n}{2\pi} \sum_{k=1}^{K_n} \frac{1}{k} \big(U_{k,n} \otimes (V_{k,n} - \sqrt{2} B_n) - (V_{k,n} - \sqrt{2} B_n) \otimes U_{k,n}\big).\end{aligned}$$

iv) Then, the approximation of $\mathrm{vec}\big[(I_n)^{\mathrm{T}}\big]$ *is computed as*

$$\mathrm{vec}\big[(I_n^{K_n})^{\mathrm{T}}\big] = \frac{h_n}{2} \big(B_n \otimes B_n - \mathrm{vec}[I_m]\big) + \mathrm{vec}\big[(A_n^{K_n})^{\mathrm{T}}\big].$$

v) For $l = 1, \dots, D$, *if* $t_n - \tau_l \geq t_0$, *determine* $q \in \{0, 1, \dots, n-1\}$ *such that* $t_q = t_n - \tau_l$, *and the approximation of* $\mathrm{vec}\big[(I_{n,\tau_l})^{\mathrm{T}}\big]$ *is computed as*

$$\mathrm{vec}\big[(I_{n,\tau_l}^{K_n})^{\mathrm{T}}\big] = \frac{h_n}{2} \big(B_q \otimes B_n\big) + \mathrm{vec}\big[(A_{n,\tau_l}^{K_n})^{\mathrm{T}}\big]$$

where

$$\begin{aligned}\mathrm{vec}\big[(A_{n,\tau_l}^{K_n})^{\mathrm{T}}\big] = \frac{h_n}{\sqrt{2}\pi} \sqrt{\frac{\pi^2}{6} - \sum_{k=1}^{K_n} \frac{1}{k^2}} \big(B_q \otimes G_{0,n} - G_{0,q} \otimes B_n\big) \\ + \frac{h_n}{2\pi} \sum_{k=1}^{K_n} \frac{1}{k} \big(U_{k,q} \otimes (V_{k,n} - \sqrt{2} B_n) - (V_{k,q} - \sqrt{2} B_q) \otimes U_{k,n}\big).\end{aligned}$$

V.2. Algorithm II: Nondelayed-Iterated Stochastic Integrals

A new algorithm for the approximation nondelayed-iterated stochastic integrals $I_{(j,j),n}$ is developed in this section. As announced in the introduction of this chapter, this algorithm lowers the computational cost significantly compared to the one of Wiktorsson in [136], see in this regard Section V.3 in particular. While Wiktorsson only considered the convergence in $L^2(\Omega;\mathbb{R})$, see [136, p. 481], we show the convergence of our method in $L^p(\Omega;\mathbb{R})$ for arbitrary $p \in [2,\infty[$.

Similarly to the algorithm of Wiktorsson in [136], we truncate series expansion (V.15) and approximate its remainder such that the order of convergence will be improved. The algorithm provided in the previous section serves as a basis at this point. Wiktorsson neglects in his algorithm known information about the distribution of the coefficient $a_{0,n}$, cf. formula (V.25). Incorporating this information is the main idea of our new algorithm and results in savings in computational costs.

In the following, let $n \in \{0,1,\ldots,N-1\}$ and $p \in [2,\infty[$ be arbitrary fixed first. The remainder, neglected in approximation (V.18) of series expansion (V.15), is given by

$$A_{(i,j),n} - A^K_{(i,j),n} = \pi \sum_{k=K+1}^{\infty} k(a^i_{k,n} b^j_{k,n} - b^i_{k,n} a^j_{k,n}) \tag{V.26}$$

P-almost surely for $i,j \in \{1,\ldots,m\}$, where the series converges in $L^p(\Omega;\mathbb{R})$, cf. Theorem V.2. As mentioned before, it holds $A_{(j,j),n} - A^K_{(j,j),n} = 0$, and thus, let $i \neq j$ unless otherwise stated. According to relation (V.16), we only have to approximate the Lévy areas $A_{(i,j),n}$ for $i < j$, cf. [136]. Therefore and for technical reasons, we introduce the matrix

$$H_m := \begin{pmatrix} 0_{m-1\times 1} & I_{m-1} & 0_{m-1\times m(m-1)} \\ 0_{m-2\times m+2} & I_{m-2} & 0_{m-2\times m(m-2)} \\ \vdots & \vdots & \vdots \\ 0_{m-j\times (j-1)m+j} & I_{m-j} & 0_{m-j\times m(m-j)} \\ \vdots & \vdots & \vdots \\ 0_{1\times (m-2)m+m-1} & 1 & 0_{1\times m} \end{pmatrix} \in \mathbb{R}^{M\times m^2}, \tag{V.27}$$

where $M := \frac{1}{2}m(m-1)$ and I_j is the identity matrix of size $j\times j$, cf. [136, pp. 477, 478, and 486]. Considering equations (V.23) and (V.24), selection matrix H_m implies

$$\begin{aligned}&\big(A^K_{(1,2),n},\ldots,A^K_{(1,m),n},A^K_{(2,3),n},\ldots,A^K_{(2,m),n},\ldots,A^K_{(j,j+1),n},\ldots,A^K_{(j,m),n},\ldots,A^K_{(m-1,m),n}\big)^{\mathrm{T}} \\ &\quad = H_m \mathrm{vec}\big[(A^K_n)^{\mathrm{T}}\big]\end{aligned} \tag{V.28}$$

P-almost surely. Let $(e_j)_{j\in\{1,\ldots,m\}}$ be the canonical orthonormal basis of $\mathbb{R}^m$. That is, e_j denotes the jth unit vector in $\mathbb{R}^m$. We define permutation matrix $P_m \in \mathbb{R}^{m^2\times m^2}$ by

$$P_m := \sum_{i,j=1}^{m} e_i e_j^{\mathrm{T}} \otimes e_j e_i^{\mathrm{T}}, \tag{V.29}$$

cf. [136, p. 478]. Then, it holds

$$P_m(x\otimes y) = y \otimes x$$

for all $x, y \in \mathbb{R}^m$, and further, we have

$$P_m = \sum_{j=1}^{m} e_j \otimes (I_m \otimes e_j^{\mathrm{T}}) = \sum_{j=1}^{m} e_j^{\mathrm{T}} \otimes (I_m \otimes e_j), \tag{V.30}$$

see [95]. The latter representation is used in [90] for example. Using selection matrix H_m and permutation matrix P_m, we have with relation (V.16) that

$$\mathrm{vec}\big[(A_n^K)^{\mathrm{T}}\big] = (I_{m^2} - P_m) H_m^{\mathrm{T}} \big(H_m \mathrm{vec}\big[(A_n^K)^{\mathrm{T}}\big]\big), \tag{V.31}$$

see [136, p. 479]. In view of equations (V.23) and (V.28), the remainder in equation (V.26) can be represented in vectorial form, and it holds

$$H_m \big(\mathrm{vec}\big[(A_n)^{\mathrm{T}}\big] - \mathrm{vec}\big[(A_n^K)^{\mathrm{T}}\big]\big) = \pi \sum_{k=K+1}^{\infty} k H_m (P_m - I_{m^2})(b_{k,n} \otimes a_{k,n}) \tag{V.32}$$

P-almost surely, where the series converges in $L^p(\Omega; \mathbb{R}^M)$ as the series in equation (V.26) is convergent in $L^p(\Omega; \mathbb{R})$.

In the following, we approximate the remainder in equation (V.32) by a suitable random variable and add that approximation to the truncated series (V.31), cf. [136]. Here, Algorithm V.4 already incorporates the normally distributed random variable $\sum_{k=K+1}^{\infty} a_{k,n}$. Thus, our approximation, obtained by Algorithm V.4, is not independent of the remainder in equation (V.26) anymore. This is the main difference and difficulty compared to [136]. In order to provide a reasonable approximation, we have to take this dependence into account in our derivation. Wiktorsson states that the conditional distribution of the remainder, which he considers in his approximation, is normal, see [136, pp. 479–480]. In our approach, the joint conditional distribution of the remainder in equation (V.32) and random variable $\sum_{k=K+1}^{\infty} a_{k,n}$ is normal, too.

Lemma V.5
Let $K \in \mathbb{N}$. Given $\{b_{k,n} : k \in \mathbb{N}$ with $k > K\}$, the conditional distribution of the random variable

$$\begin{pmatrix} \displaystyle\sum_{k=K+1}^{\infty} a_{k,n} \\ \displaystyle\pi \sum_{k=K+1}^{\infty} k H_m (P_m - I_{m^2})(b_{k,n} \otimes a_{k,n}) \end{pmatrix} : \Omega \to \mathbb{R}^{m+M} \tag{V.33}$$

is normal with conditional expectation $0_{(m+M)\times 1}$ P-almost surely and conditional covariance

$$\begin{pmatrix} \Sigma_{1,n}^K & (\Sigma_{2,n}^K)^{\mathrm{T}} \\ \Sigma_{2,n}^K & \Sigma_{3,n}^K \end{pmatrix} : \Omega \to \mathbb{R}^{(m+M)\times(m+M)},$$

where

$$\Sigma_{1,n}^K = \Big(\sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^2} \Big) I_m, \tag{V.34}$$

$$\Sigma_{2,n}^K = \frac{h_n}{2\pi} H_m(P_m - I_{m^2})\Big(\sum_{k=K+1}^{\infty} \frac{b_{k,n}}{k} \otimes I_m \Big), \tag{V.35}$$

and

$$\Sigma_{3,n}^K = \frac{h_n}{2} H_m(P_m - I_{m^2})\Big(\sum_{k=K+1}^{\infty} b_{k,n} b_{k,n}^{\mathrm{T}} \otimes I_m \Big)(P_m - I_{m^2}) H_m^{\mathrm{T}} \tag{V.36}$$

P-*almost surely. Here, we have* $\Sigma_{1,n}^K \in \mathbb{R}^{m\times m}$, $\Sigma_{2,n}^K \colon \Omega \to \mathbb{R}^{M\times m}$, *and* $\Sigma_{3,n}^K \colon \Omega \to \mathbb{R}^{M\times M}$. *The series* $\sum_{k=K+1}^{\infty} \frac{1}{k} b_{k,n}^j$ *and* $\sum_{k=K+1}^{\infty} b_{k,n}^i b_{k,n}^j$ *converge absolutely in the* $L^p(\Omega;\mathbb{R})$*-norm for every* $p \in [1,\infty[$ *as well as* P-*almost surely for all* $i, j \in \{1,\dots,m\}$.

Proof. The proof is stated in Section V.5, see p. 162. □

In the following, we consider the conditional covariance matrix in Lemma V.5 more closely. Introducing the Schur complement

$$S_n^K \coloneqq \Sigma_{3,n}^K - \Sigma_{2,n}^K (\Sigma_{1,n}^K)^{-1} (\Sigma_{2,n}^K)^{\mathrm{T}}, \tag{V.37}$$

we P-almost surely have

$$\begin{aligned} S_n^K = {} & \frac{h_n}{2} H_m(P_m - I_{m^2}) \\ & \times \Bigg(\Bigg(\sum_{k=K+1}^{\infty} b_{k,n} b_{k,n}^{\mathrm{T}} - \Bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2} \Bigg)^{-1} \Bigg(\sum_{k=K+1}^{\infty} \frac{b_{k,n}}{k} \Bigg) \Bigg(\sum_{k=K+1}^{\infty} \frac{b_{k,n}^{\mathrm{T}}}{k} \Bigg) \Bigg) \otimes I_m \Bigg) \\ & \times (P_m - I_{m^2}) H_m^{\mathrm{T}}, \end{aligned} \tag{V.38}$$

and

$$\begin{pmatrix} \Sigma_{1,n}^K & (\Sigma_{2,n}^K)^{\mathrm{T}} \\ \Sigma_{2,n}^K & \Sigma_{3,n}^K \end{pmatrix} = \begin{pmatrix} I_m & 0_{m\times M} \\ \Sigma_{2,n}^K (\Sigma_{1,n}^K)^{-1} & I_M \end{pmatrix} \begin{pmatrix} \Sigma_{1,n}^K & (\Sigma_{2,n}^K)^{\mathrm{T}} \\ 0_{M\times m} & S_n^K \end{pmatrix}. \tag{V.39}$$

Further, let the matrix square roots of $\Sigma_{1,n}^K$ and S_n^K be denoted by $\sqrt{\Sigma_{1,n}^K}$ and $\sqrt{S_n^K}$ so that we have $\Sigma_{1,n}^K = \sqrt{\Sigma_{1,n}^K}\sqrt{\Sigma_{1,n}^K}$ and $S_n^K = \sqrt{S_n^K}\sqrt{S_n^K}$. Using equation (V.39) and the previous introduced matrix square roots, we can calculate a covariance decomposition

$$\begin{aligned} \begin{pmatrix} \Sigma_{1,n}^K & (\Sigma_{2,n}^K)^{\mathrm{T}} \\ \Sigma_{2,n}^K & \Sigma_{3,n}^K \end{pmatrix} &= \begin{pmatrix} I_m & 0_{m\times M} \\ \Sigma_{2,n}^K (\Sigma_{1,n}^K)^{-1} & I_M \end{pmatrix} \begin{pmatrix} \Sigma_{1,n}^K & 0_{m\times M} \\ 0_{M\times m} & S_n^K \end{pmatrix} \begin{pmatrix} I_m & (\Sigma_{1,n}^K)^{-1} (\Sigma_{2,n}^K)^{\mathrm{T}} \\ 0_{M\times m} & I_M \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\Sigma_{1,n}^K} & 0_{m\times M} \\ \Sigma_{2,n}^K (\Sigma_{1,n}^K)^{-1} \sqrt{\Sigma_{1,n}^K} & \sqrt{S_n^K} \end{pmatrix} \begin{pmatrix} \sqrt{\Sigma_{1,n}^K} & \sqrt{\Sigma_{1,n}^K} (\Sigma_{1,n}^K)^{-1} (\Sigma_{2,n}^K)^{\mathrm{T}} \\ 0_{M\times m} & \sqrt{S_n^K} \end{pmatrix}. \end{aligned}$$

According to this decomposition of the conditional covariance matrix and to [74, Corollary 6.11], there exists a $\mathrm{N}(0_{(m+M)\times 1}, I_{m+M})$-distributed random variable G_n such that

$$\begin{pmatrix} \displaystyle\sum_{k=K+1}^{\infty} a_{k,n} \\ \displaystyle\pi \sum_{k=K+1}^{\infty} k H_m(P_m - I_{m^2})(b_{k,n} \otimes a_{k,n}) \end{pmatrix} = \begin{pmatrix} \sqrt{\Sigma_{1,n}^K} & 0_{m\times M} \\ \Sigma_{2,n}^K (\Sigma_{1,n}^K)^{-1} \sqrt{\Sigma_{1,n}^K} & \sqrt{S_n^K} \end{pmatrix} G_n$$

P-almost surely. Here, random variable G_n is stochastically independent of Fourier coefficients $b_{k,n}$, $k \in \mathbb{N}$, or is, more precisely, only depending on Fourier coefficients $a_{k,n}$, $k \in \mathbb{N}$ where $k > K$. Further, it P-almost surely holds $\mathrm{E}[G_n|\mathscr{F}_{t_{n+1}}] = G_n$ and $\mathrm{E}[G_n|\mathscr{F}_{t_n}] = 0_{(m+M)\times 1}$.

Writing $G_n = (G_{0,n}^{\mathrm{T}}, G_{1,n}^{\mathrm{T}})^{\mathrm{T}}$ where $G_{0,n} \sim \mathrm{N}(0_{m\times 1}, I_m)$ and $G_{1,n} \sim \mathrm{N}(0_{M\times 1}, I_M)$, the remainder in equation (V.32) can be represented as

$$\pi \sum_{k=K+1}^{\infty} k H_m(P_m - I_{m^2})(b_{k,n} \otimes a_{k,n}) = \varSigma_{2,n}^K (\varSigma_{1,n}^K)^{-1} \sqrt{\varSigma_{1,n}^K} G_{0,n} + \sqrt{S_n^K} G_{1,n} \tag{V.40}$$

P-almost surely, where $\sqrt{\varSigma_{1,n}^K} G_{0,n} = \sum_{k=K+1}^{\infty} a_{k,n}$ P-almost surely.

Similarly to the approach of Wiktorsson in [136], we replace $\sqrt{S_n^K}$ with $\sqrt{\mathrm{E}[S_n^K]}$. The next lemma provides an explicit expression of matrix square root $\sqrt{\mathrm{E}[S_n^K]}$.

Lemma V.6
Let $K \in \mathbb{N}$ and $n \in \{0, 1, \ldots, N-1\}$. Consider Schur complement S_n^K, see equations (V.37) *and* (V.38)*. Matrix square root $\sqrt{\mathrm{E}[S_n^K]}$ of matrix $\mathrm{E}[S_n^K] \in \mathbb{R}^{M\times M}$ satisfies*

$$\sqrt{\mathrm{E}[S_n^K]} = \frac{h_n}{\sqrt{2}\pi} \left(\left(\sum_{k=K+1}^{\infty} \frac{1}{k^2} \right) - \left(\sum_{k=K+1}^{\infty} \frac{1}{k^2} \right)^{-1} \left(\sum_{k=K+1}^{\infty} \frac{1}{k^4} \right) \right)^{\frac{1}{2}} I_M.$$

Proof. The proof is stated in Section V.5, see p. 165. □

We emphasize that square root matrix $\sqrt{\mathrm{E}[S_n^K]}$ is diagonal. Hence, less computational effort is needed to compute $\sqrt{\mathrm{E}[S_n^K]}$ than for the computation of random matrix $\sqrt{\Sigma_\infty}$ in Wiktorsson's article, cf. [136, Equations (4.5) and (4.7)].

By replacing the random square root matrix $\sqrt{S_n^K}$ with the deterministic and diagonal square root matrix $\sqrt{\mathrm{E}[S_n^K]}$ in equation (V.40), the remainder is only simulated approximately. The following theorem provides an estimate of the error that results from this procedure.

Theorem V.7
Let $n \in \{0, 1, \ldots, N-1\}$. It holds

$$\max_{i\in\{1,\ldots,M\}} \left\| e_i^{\mathrm{T}} \left(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]} \right) G_{1,n} \right\|_{L^2(\Omega;\mathbb{R})} \leq \frac{\sqrt{m} h_n}{\sqrt{12}\pi K}$$

for all $K \in \mathbb{N}$, where e_i is the ith unit vector of $\mathbb{R}^M$.

Proof. The proof is stated in Section V.5, see p. 166. □

We improve Algorithm V.4 by adding the approximation of the remainder (V.26), also see equation (V.40). Similarly to approximation

$$\mathrm{vec}\left[(I_n^K)^{\mathrm{T}}\right] = \frac{1}{2}\left(\Delta W_n \otimes \Delta W_n - \mathrm{vec}[h_n I_m]\right) + \mathrm{vec}\left[(A_n^K)^{\mathrm{T}}\right]$$

of the iterated stochastic integrals from the previous section, we define the approximation

$$\operatorname{vec}\big[(I_n^{K+})^{\mathrm{T}}\big] \coloneqq \operatorname{vec}\big[(I_n^{K})^{\mathrm{T}}\big] + (I_{m^2} - P_m)H_m^{\mathrm{T}}\Big(\Sigma_{2,n}^K(\Sigma_{1,n}^K)^{-1}\sqrt{\Sigma_{1,n}^K}G_{0,n} + \sqrt{\mathrm{E}[S_n^K]}G_{1,n}\Big). \tag{V.41}$$

This additional term increases the order of convergence in K as $K \to \infty$ compared to approximation $I_{(i,j),n}^K$, which is of order $\mathcal{O}(K^{-\frac{1}{2}})$ as $K \to \infty$, see inequality (V.22) in Theorem V.2.

Theorem V.8
Let $n \in \{0, 1, \ldots, N-1\}$. Consider approximation $I_{(i,j),n}^{K+}$ defined by equation (V.41)*, where $K \in \mathbb{N}$ and $i, j \in \{1, \ldots, m\}$. It holds*

$$\max_{i,j\in\{1,\ldots,m\}} \|I_{(i,j),n} - I_{(i,j),n}^{K+}\|_{L^2(\Omega;\mathbb{R})} \leq \frac{\sqrt{m}h_n}{\sqrt{12}\pi K},$$

where in particular

$$\max_{j\in\{1,\ldots,m\}} \|I_{(j,j),n} - I_{(j,j),n}^{K+}\|_{L^2(\Omega;\mathbb{R})} = 0.$$

Proof. The result directly follows from Theorem V.7 because

$$\begin{aligned}
&\max_{i,j\in\{1,\ldots,m\}} \|I_{(i,j),n} - I_{(i,j),n}^{K+}\|_{L^2(\Omega;\mathbb{R})} \\
&\quad= \max_{i\in\{1,\ldots,M\}} \big\|e_i^{\mathrm{T}}\big(H_m\operatorname{vec}\big[(I_n)^{\mathrm{T}}\big] - H_m\operatorname{vec}\big[(I_n^{K+})^{\mathrm{T}}\big]\big)\big\|_{L^2(\Omega;\mathbb{R})} \\
&\quad= \max_{i\in\{1,\ldots,M\}} \big\|e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big)G_{1,n}\big\|_{L^2(\Omega;\mathbb{R})}
\end{aligned} \tag{V.42}$$

where $I_{(j,j),n} = I_{(j,j),n}^{K+}$ for $j \in \{1, \ldots, m\}$. ☐

Let us already remark here that the error bound $\frac{\sqrt{m}h_n}{\sqrt{12}\pi K}$ in the theorem above is smaller than the one of Wiktorsson's algorithm, cf. [136, Theorem 4.1]. Wiktorsson proved that his approximation $I_{(i,j),n}^{(K)'}$ satisfies

$$\max_{i,j\in\{1,\ldots,m\}} \|I_{(i,j),n} - I_{(i,j),n}^{(K)'}\|_{L^2(\Omega;\mathbb{R})} \leq \frac{\sqrt{5m^2(m-1)}h_n}{\sqrt{24}\pi K}. \tag{V.43}$$

Thus, our algorithm improves that error bound by a factor of $\frac{\sqrt{5m(m-1)}}{\sqrt{2}}$. This lowers the costs significantly. For more details on this, we refer to Section V.3.

The following theorem generalizes the results from Theorem V.8 to arbitrary $p \in \,]2, \infty[$.

Theorem V.9
Let $p \in \,]2, \infty[$ and $n \in \{0, 1, \ldots, N-1\}$. Consider approximation $I_{(i,j),n}^{K+}$ defined by equation (V.41)*, where $K \in \mathbb{N}$ and $i, j \in \{1, \ldots, m\}$. It holds*

$$\begin{aligned}
&\max_{i,j\in\{1,\ldots,m\}} \|I_{(i,j),n} - I_{(i,j),n}^{K+}\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}\left(\left(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} + 1\right)^2 + 2(m-2)\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\right)^{\frac{1}{2}} \frac{\big(\sqrt{3}\sqrt{p-1}+1\big)h_n}{3\pi K},
\end{aligned}$$

where in particular

$$\max_{j\in\{1,\ldots,m\}} \|I_{(j,j),n} - I^{K+}_{(j,j),n}\|_{L^p(\Omega;\mathbb{R})} = 0.$$

Proof. The proof is stated in Section V.5, see p. 174. □

This theorem of course holds true for $p = 2$ as well. However, the constant is greater than the one in Theorem V.8 because the estimates with respect to $L^p(\Omega;\mathbb{R})$-norms neglect that covariances may vanish or cancel out each other as in case of $p = 2$, cf. formulas (V.81), (V.82), (V.83), and (V.86) in the proof of Theorem V.7.

Using the notations from the previous Section V.1, we provide an algorithm for simulating the nondelayed-iterated stochastic integrals $I_{(i,j),n}$, $i, j \in \{1, \ldots, m\}$ and $n \in \{0, 1, \ldots, N-1\}$, approximately. Covariance $\Sigma^K_{2,n}$ contains random variable

$$\sum_{k=K+1}^{\infty} \frac{b_{k,n}}{k} \sim \mathrm{N}\bigg(0_{m\times 1}, \Big(\sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^4}\Big) I_m\bigg).$$

According to [74, Corollary 6.11], there exists a $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variable $G_{2,n}$ such that

$$\sum_{k=K+1}^{\infty} \frac{b_{k,n}}{k} = \Big(\frac{h_n}{2\pi^2} \sum_{k=K+1}^{\infty} \frac{1}{k^4}\Big)^{\frac{1}{2}} G_{2,n}$$

and

$$\Sigma^K_{2,n} = \frac{h_n}{2\pi}\Big(\frac{h_n}{2\pi^2} \sum_{k=K+1}^{\infty} \frac{1}{k^4}\Big)^{\frac{1}{2}} H_m(P_m - I_{m^2})(G_{2,n} \otimes I_m)$$

P-almost surely.

Algorithm V.10
Let $\{t_0, t_1, \ldots, t_N\}$ be a discretization of $[t_0, T]$ and $p \in [2, \infty[$. In order to simulate ΔW^j_n and $I_{(i,j),n}$ for $i, j \in \{1, \ldots, m\}$ and $n \in \{0, 1, \ldots, N-1\}$ such that

$$\max_{i,j\in\{1,\ldots,m\}} \|I_{(i,j),n} - I^{K_n+}_{(i,j),n}\|_{L^p(\Omega;\mathbb{R})} \leq \varepsilon$$

for some error bound $\epsilon > 0$, proceed as follows. For $n = 0, 1, \ldots, N-1$,

i) set

$$K_n = \begin{cases} \left\lceil \dfrac{\sqrt{m} h_n}{\sqrt{12}\pi\varepsilon} \right\rceil & \text{if } p = 2 \text{ and} \\ \Bigg\lceil \dfrac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} \bigg(\Big(\dfrac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} + 1\Big)^2 + 2(m-2)\dfrac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\bigg)^{\frac{1}{2}} & \\ \qquad \times \dfrac{\big(\sqrt{3}\sqrt{p-1}+1\big)h_n}{3\pi\varepsilon} \Bigg\rceil & \text{if } p \in \,]2, \infty[, \end{cases}$$

where $\lceil\cdot\rceil$ *is the ceiling function, and compute*

$$\sigma_2^{K_n} = \sum_{k=K_n+1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} - \sum_{k=1}^{K_n} \frac{1}{k^2} = \psi^{(1)}(K_n+1)$$

and

$$\sigma_4^{K_n} = \sum_{k=K_n+1}^{\infty} \frac{1}{k^4} = \frac{\pi^4}{90} - \sum_{k=1}^{K_n} \frac{1}{k^4} = \frac{1}{6}\psi^{(3)}(K_n+1),$$

where $\psi^{(i)}$ *is the polygamma function of order* i*, see e. g. [1, p. 260].*

ii) Generate the independently $\mathrm{N}(0_{m\times 1}, I_m)$*-distributed random variables* B_n*,* $G_{0,n}$*,* $U_{k,n}$*, and* $V_{k,n}$ *for* $k \in \{1,\dots,K_n\}$*.*

iii) Set $\Delta W_n = \sqrt{h_n} B_n$*, and compute*

$$\begin{aligned}\operatorname{vec}\big[(A_n^{K_n})^{\mathrm{T}}\big] &= \frac{h_n}{\sqrt{2}\pi}\sqrt{\sigma_2^{K_n}}\big(B_n \otimes G_{0,n} - G_{0,n} \otimes B_n\big) \\ &\quad + \frac{h_n}{2\pi}\sum_{k=1}^{K_n}\frac{1}{k}\big(U_{k,n} \otimes (V_{k,n} - \sqrt{2}B_n) - (V_{k,n} - \sqrt{2}B_n) \otimes U_{k,n}\big).\end{aligned}$$

iv) Independently generate $G_{1,n} \sim \mathrm{N}(0_{M\times 1}, I_M)$ *where* $M = \frac{1}{2}m(m-1)$ *and* $G_{2,n} \sim \mathrm{N}(0_{m\times 1}, I_m)$*, and approximate* $\operatorname{vec}\big[(A_n)^{\mathrm{T}}\big]$ *by*

$$\begin{aligned}\operatorname{vec}\big[(A_n^{K_n+})^{\mathrm{T}}\big] &= \operatorname{vec}\big[(A_n^{K_n})^{\mathrm{T}}\big] + \frac{h_n}{2\pi}\sqrt{\frac{\sigma_4^{K_n}}{\sigma_2^{K_n}}}(P_m - I_{m^2})(G_{2,n} \otimes I_m)G_{0,n} \\ &\quad - \frac{h_n}{\sqrt{2}\pi}\sqrt{\sigma_2^{K_n} - \frac{\sigma_4^{K_n}}{\sigma_2^{K_n}}}(P_m - I_{m^2})H_m^{\mathrm{T}} G_{1,n},\end{aligned}$$

where H_m *and* P_m *are defined in formulas* (V.27) *and* (V.29)*.*

v) Then, the approximation of $\operatorname{vec}\big[(I_n)^{\mathrm{T}}\big]$ *is computed as*

$$\operatorname{vec}\big[(I_n^{K_n})^{\mathrm{T}}\big] = \frac{h_n}{2}\big(B_n \otimes B_n - \operatorname{vec}[I_m]\big) + \operatorname{vec}\big[(A_n^{K_n+})^{\mathrm{T}}\big].$$

At this point, we remark that term

$$\Sigma_{2,n}^K (\Sigma_{1,n}^K)^{-1}\sqrt{\Sigma_{1,n}^K}G_{0,n} = \frac{h_n}{2\pi}\sqrt{\frac{\sigma_4^K}{\sigma_2^K}}H_m(P_m - I_{m^2})(G_{2,n} \otimes I_m)G_{0,n}$$

P-almost surely is also of order $\mathcal{O}(K^{-1})$ in $L^p(\Omega;\mathbb{R}^M)$ as $K \to \infty$. However, neglecting this term results in a larger error. Define that approximation by

$$\operatorname{vec}\big[(\tilde{I}_n^{K+})^{\mathrm{T}}\big] \coloneqq \operatorname{vec}\big[(I_n^K)^{\mathrm{T}}\big] + (I_{m^2} - P_m)H_m^{\mathrm{T}}\sqrt{\mathrm{E}[S_n^K]}G_{1,n}. \qquad \text{(V.44)}$$

Theorem V.11
Let $n \in \{0,1,\ldots,N-1\}$. Consider approximation $I^{K+}_{(i,j),n}$ defined by equation (V.44)*, where $K \in \mathbb{N}$ and $i,j \in \{1,\ldots,m\}$. It holds*

$$\max_{i\in\{1,\ldots,M\}} \big\| e_i^{\mathrm{T}} \Sigma^K_{2,n} (\Sigma^K_{1,n})^{-1} \sqrt{\Sigma^K_{1,n}} G_{0,n} \big\|_{L^2(\Omega;\mathbb{R})} \leq \frac{h_n}{\sqrt{6}\pi K},$$

and thus

$$\max_{i,j\in\{1,\ldots,m\}} \| I_{(i,j),n} - \tilde{I}^{K+}_{(i,j),n} \|_{L^2(\Omega;\mathbb{R})} \leq \frac{\big(\sqrt{m}+\sqrt{2}\big) h_n}{\sqrt{12}\pi K},$$

where in particular

$$\max_{j\in\{1,\ldots,m\}} \| I_{(j,j),n} - \tilde{I}^{K+}_{(j,j),n} \|_{L^2(\Omega;\mathbb{R})} = 0.$$

Proof. The proof is stated in Section V.5, see p. 179. □

Similarly to Theorem V.9, we can also extend the result of the previous theorem to general $p \in \,]2,\infty[$.

Theorem V.12
Let $p \in \,]2,\infty[$ and $n \in \{0,1,\ldots,N-1\}$. Consider approximation $I^{K+}_{(i,j),n}$ defined by equation (V.44)*, where $K \in \mathbb{N}$ and $i,j \in \{1,\ldots,m\}$. It holds*

$$\max_{i\in\{1,\ldots,M\}} \big\| e_i^{\mathrm{T}} \Sigma^K_{2,n} (\Sigma^K_{1,n})^{-1} \sqrt{\Sigma^K_{1,n}} G_{0,n} \big\|_{L^p(\Omega;\mathbb{R})} \leq \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \frac{2h_n}{\sqrt{3}\pi K},$$

and thus

$$\begin{aligned}
&\max_{i,j\in\{1,\ldots,m\}} \| I_{(i,j),n} - \tilde{I}^{K+}_{(i,j),n} \|_{L^p(\Omega;\mathbb{R})} \\
&\leq \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} \Bigg(\Bigg(\bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} + 1 \bigg)^2 + 2(m-2) \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}} \Bigg)^{\frac{1}{2}} \big(\sqrt{3}\sqrt{p-1}+1\big) \\
&\qquad + \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} \sqrt{12} \Bigg) \frac{h_n}{3\pi K},
\end{aligned}$$

where in particular

$$\max_{j\in\{1,\ldots,m\}} \| I_{(j,j),n} - \tilde{I}^{K+}_{(j,j),n} \|_{L^p(\Omega;\mathbb{R})} = 0.$$

Proof. The proof is stated in Section V.5, see p. 181. □

Approximating $I_{(i,j),n}$ by $\tilde{I}^{K+}_{(i,j),n}$ instead of $I^{K+}_{(i,j),n}$, we do not need to generate $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variable $G_{2,n}$. However, as we see in Section V.3, the savings in computational costs are not large enough to compensate for the larger error as $K \to \infty$ unless dimension m of Wiener process W is high.

V.3. Analysis of the Computational Costs

The algorithms introduced in the previous sections are analyzed below with respect to their computational costs. By computational costs, we mean the number of independently N(0, 1)-distributed random variables that need to be generated. The cost for one standard-normally distributed random variable is set to one.

The first cost analysis of different algorithms approximating nondelayed-iterated stochastic integrals was done by Milstein [105]. He compared a rectangle method, a trapezium method, and the Fourier method, cf. Algorithm V.4. As a result of this, it turned out that the Fourier method has the least computing effort, see [105, p. 100]. This is the reason why we only considered the Fourier method and not a rectangle method or a trapezium method as well. Moreover, the computational effort of Algorithm V.4 does not increase by modeling the delayed-iterated stochastic integrals since the random variables that were already generated are reused.

In the following, we compare the costs caused by the computation of approximations $I^K_{(i,j),n}$, $I^{K+}_{(i,j),n}$, and $\tilde{I}^{K+}_{(i,j),n}$ as well as approximation $I^{(K)'}_{(i,j),n}$ of Wiktorsson [136] for $i, j \in \{1, \dots, m\}$, $K \in \mathbb{N}$, and an arbitrary $n \in \{0, 1, \dots, N-1\}$. The approximations $I^K_{(i,j),n}$, $I^{K+}_{(i,j),n}$, and $\tilde{I}^{K+}_{(i,j),n}$ are defined in formulas (V.20), (V.41), and (V.44), respectively.

Given some $K \in \mathbb{N}$, the Fourier method $I^K_{(i,j),n}$ involves $2(K+1)m$ independent, N(0, 1)-distributed random variables in order to approximate the iterated stochastic integrals $I_{(i,j),n}$ for $i, j \in \{1, \dots, m\}$, cf. Algorithm V.4. We write

$$\mathrm{cost}[I^K_n] = (2K+2)m. \tag{V.45}$$

Furthermore, we have

$$\mathrm{cost}[I^{K+}_n] = (2K+3)m + \frac{m(m-1)}{2}, \tag{V.46}$$

cf. Algorithm V.10, and

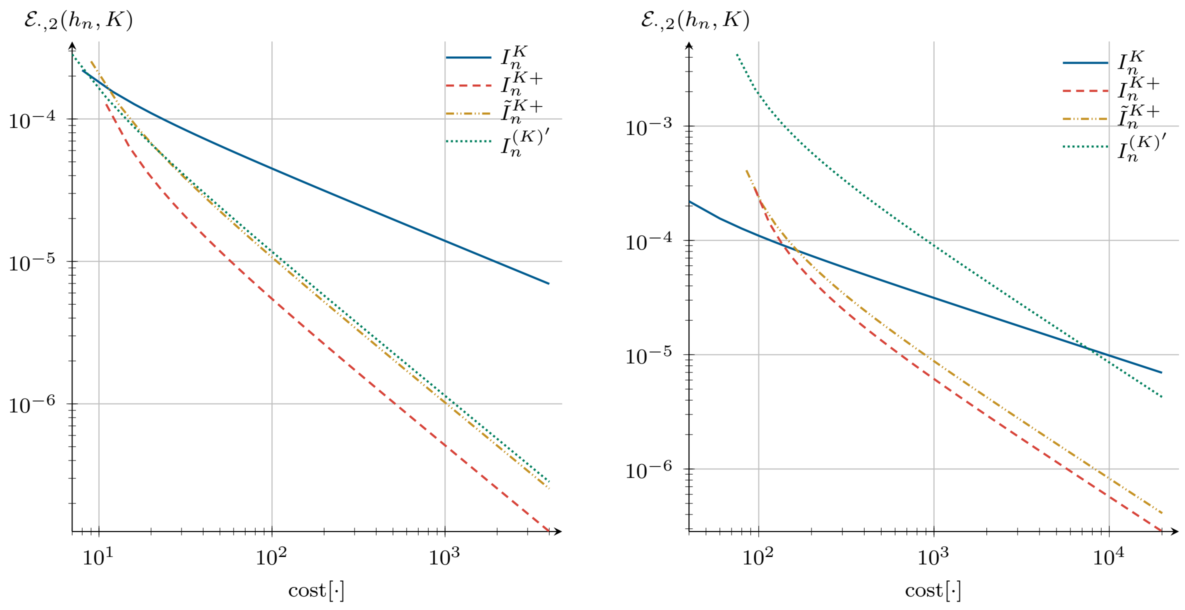
$$\mathrm{cost}[\tilde{I}^{K+}_n] = (2K+2)m + \frac{m(m-1)}{2}. \tag{V.47}$$

For approximation $I^{(K)'}_{(i,j),n}$ proposed by Wiktorsson in [136], it holds

$$\mathrm{cost}[I^{(K)'}_n] = (2K+1)m + \frac{m(m-1)}{2}. \tag{V.48}$$

In Figure V.13, we present the error bounds $\mathcal{E}_{\cdot,2}(h_n, K) \geq \max_{i,j \in \{1,\dots,m\}} \|I_{(i,j),n} - \cdot\|_{L^2(\Omega;\mathbb{R})}$ of Theorem V.2, Theorem V.8, Theorem V.11, and inequality (V.43) versus these computational costs for $K \in \{1, \dots, 1000\}$ where $h_n = 2^{-10}$. The Wiener process is $m = 2$ dimensional in Figure V.13 i) on the left and $m = 10$ dimensional in Figure V.13 ii) on the right. At first, we see that approximations I^{K+}_n, $\tilde{I}^{K+}_n$, and $I^{(K)'}_n$ reduce the root mean square error in the same and higher order compared to the simple approximation I^K_n from Algorithm V.4 (blue-solid line) without the approximation of the remainder of the Fourier series. Although we have

$$\mathrm{cost}[I^{K+}_n] > \mathrm{cost}[\tilde{I}^{K+}_n] > \mathrm{cost}[I^{(K)'}_n]$$



i) The dimension of Wiener process W is $m = 2$. **ii)** The dimension of Wiener process W is $m = 10$.

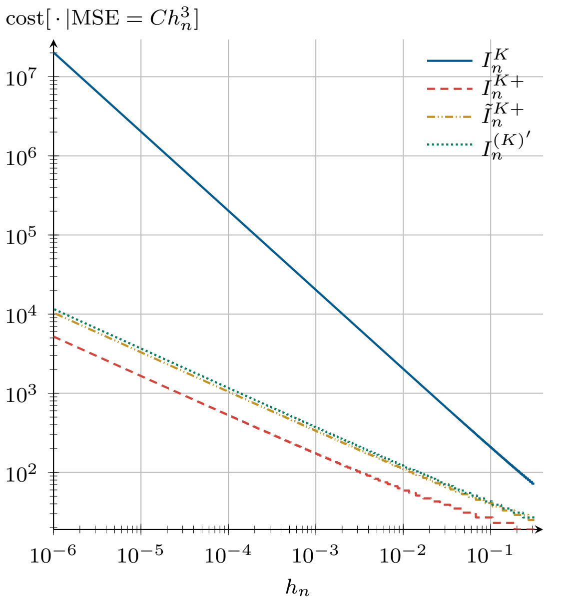
Figure V.13. Consider the approximations $I_{(i,j),n}^{K}$, $I_{(i,j),n}^{K+}$, $\tilde{I}_{(i,j),n}^{K+}$, and $I_{(i,j),n}^{(K)'}$ of the iterated stochastic integral $I_{(i,j),n}$ for $i,j \in \{1,\dots,m\}$ and an arbitrary fixed $n \in \{0,1,\dots,N-1\}$. The error bounds $\mathcal{E}_{\cdot,2}(h_n,K) \geq \max_{i,j\in\{1,\dots,m\}} \|I_{(i,j),n} - \cdot\,\|_{L^2(\Omega;\mathbb{R})}$ of these approximations versus their computational costs cost[·] are presented for parameter $K \in \{1,\dots,1000\}$ and step size $h_n = 2^{-10}$. We refer to Theorem V.2, Theorem V.8, Theorem V.11, and inequality (V.43) for the error bounds and to equations (V.45), (V.46), (V.47), and (V.48) for the computational costs. In Figure i) and Figure ii), the scales of both axes are logarithmic.

for all $K \in \mathbb{N}$, the inverse ordering of the error bounds

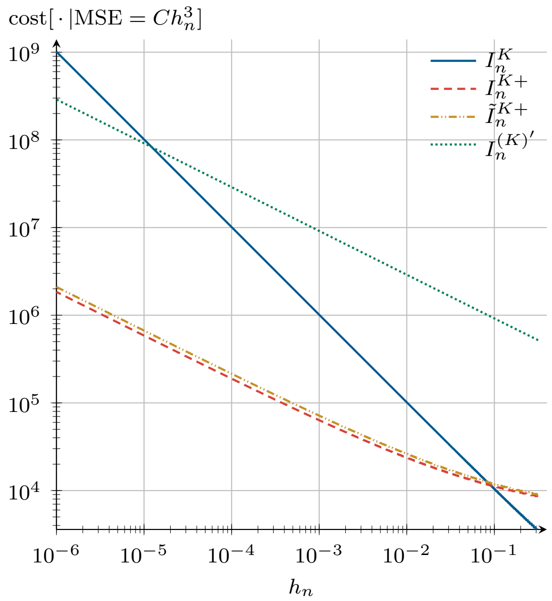
$$\frac{\sqrt{m}h_n}{\sqrt{12}\pi K} < \frac{(\sqrt{m}+\sqrt{2})h_n}{\sqrt{12}\pi K} < \frac{\sqrt{5m^2(m-1)}h_n}{\sqrt{24}\pi K}$$

of the algorithms I_n^{K+}, $\tilde{I}_n^{K+}$, and $I_n^{(K)'}$ yields that our algorithm I_n^{K+} (red-dashed line) is (asymptotically) the most efficient one. Only for larger error bounds and higher dimensions m of the Wiener process, the simple approximation I_n^K from Algorithm V.4 (blue-solid line) is preferable, see Figure V.13 ii). However, both our algorithms I_n^{K+} and $\tilde{I}_n^{K+}$ are more efficient than algorithm $I_n^{(K)'}$ (green-dotted line) proposed by Wiktorsson in [136].

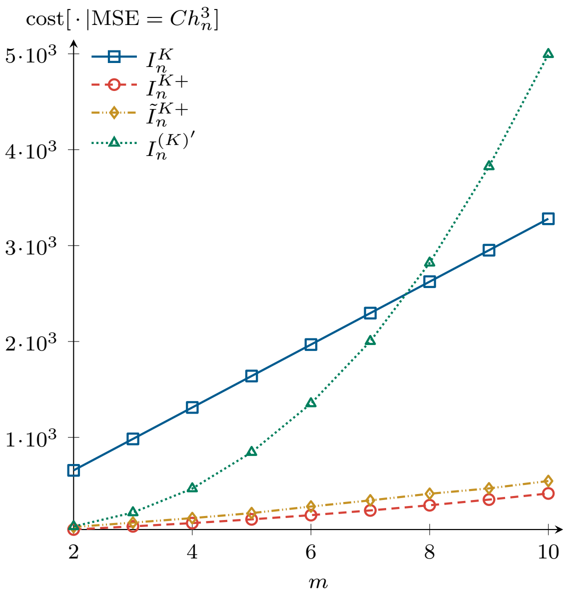
In the following, we analyze the computational cost under the change of step size h_n and dimension m of the Wiener process. However, the computational costs introduced above are not really meaningful in this regard. The different algorithms have different error bounds, and the costs in equations (V.45), (V.46), (V.47), and (V.48) do only depend on K and m. Therefore, we compare the computational effort that is required in order to ensure a mean square error $C \cdot h_n^3$, where $C > 0$ is a constant. This mean square error is motivated by the convergence analysis of the Milstein scheme. A mean square error of $C \cdot h_n^3$ in the modeling of the iterated stochastic integrals ensures the convergence of order $\mathcal{O}(h)$ of the Milstein scheme as maximum step size $h \to 0$, cf. [105, Theorem 7.1] in case of SODEs as well as Lemma V.16 of the following section in case of SDDEs.



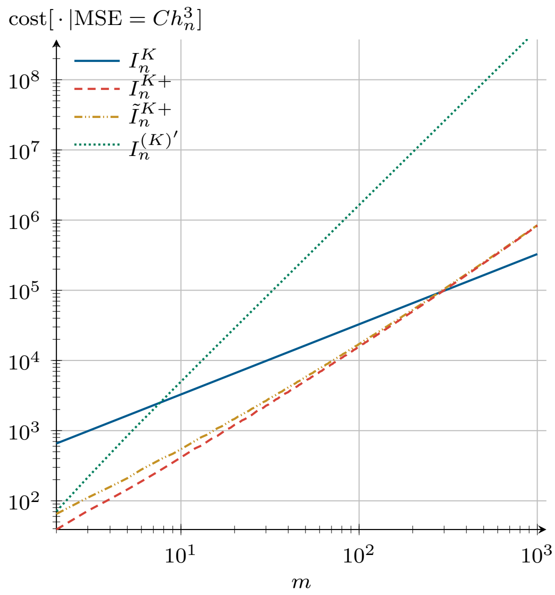
i) The dimension of Wiener process W is $m = 2$. The scales of both axes are logarithmic.



ii) The dimension of Wiener process W is $m = 100$. The scales of both axes are logarithmic.



iii) The step size is $h_n = 2^{-5}$.



iv) The step size is $h_n = 2^{-5}$. The scales of both axes are logarithmic.

Figure V.14. Consider the approximations $I_{(i,j),n}^{K}$, $I_{(i,j),n}^{K+}$, $\tilde{I}_{(i,j),n}^{K+}$, and $I_{(i,j),n}^{(K)'}$ of the iterated stochastic integral $I_{(i,j),n}$ for $i, j \in \{1, \ldots, m\}$ and an arbitrary fixed $n \in \{0, 1, \ldots, N-1\}$. Given a mean square error of $C \cdot h_n^3$ where $C = \frac{1}{100}$, the conditional computational costs, see equations (V.49), (V.50), (V.51), and (V.52), versus step size h_n and dimension m of the Wiener process, respectively, are presented.

In order to ensure a mean square error $C \cdot h_n^3$ of the algorithm I_n^K, that is

$$\max_{i,j \in \{1,\dots,m\}} \|I_{(i,j),n} - I_{(i,j),n}^K\|_{L^2(\Omega;\mathbb{R})}^2 \leq C \cdot h_n^3,$$

we have to choose $K = \lceil (2\pi^2 C h_n)^{-1} \rceil$ according to inequality (V.22). We write

$$\text{cost}[I_n^K|\text{MSE} = Ch_n^3] = \left(2 \cdot \left\lceil \frac{1}{2\pi^2 C h_n} \right\rceil + 2\right) m \tag{V.49}$$

for the computational costs that are needed to ensure the mean square error $C \cdot h_n^3$. Similarly, we have by Theorem V.8, Theorem V.11, and [136, Inequality (4.9)] that

$$\text{cost}[I_n^{K+}|\text{MSE} = Ch_n^3] = \left(2 \cdot \left\lceil \frac{\sqrt{m}}{\sqrt{12}\pi\sqrt{Ch_n}} \right\rceil + 3\right) m + \frac{m(m-1)}{2}, \tag{V.50}$$

$$\text{cost}[\tilde{I}_n^{K+}|\text{MSE} = Ch_n^3] = \left(2 \cdot \left\lceil \frac{\sqrt{m}+\sqrt{2}}{\sqrt{12}\pi\sqrt{Ch_n}} \right\rceil + 2\right) m + \frac{m(m-1)}{2}, \tag{V.51}$$

and

$$\text{cost}[I_n^{(K)'}|\text{MSE} = Ch_n^3] = \left(2 \cdot \left\lceil \frac{\sqrt{5m^2(m-1)}}{\sqrt{24}\pi\sqrt{Ch_n}} \right\rceil + 1\right) m + \frac{m(m-1)}{2}, \tag{V.52}$$

respectively.

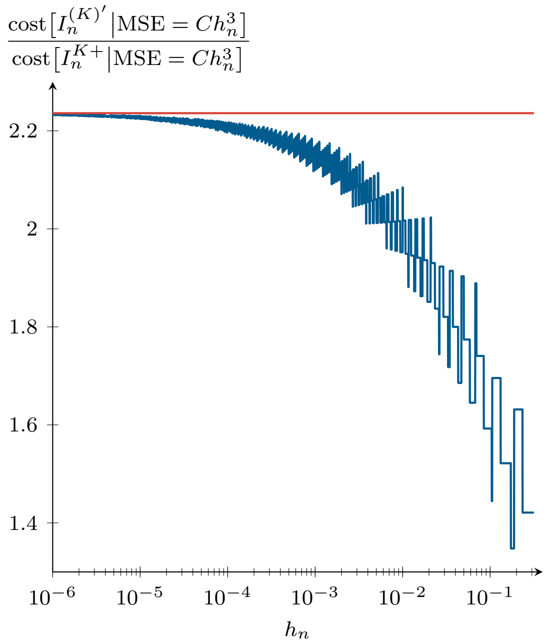
Here, we see for simple Fourier method I_n^K that $\text{cost}[I_n^K|\text{MSE} = Ch_n^3] = \mathcal{O}(h_n^{-1})$ as $h_n \to 0$ whereas the other algorithms are of order $\mathcal{O}(h_n^{-\frac{1}{2}})$ as $h_n \to 0$. Moreover, we have for the lower bounds of the cut off parameter $K \in \mathbb{N}$ of algorithms I_n^{K+}, $\tilde{I}_n^{K+}$, and $I_n^{(K)'}$ that

$$\frac{\sqrt{m}}{\sqrt{12}\pi\sqrt{Ch_n}} < \frac{\sqrt{m}+\sqrt{2}}{\sqrt{12}\pi\sqrt{Ch_n}} < \frac{\sqrt{5m^2(m-1)}}{\sqrt{24}\pi\sqrt{Ch_n}}$$

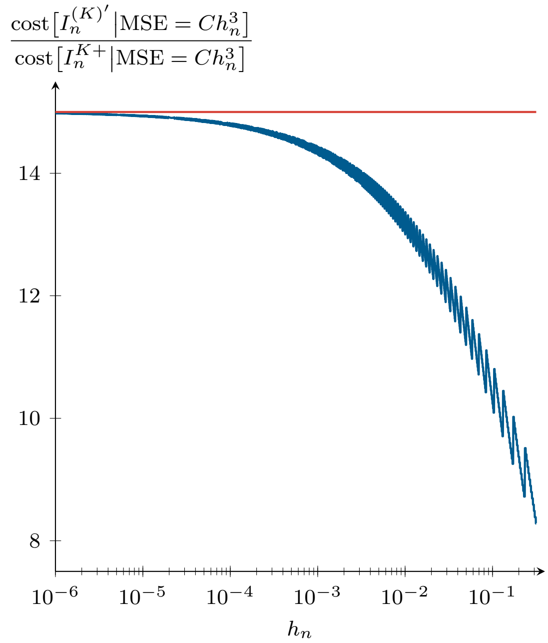
for all $m \in \mathbb{N} \setminus \{1\}$, $C > 0$, and $h_n > 0$.

These properties are illustrated in Figure V.14 i) and Figure V.14 ii), where $m = 2$ and $m = 100$, respectively. The stair steps in the presented figures result from the ceil function. Figure V.14 i) and Figure V.14 ii) clarify that our algorithm I_n^{K+} (red-dashed line) asymptotically has the lowest computational costs and is therefore preferable to the others. In particular for larger dimensions m of the Wiener process, we see in Figure V.14 ii) that Wiktorsson's algorithm $I_n^{(K)'}$ (green-dotted line) only has lower computational costs compared to simple approximation I_n^K (blue-solid line) for very small step sizes h_n.

Considering the growth of the computational cost given a mean square error as $m \to \infty$, we see from the conditional costs above that $\text{cost}[I_n^K|\text{MSE} = Ch_n^3] = \mathcal{O}(m)$ as $m \to \infty$. Further, our algorithms I_n^{K+} and $\tilde{I}_n^{K+}$ are of order $\mathcal{O}(m^2)$ whereas the algorithm $I_n^{(K)'}$ derived by Wiktorsson is of order $\mathcal{O}(m^{\frac{5}{2}})$ as $m \to \infty$. See Figure V.14 iii) and Figure V.14 iv). Here, we especially see the advantage of our algorithms I_n^{K+} (red-dashed line) and $\tilde{I}_n^{K+}$ (yellow-dash-dotted line) derived in the previous sections in comparison to algorithm $I_n^{(K)'}$ (green-dotted line) of Wiktorsson. Only in case of a fixed moderate mean square error the simple algorithm I_n^K (blue-solid line) may be preferable for very high dimension m whereas Wiktorsson's algorithm



i) The dimension of Wiener process W is $m = 2$. The red line represents the factor $\frac{\sqrt{5m(m-1)}}{\sqrt{2}}$ of asymptotic savings in computational effort, which equals $\sqrt{5}$.



ii) The dimension of Wiener process W is $m = 10$. The red line represents the factor $\frac{\sqrt{5m(m-1)}}{\sqrt{2}}$ of asymptotic savings in computational effort, which equals 15.

Figure V.15. Consider the approximations $I_{(i,j),n}^{K+}$ and $I_{(i,j),n}^{(K)'}$ of the iterated stochastic integral $I_{(i,j),n}$ for $i, j \in \{1, \ldots, m\}$ and an arbitrary fixed $n \in \{0, 1, \ldots, N-1\}$. Given a mean square error of $C \cdot h_n^3$ where $C = \frac{1}{100}$, the ratio of the conditional computational costs of Wiktorsson's algorithm $I_{(i,j),n}^{(K)'}$ in [136] to Algorithm V.10 versus the step size h_n is shown. The abscissa is logarithmically scaled in both Figure i) and Figure ii).

is more and more inefficient, see Figure V.14 iv). Moreover, Figure V.14 iv) illustrates that for higher dimensions m of the Wiener process, the computational costs of our algorithms I_n^{K+} and $\tilde{I}_n^{K+}$ are of similar size. Hence, for higher dimensions m, it might be reasonable to use algorithm $\tilde{I}_n^{K+}$ because we do not need to generate random variable $G_{2,n} \sim \mathrm{N}(0_{m\times 1}, I_m)$, and thus, we have not to calculate the term

$$\frac{h_n}{2\pi}\sqrt{\frac{\sigma_4^{K_n}}{\sigma_2^{K_n}}}(P_m - I_{m^2})(G_{2,n} \otimes I_m)G_{0,n}$$

in Algorithm V.10 step *iv)*. Moreover, in case of an equidistant discretization, we can choose $K_n = K$ for all $n \in \{0, 1, \ldots, N-1\}$, and thus, the matrix

$$\frac{h_n}{\sqrt{2}\pi}\sqrt{\sigma_2^{K_n} - \frac{\sigma_4^{K_n}}{\sigma_2^{K_n}}}(P_m - I_{m^2})H_m^{\mathrm{T}}$$

in Algorithm V.10 step *iv)* only needs to be computed once. In contrast to this, the random matrix $\sqrt{\Sigma_\infty}$ in Wiktorsson's algorithm $I_n^{(K)'}$, cf. [136, Equation (4.5) and (4.7)], needs to be computed in every time step, even in case of an equidistant discretization. Thus, algorithm $\tilde{I}_n^{K+}$ is easier to implement and requires fewer arithmetic operations.

As already mentioned in the previous section, our method I_n^{K+} improves, compared to Wiktorsson's algorithm $I_n^{(K)'}$, the error bound by a factor of $\frac{\sqrt{5m(m-1)}}{\sqrt{2}}$. This asymptotically reduces the computational cost by the factor $\frac{\sqrt{5m(m-1)}}{\sqrt{2}}$ as $h_n \to 0$. This is illustrated in Figure V.15. Figure V.15 i) and Figure V.15 ii) show that we need $\sqrt{5}$-times and 15-times fewer standard-normally distributed random variables in case of $m = 2$ and $m = 10$, respectively. These savings in computational costs result in a much more efficient simulation of iterated stochastic integrals and reduces the computing time significantly.

Even if we consider the error criteria

$$\sum_{\substack{i,j=1 \\ i<j}}^{m} \| I_{(i,j),n} - I_{(i,j),n}^{K+} \|_{L^2(\Omega;\mathbb{R})}^2 \leq \frac{m^2(m-1)h_n^2}{24\pi^2 K^2} \tag{V.53}$$

and

$$\sum_{\substack{i,j=1 \\ i<j}}^{m} \| I_{(i,j),n} - I_{(i,j),n}^{(K)'} \|_{L^2(\Omega;\mathbb{R})}^2 \leq \frac{5m^2(m-1)h_n^2}{24\pi^2 K^2},$$

cf. Theorem V.8 and [136, Theorem 4.1], our algorithm I_n^{K+} improves the error bound by a constant factor of 5 for all $m \in \mathbb{N}$ with $m \geq 2$. Thus, the computational cost of Wiktorsson's algorithm is asymptotically $\sqrt{5}$ times larger as $h_n \to 0$. Hence, for smaller step sizes h_n our method I_n^{K+} approximately halves the number of standard-normally distributed random variables that need to be generated compared to Wiktorsson's algorithm $I_n^{(K)'}$.

If we however consider the stronger error estimates presented in Theorem V.8 and Theorem V.11, the savings in computational costs of our method I_n^{K+} are even greater for $m \in \mathbb{N}$ with $m > 2$, cf. Figure V.15 ii). These stronger estimates in Theorem V.8 and Theorem V.11 are in particular valuable if not all iterated stochastic integrals $I_{(i,j),n}$, $i, j \in \{1, \dots, m\}$ where $i \neq j$, have to be simulated.

Consider for example the following SODE where $d = 1$ and $m = 3$. Let $a(x) = 0$, $b^1(x) = 2x$, $b^2(x) = x$, and $b^3(x) = 1$ for all $x \in \mathbb{R}$, that is

$$X_t = \begin{cases} 1 & \text{if } t = 0 \text{ and} \\ 1 + \int_0^t 2X_s \, \mathrm{d}W_s^1 + \int_0^t X_s \, \mathrm{d}W_s^2 + \int_0^t 1 \, \mathrm{d}W_s^3 & \text{if } t \in [t_0, T]. \end{cases}$$

Since

$$\frac{\mathrm{d}b^1(x)}{\mathrm{d}x} b^2(x) = 2 \cdot x = 1 \cdot 2x = \frac{\mathrm{d}b^2(x)}{\mathrm{d}x} b^1(x),$$

cf. commutativity condition (V.1), and since

$$\frac{\mathrm{d}b^3(x)}{\mathrm{d}x} = 0$$

for all $x \in \mathbb{R}$, we only need to simulate the iterated stochastic integrals $I_{(1,3),n}$ and $I_{(2,3),n}$. The cut off parameter $K \in \mathbb{N}$, determined by the estimates in Theorem V.8 and Theorem V.11, is much smaller than the one resulting from estimate (V.53) for algorithm I_n^{K+} or a similar one for method $\tilde{I}_n^{K+}$. This additionally emphasizes the value of our results in the previous section.

V.4. Milstein Scheme with Approximated Iterated Stochastic Integrals

We now consider the Milstein scheme from Chapter IV, see equation (IV.33). If the underlying SDDE does not have additive noise nor satisfies commutative condition (V.1), the iterated stochastic integrals in Milstein scheme (IV.33) have to be substituted by approximations. We provide conditions so that the Milstein scheme still converges strongly with order $\alpha = 1$.

Let $\bar{I}^{K_n}_{(i,j),n,\tau_l} \in L^p(\Omega;\mathbb{R})$ be an approximation of iterated stochastic integral $I_{(i,j),n,\tau_l}$ such that

$$\max_{\substack{i,j\in\{1,\dots,m\}\\ l\in\{0,1,\dots,D\}}} \big\|\bar{I}^{K_n}_{(i,j),n,\tau_l}\big\|_{L^p(\Omega;\mathbb{R})} \le C_{\bar{I},p} h_n \tag{V.54}$$

and

$$\max_{\substack{i,j\in\{1,\dots,m\}\\ l\in\{0,1,\dots,D\}}} \big\|I_{(i,j),n,\tau_l} - \bar{I}^{K_n}_{(i,j),n,\tau_l}\big\|_{L^p(\Omega;\mathbb{R})} \le \mathcal{E}_{\bar{I},p}(h_n, K_n) \tag{V.55}$$

for all $n \in \{0,1,\dots,N-1\}$, where $K_n \in \mathbb{N}$, $C_{\bar{I},p} > 0$ is a constant, and $\mathcal{E}_{\bar{I},p} \colon \mathbb{R}_+ \times \mathbb{N} \to \mathbb{R}_+$ is a function. We still assume that discretization $\{t_0, t_1, \dots, t_M\} \subset [t_0, T]$ of form (V.6) whenever $D > 0$ as stated in the introduction of this chapter. The Milstein scheme with approximated iterated stochastic integrals is defined by

$$\begin{aligned} &\overline{Y}_t = \xi_t \quad \text{for } t \in [t_0 - \tau, t_0] \text{ and} \\ &\overline{Y}_{t_{n+1}} = \overline{Y}_{t_n} + a(\mathcal{T}(t_n, \overline{Y}_{t_n}))h_n + \sum_{j=1}^{m} b^j(\mathcal{T}(t_n, \overline{Y}_{t_n}))\Delta W_n^j \\ &\qquad + \sum_{l=0}^{D} \sum_{j_1,j_2=1}^{m} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_n, \overline{Y}_{t_n})) b^{i,j_2}(\mathcal{T}((t_n - \tau_l) \vee t_0, \overline{Y}_{(t_n-\tau_l)\vee t_0}))\bar{I}^{K_n}_{(j_2,j_1),n,\tau_l} \\ &\text{for } n = 0,1,\dots,N-1. \end{aligned} \tag{V.56}$$

The properties on the convergence of this approximation of SDDE (II.1) are stated in the following lemma.

Lemma V.16
Let the Borel-measurable coefficients of SDDE (II.1) *fulfill Assumption IV.8 ii), iii), and iv), where $b^j(t, t-\tau_1, \dots, t-\tau_D, \cdot, \dots, \cdot) \in C^1(\mathbb{R}^{d\times(D+1)};\mathbb{R}^d)$ for all $t \in [t_0, T]$ and $j \in \{1, \dots, m\}$. Further, let initial condition ξ belong to $S^{2(\beta+1)p}([t_0 - \tau, t_0] \times \Omega; \mathbb{R}^d)$ for some $p \in [2, \infty[$, where $\beta \in [0, \infty[$ is determined by Assumption IV.8 iii).*

For $i, j \in \{1, \dots, m\}$, $l \in \{0, 1, \dots, D\}$, and $n \in \{0, 1, \dots, N-1\}$, let approximation $\bar{I}^{K_n}_{(i,j),n,\tau_l}$ fulfill assumptions (V.54) *and* (V.55)*, be $\mathscr{F}_{t_{n+1}}/\mathscr{B}(\mathbb{R})$-measurable, be independent of σ-algebra $\mathscr{F}_{(t_n-\tau_l)\vee t_0}$ and satisfy $\mathrm{E}\big[\bar{I}^{K_n}_{(i,j),n,\tau_l}\big|\mathscr{F}_{t_n}\big] = 0$ P-almost surely.*

Consider the families of Milstein approximations $(Y^h)_{h\in]0,T-t_0]}$ and $(\overline{Y}^h)_{h\in]0,T-t_0]}$ regarding SDDE (II.1) *from equations* (IV.33) *and* (V.56)*, where both schemes Y^h and $\overline{Y}^h$ have maximum* *step size h. Let there exists a constant $C_{\text{Milstein}} > 0$, independent of h and N, such that*

$$\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - Y_t^h\| \Big\|_{L^p(\Omega;\mathbb{R})} \leq C_{\text{Milstein}} h \tag{V.57}$$

for all $h \in\,]0, T-t_0]$, where X is the solution of SDDE (II.1)*.*

Then, it holds

$$\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - \overline{Y}_t^h\| \Big\|_{L^p(\Omega;\mathbb{R})} \leq \overline{C}_1 h + \overline{C}_2 \Big(\sum_{n=0}^{N-1} \big(\mathcal{E}_{\bar{I},p}(h_n, K_n)\big)^2 \Big)^{\frac{1}{2}}$$

for all $h \in\,]0, T-t_0]$, where $\overline{C}_1, \overline{C}_2 > 0$ are constants that are independent of h and N.

If $\sum_{n=0}^{N-1} \big(\mathcal{E}_{\bar{I},p}(h_n, K_n)\big)^2 \in \mathcal{O}(h^2)$ as $h \to 0$, the family of Milstein schemes $(\overline{Y}^h)_{h\in]0,T-t_0]}$ is strongly convergent with order $\alpha = 1$ to solution X of SDDE (II.1) *as $h \to 0$. That is, there exists a constant $\overline{C}_{\text{Milstein}} > 0$, independent of h and N, such that*

$$\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - \overline{Y}_t^h\| \Big\|_{L^p(\Omega;\mathbb{R})} \leq \overline{C}_{\text{Milstein}} h$$

for all $h \in\,]0, T-t_0]$.

Proof. We refer to Section V.5, see p. 182, for the proof of this lemma and details on the constants. □

We remark that assumption (V.57) in Lemma V.16 holds true when the conditions supposed in Theorem IV.9 on the convergence of Milstein approximation (IV.33) are fulfilled.

If the assumptions of the previous lemma are fulfilled for all $p \in [2, \infty[$, we also obtain by Lemma IV.3 the pathwise convergence of Milstein scheme $\overline{Y}$ with approximated iterated stochastic integrals.

Corollary V.17
Let the assumptions of Lemma V.16 be fulfilled for all $p \in [2, \infty[$. Consider the family of Milstein approximations $(\overline{Y}^{h_N})_{N\in\mathbb{N}}$ regarding SDDE (II.1) *from equation* (V.56)*, where $(h_N)_{N\in\mathbb{N}} \subset]0, T-t_0]$. Let $q_\varepsilon \in [1, \infty[$ for all $\varepsilon > 0$ be independent of N and such that $\sum_{N=1}^{\infty} h_N^{\varepsilon q_\varepsilon} < \infty$. Moreover, let*

$$\sum_{n=0}^{N-1} \big(\mathcal{E}_{\bar{I},p}(h_n, K_n)\big)^2 \in \mathcal{O}(h_N^2),$$

where $h_N = \max_{n\in\{0,1,\dots,N-1\}} h_n$ is the maximum step size of discretization $\{t_0, t_1, \dots, t_N\}$.

Then, the family of Milstein approximations $(\overline{Y}^{h_N})_{N\in\mathbb{N}}$ converges pathwise with order $\alpha = 1-\varepsilon$ to solution X of SDDE (II.1) *for arbitrary $\varepsilon > 0$ as $N \to \infty$. That is, for all $\varepsilon > 0$, there exists a positive random variable Z_ε, which belongs to $L^p(\Omega;\mathbb{R})$ for all $p \in [1, \infty[$, such that*

$$\sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - \overline{Y}_t^{h_N}\| \leq Z_\varepsilon\, h_N^{1-\varepsilon}$$

P*-almost surely for all $N \in \mathbb{N}$.*

Using Lemma V.16 and Corollary V.17, we obtain the strong and pathwise convergence of Milstein scheme (V.56) where the iterated stochastic integrals are approximated by Algorithm V.4 and Algorithm V.10, respectively.

Considering Milstein scheme (V.56) where $\bar{I}^{K_n}_{(j_2,j_1),n,\tau_l} = I^{K_n}_{(j_2,j_1),n,\tau_l}$, see Algorithm V.4, the following holds.

Theorem V.18
Let the Borel-measurable coefficients of SDDE (II.1) *fulfill Assumption IV.8 ii), iii), and iv), where* $b^j(t, t-\tau_1, \dots, t-\tau_D, \cdot, \dots, \cdot) \in C^1(\mathbb{R}^{d\times(D+1)}; \mathbb{R}^d)$ *for all* $t \in [t_0, T]$ *and* $j \in \{1, \dots, m\}$. *Further, let initial condition* ξ *belong to* $S^{2(\beta+1)p}([t_0-\tau, t_0]\times\Omega; \mathbb{R}^d)$ *for some* $p \in [2, \infty[$, *where* $\beta \in [0, \infty[$ *is determined by Assumption IV.8 iii).*

Consider the family of Milstein approximations $(Y^h)_{h\in]0,T-t_0]}$ *regarding SDDE* (II.1) *from equations* (IV.33). *Let there exist a constant* $C_{\text{Milstein}} > 0$, *independent of* h *and* N, *such that*

$$\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - Y^h_t\| \Big\|_{L^p(\Omega;\mathbb{R})} \le C_{\text{Milstein}} h$$

for all $h \in]0, T-t_0]$, *where* X *is the solution of SDDE* (II.1).

Consider the family of Milstein approximations $(\overline{Y}^h)_{h\in]0,T-t_0]}$ *defined in equation* (V.56) *with approximated iterated stochastic integrals* $\bar{I}^{K_n}_{(j_2,j_1),n,\tau_l} = I^{K_n}_{(j_2,j_1),n,\tau_l}$ *from Algorithm V.4, where* $\overline{Y}^h$ *has the maximum step size* h. *Let there exist a constant* $C > 0$, *independent of* h, h_n, n, *and* N, *such that parameter* $K_n \ge C h^{-1}$ *for all* $n \in \{0, 1, \dots, N-1\}$.

Then, the family of Milstein approximations $(\overline{Y}^h)_{h\in]0,T-t_0]}$ *is strongly convergent with order* $\alpha = 1$ *to solution* X *of SDDE* (II.1) *as* $h \to 0$. *That is, there exists a constant* $\overline{C}_{\text{Milstein}} > 0$, *independent of* h *and* N, *such that*

$$\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - \overline{Y}^h_t\| \Big\|_{L^p(\Omega;\mathbb{R})} \le \overline{C}_{\text{Milstein}} h$$

for all $h \in]0, T-t_0]$.

Furthermore, consider the subfamily of Milstein approximations $(\overline{Y}^{h_N})_{N\in\mathbb{N}}$, *where* $(h_N)_{N\in\mathbb{N}} \subset]0, T-t_0]$. *Let* $q_\varepsilon \in [1, \infty[$ *for all* $\varepsilon > 0$ *be independent of* N *and such that* $\sum_{N=1}^\infty h_N^{\varepsilon q_\varepsilon} < \infty$.

If the assumptions above are fulfilled for all $p \in [2, \infty[$, *the subfamily of Milstein approximations* $(\overline{Y}^{h_N})_{N\in\mathbb{N}}$ *converges pathwise with order* $\alpha = 1-\varepsilon$ *to solution* X *of SDDE* (II.1) *for arbitrary* $\varepsilon > 0$ *as* $N \to \infty$. *That is, for all* $\varepsilon > 0$, *there exists a positive random variable* Z_ε, *which belongs to* $L^p(\Omega; \mathbb{R})$ *for all* $p \in [1, \infty[$, *such that*

$$\sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - \overline{Y}^{h_N}_t\| \le Z_\varepsilon\, h_N^{1-\varepsilon}$$

P-*almost surely for all* $N \in \mathbb{N}$.

Proof. The proof is stated in Section V.5, see p. 194. □

In [60], only the convergence in $L^2(\Omega;\mathbb{R})$ of the Milstein scheme with approximated iterated stochastic integrals is considered. Our Theorem V.18 extends the results in [60] not only to the convergence in $L^p(\Omega;\mathbb{R})$ for arbitrary $p \in [2,\infty[$ but also to the pathwise convergence. Moreover, we improve the results from [60, Appendix B] even in case of $p = 2$. The authors in [60] assume that $K_n = \mathcal{O}(h^{-2})$ as $h \to 0$ whereas we can only suppose $K_n = \mathcal{O}(h^{-1})$. This reduces the computational cost significantly, cf. Algorithm V.4 and Section V.3.

If the diffusion coefficients of SDDE (II.1) do not depend on the past history of the solution, that is, they are of the form $t \mapsto b^j(t, t-\tau_1, \dots, t-\tau_D, X_t)$ for all $t \in [t_0, T]$ and $j \in \{1, \dots, m\}$, only the nondelayed-iterated stochastic integrals $I_{(i,j),n}$ for $i, j \in \{1, \dots, m\}$ appear in the Milstein scheme and have to be modeled. Then, the computational cost can be further reduced by using Algorithm V.10. In this case, the Milstein scheme (V.56) reads as

$$\begin{aligned}
&\overline{Y}_t = \xi_t \quad \text{for } t \in [t_0 - \tau, t_0] \text{ and} \\
&\overline{Y}_{t_{n+1}} = \overline{Y}_{t_n} + a(\mathcal{T}(t_n, \overline{Y}_{t_n})))h_n + \sum_{j=1}^{m} b^j(t_n, t_n - \tau_1, \dots, t_n - \tau_D, \overline{Y}_{t_n}))\Delta W_n^j \\
&\qquad + \sum_{j_1,j_2=1}^{m} \sum_{i=1}^{d} \partial_{x_0^i} b^{j_1}(t_n, t_n - \tau_1, \dots, t_n - \tau_D, \overline{Y}_{t_n})) \\
&\qquad\quad \times b^{i,j_2}(t_n, t_n - \tau_1, \dots, t_n - \tau_D, \overline{Y}_{t_n})) I^{K_n+}_{(j_2,j_1),n} \\
&\text{for } n = 0, 1, \dots, N-1.
\end{aligned} \tag{V.58}$$

Here, we can even suppose that $K_n = \mathcal{O}(h^{-\frac{1}{2}})$ as $h \to 0$, only, in order to achieve a strong convergence of order $\mathcal{O}(h)$ for the Milstein scheme. Consult the following theorem for more details.

Theorem V.19
Let the Borel-measurable coefficients of SDDE (II.1) *fulfill Assumption IV.8 ii), iii), and iv), where the diffusion coefficients are of the form $t \mapsto b^j(t, t-\tau_1, \dots, t-\tau_D, X_t)$ and $b^j(t, t-\tau_1, \dots, t-\tau_D, \cdot) \in C^1(\mathbb{R}^d;\mathbb{R}^d)$ for all $t \in [t_0, T]$ and $j \in \{1, \dots, m\}$. Further, let initial condition ξ belong to $S^{2(\beta+1)p}([t_0 - \tau, t_0] \times \Omega;\mathbb{R}^d)$ for some $p \in [2,\infty[$, where $\beta \in [0,\infty[$ is determined by Assumption IV.8 iii).*

Consider the family of Milstein approximations $(Y^h)_{h\in]0,T-t_0]}$ regarding SDDE (II.1) *from equations* (IV.33). *Let there exist a constant $C_{\text{Milstein}} > 0$, independent of h and N, such that*

$$\left\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - Y_t^h\| \right\|_{L^p(\Omega;\mathbb{R})} \le C_{\text{Milstein}} h$$

for all $h \in]0, T - t_0]$, where X is the solution of SDDE (II.1).

Consider the family of Milstein approximations $(\overline{Y}^h)_{h\in]0,T-t_0]}$ defined in equation (V.58) *with approximated iterated stochastic integrals $I^{K_n+}_{(j_2,j_1),n}$ from Algorithm V.10, where $\overline{Y}^h$ has the maximum step size h. Let there exist a constant $C > 0$, independent of h, h_n, n, and N, such that parameter $K_n \ge Ch^{-\frac{1}{2}}$ for all $n \in \{0, 1, \dots, N-1\}$.*

Then, the family of Milstein approximations $(\overline{Y}^h)_{h\in]0,T-t_0]}$ is strongly convergent with order $\alpha = 1$ to solution X of SDDE (II.1) *as $h \to 0$. That is, there exists a constant $\overline{C}_{\text{Milstein}} > 0$,* *independent of h and N, such that*

$$\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|X_t - \overline{Y}_t^h\| \Big\|_{L^p(\Omega;\mathbb{R})} \leq \overline{C}_{\text{Milstein}} h$$

for all $h \in \,]0, T-t_0]$.

Furthermore, consider the subfamily of Milstein approximations $(\overline{Y}^{h_N})_{N\in\mathbb{N}}$, where $(h_N)_{N\in\mathbb{N}} \subset \,]0, T-t_0]$. Let $q_\varepsilon \in [1,\infty[$ for all $\varepsilon > 0$ be independent of N and such that $\sum_{N=1}^{\infty} h_N^{\varepsilon q_\varepsilon} < \infty$.

If the assumptions above are fulfilled for all $p \in [2,\infty[$, the subfamily of Milstein approximations $(\overline{Y}^{h_N})_{N\in\mathbb{N}}$ converges pathwise with order $\alpha = 1-\varepsilon$ to solution X of SDDE (II.1) *for arbitrary $\varepsilon > 0$ as $N \to \infty$. That is, for all $\varepsilon > 0$, there exists a positive random variable Z_ε, which belongs to $L^p(\Omega;\mathbb{R})$ for all $p \in [1,\infty[$, such that*

$$\sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|X_t - \overline{Y}_t^{h_N}\| \leq Z_\varepsilon\, h_N^{1-\varepsilon}$$

P-*almost surely for all $N \in \mathbb{N}$.*

Proof. The proof is stated in Section V.5, see p. 195. □

From the proof of Theorem V.19, it is evident that the statement of Theorem V.19 also holds true for the iterated stochastic integral approximations $\tilde{I}^{K_n+}_{(j_2,j_1),n}$, $j_1, j_2 \in \{1,\ldots,m\}$, from Theorem V.11 and Theorem V.12.

In the following, we compare the computational costs of the Euler-Maruyama scheme as well as the Milstein schemes in Theorem V.18 and Theorem V.19. The computational cost of a scheme, as in Section V.3, is measured by the number of standard-normally distributed random variables that need to be generated.

In order to achieve an error of order $\mathcal{O}(h)$ as $h \to 0$, Euler-Maruyama scheme (IV.13) with a maximum step size h^2 has to be applied. This results in a computational cost of $\mathcal{O}(h^{-2})$. Using the Milstein scheme in Theorem V.18 with maximum step size h, where the iterated stochastic integrals are approximated by Algorithm V.4, the computational effort is likewise $\mathcal{O}(h^{-2})$ due to the computational cost of the integral approximations, see equation (V.45). If the diffusion coefficients do not depend on the past history of the solution, the Milstein scheme in Theorem V.19, where we used Algorithm V.10, can be applied. Its computational cost is of order $\mathcal{O}(h^{-\frac{3}{2}})$ only. Thus, if the assumptions of Theorem V.19 are fulfilled, the Milstein scheme in Theorem V.19, where the iterated stochastic integrals approximations are obtained by Algorithm V.10, is the method of choice. See also [136, p. 472] for the discussion on the computational effort in case of SODEs.

In [136, p. 472], Wiktorsson argues that there is no gain using a Milstein scheme like in Theorem V.18 instead of the Euler-Maruyama scheme because they have the same order of computational complexity $\mathcal{O}(h^{-2})$, and because the Euler-Maruyama scheme is easier to implement. Wiktorsson further argues that the Euler-Maruyama scheme needs less computational effort in practice if the evaluation of the SDDEs' coefficients is not too time-consuming compared to the generation of normally distributed random variables.

However, the Milstein scheme needs only $\mathcal{O}(h^{-1})$ steps in time whereas the Euler-Maruyama scheme needs $\mathcal{O}(h^{-2})$ steps, which can only be computed sequentially. Thus, on parallel computers, the Milstein scheme can reduce the computing time as the normally distributed random variables in each step in time can be generated in parallel. Thus, the Milstein scheme in Theorem V.18 is preferable to the Euler-Maruyama scheme in certain situations.

V.5. Proofs

Proof of Theorem V.2

The proof needs the following result on absolute moments of normally distributed random variables. This result is of course not new, however, we did not find any paper or book to cite.

Lemma V.20
Let $p \in [1, \infty[$, and let G be a $\mathrm{N}(0, \sigma^2)$-distributed random variable, where $\sigma \in \mathbb{R}$ with $\sigma > 0$. It holds

$$\|G\|_{L^p(\Omega;\mathbb{R})} = \frac{\sqrt{2}\sigma\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}.$$

Proof. Using the symmetry of the normal distribution and the substitution $y = \frac{x^2}{2\sigma^2}$, we have

$$\mathrm{E}[|G|^p] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} |x|^p \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \,\mathrm{d}x = \frac{2}{\sqrt{2\pi\sigma^2}} \int_0^{\infty} x^p \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \,\mathrm{d}x = \frac{(2\sigma^2)^{\frac{p}{2}}}{\sqrt{\pi}} \int_0^{\infty} y^{\frac{p-1}{2}} \mathrm{e}^{-y} \,\mathrm{d}y.$$

Since $\int_0^{\infty} y^{\frac{p-1}{2}} \mathrm{e}^{-y} \,\mathrm{d}y = \Gamma(\frac{p+1}{2})$, the assertion follows by taking the pth root. □

Proof of Theorem V.2. It is evident that $\|I_{(j,j),n} - I^K_{(j,j),n}\|_{L^p(\Omega;\mathbb{R})} = 0$ for $j \in \{1, \ldots, m\}$. In the following, we therefore assume that $i \neq j$ in case of $l = 0$. By rewriting, cf. equations (V.13) and (V.14) as well as equations (V.20) and (V.21), we obtain at first that

$$\begin{aligned}
&\|I_{(i,j),n,\tau_l} - I^K_{(i,j),n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})} \\
&\quad= \bigg\| \int_{t_n}^{t_{n+1}} \int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W^i_u \,\mathrm{d}W^j_s - \bigg(\frac{1}{2} \Delta W^i_{n,\tau_l} \Delta W^j_n + \frac{a^i_{0,n,\tau_l}}{2} \Delta W^j_n - \frac{a^j_{0,n}}{2} \Delta W^i_{n,\tau_l} \\
&\qquad + \pi \sum_{k=1}^{K} k(a^i_{k,n,\tau_l} b^j_{k,n} - b^i_{k,n,\tau_l} a^j_{k,n}) \bigg) \bigg\|^2_{L^p(\Omega;\mathbb{R})} \\
&\quad= \bigg\| \int_{t_n}^{t_{n+1}} \int_{t_n-\tau_l}^{s-\tau_l} \mathrm{d}W^i_u - \bigg(\frac{s-t_n}{h_n} \Delta W^i_{n,\tau_l} + \frac{a^i_{0,n,\tau_l}}{2} \\
&\qquad + \sum_{k=1}^{K} a^i_{k,n,\tau_l} \cos\Big(\frac{2\pi}{h_n} k(s-t_n) \Big) + b^i_{k,n,\tau_l} \sin\Big(\frac{2\pi}{h_n} k(s-t_n) \Big) \bigg) \,\mathrm{d}W^j_s \bigg\|^2_{L^p(\Omega;\mathbb{R})}.
\end{aligned}$$

Next, we apply Burkholder's inequality, see Theorem II.4. Here, we can use our usual filtration $(\mathscr{F}_t)_{t\in[t_n,t_{n+1}]}$ in case of $l \in \{1, \ldots, D\}$. In case of $l = 0$, the stochastic integral inside the

$L^p(\Omega;\mathbb{R})$-norm on the right-hand side of the above equation is $\mathscr{F}_{t_{n+1}}/\mathscr{B}(\mathbb{R})$-measurable but as process not adapted with respect to $(\mathscr{F}_t)_{t\in[t_n,t_{n+1}]}$ anymore. However, it is still martingale with respect to a filtration $(\mathscr{F}_t^{i,j})_{t\in[t_n,t_{n+1}]}$ defined by

$$\mathscr{F}_t^{i,j} = \sigma\big(\{W_s^j - W_{t_0}^j : s \in [t_n, t]\} \cup \{W_s^i - W_{t_0}^i : s \in [t_n, t_{n+1}]\} \cup \mathcal{N}\big),$$

where $\mathcal{N} = \{A \in \mathscr{F} : \mathrm{P}[A] = 0\}$. Thus, Burkholder's inequality is applicable, and we obtain

$$\begin{aligned}
&\|I_{(i,j),n,\tau_l} - I^K_{(i,j),n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})} \\
&\quad\le (p-1)^2\bigg\|\int_{t_n}^{t_{n+1}}\bigg|\int_{t_n-\tau_l}^{s-\tau_l}\mathrm{d}W_u^i - \bigg(\frac{s-t_n}{h_n}\Delta W^i_{n,\tau_l} + \frac{a^i_{0,n,\tau_l}}{2} \\
&\qquad + \sum_{k=1}^{K} a^i_{k,n,\tau_l}\cos\Big(\frac{2\pi}{h_n}k(s-t_n)\Big) + b^i_{k,n,\tau_l}\sin\Big(\frac{2\pi}{h_n}k(s-t_n)\Big)\bigg)\bigg|^2\mathrm{d}s\bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \\
&\quad= (p-1)^2\bigg\|\int_{t_n}^{t_{n+1}}\bigg|\sum_{k=K+1}^{\infty} a^i_{k,n,\tau_l}\cos\Big(\frac{2\pi}{h_n}k(s-t_n)\Big) + b^i_{k,n,\tau_l}\sin\Big(\frac{2\pi}{h_n}k(s-t_n)\Big)\bigg|^2\mathrm{d}s\bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})}.
\end{aligned} \tag{V.59}$$

The equality in the last formula holds according to expansion (V.8). Using Parseval's formula, that is

$$\begin{aligned}
&\frac{2}{h_n}\int_{t_n}^{t_{n+1}}\bigg|\sum_{k=K+1}^{\infty} a^i_{k,n,\tau_l}\cos\Big(\frac{2\pi}{h_n}k(s-t_n)\Big) + b^i_{k,n,\tau_l}\sin\Big(\frac{2\pi}{h_n}k(s-t_n)\Big)\bigg|^2\mathrm{d}s \\
&\quad= \sum_{k=K+1}^{\infty}|a^i_{k,n,\tau_l}|^2 + |b^i_{k,n,\tau_l}|^2
\end{aligned}$$

P-almost surely, cf. [141, p. 37 in Volume I], and using the triangle inequality, it follows

$$\begin{aligned}
\|I_{(i,j),n,\tau_l} - I^K_{(i,j),n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})} &\le (p-1)^2\frac{h_n}{2}\bigg\|\sum_{k=K+1}^{\infty}|a^i_{k,n,\tau_l}|^2 + |b^i_{k,n,\tau_l}|^2\bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \\
&\le (p-1)^2\frac{h_n}{2}\sum_{k=K+1}^{\infty}\|a^i_{k,n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})} + \|b^i_{k,n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})}.
\end{aligned} \tag{V.60}$$

Since, for $k \in \mathbb{N}$, Fourier coefficients a^i_{k,n,τ_l} and b^i_{k,n,τ_l} are $\mathrm{N}\big(0, \frac{h_n}{2\pi^2k^2}\big)$-distributed random variables, Lemma V.20 implies

$$\|a^i_{k,n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})} = \|b^i_{k,n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})} = \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}h_n}{\pi^{\frac{2p+1}{p}}k^2}.$$

Inserting this into inequality (V.60) and estimating the series, we have

$$\begin{aligned}
\|I_{(i,j),n,\tau_l} - I^K_{(i,j),n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})} &\le (p-1)^2\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}h_n^2}{\pi^{\frac{2p+1}{p}}}\sum_{k=K+1}^{\infty}\frac{1}{k^2} \\
&\le (p-1)^2\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}h_n^2}{\pi^{\frac{2p+1}{p}}}\int_K^{\infty}\frac{1}{x^2}\,\mathrm{d}x \\
&= (p-1)^2\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}h_n^2}{\pi^{\frac{2p+1}{p}}K}.
\end{aligned} \tag{V.61}$$

Since the right-hand side of the above inequality is independent of i, j, and l, we obtain

$$\max_{\substack{i,j\in\{1,\dots,m\}\\ l\in\{0,1,\dots,D\}}} \|I_{(i,j),n,\tau_l} - I^K_{(i,j),n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})} \leq (p-1)^2 \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}} h_n^2}{\pi^{\frac{2p+1}{p}} K},$$

and the assertion of Theorem V.2 follows by taking the square root. □

Proof of Corollary V.3. In case of $p = 2$, inequality (V.59) becomes an identity by Itô's isometry. The triangle inequality in formula (V.60) needs not to be applied since the norm simplifies to the expected value. There, we use the monotone convergence theorem and obtain

$$\|I_{(i,j),n,\tau_l} - I^K_{(i,j),n,\tau_l}\|^2_{L^2(\Omega;\mathbb{R})} = \frac{h_n}{2} \sum_{k=K+1}^{\infty} \|a^i_{k,n,\tau_l}\|^2_{L^2(\Omega;\mathbb{R})} + \|b^i_{k,n,\tau_l}\|^2_{L^2(\Omega;\mathbb{R})} = \frac{h_n^2}{2\pi^2} \sum_{k=K+1}^{\infty} \frac{1}{k^2}.$$

Then, the corollary follows from

$$\sum_{k=K+1}^{\infty} \frac{1}{k^2} = \sum_{k=1}^{\infty} \frac{1}{k^2} - \sum_{k=1}^{K} \frac{1}{k^2} = \frac{\pi^2}{6} - \sum_{k=1}^{K} \frac{1}{k^2}$$

and taking the square root. □

Proof of Lemma V.5

Proof of Lemma V.5. At first, let

$$B_k := \begin{pmatrix} I_m \\ \pi k H_m (P_m - I_{m^2})(b_{k,n} \otimes I_m) \end{pmatrix},$$

and for $k \in \mathbb{N}$, consider the random variable

$$B_k a_{k,n} = \begin{pmatrix} a_{k,n} \\ \pi k H_m (P_m - I_{m^2})(b_{k,n} \otimes a_{k,n}) \end{pmatrix}, \tag{V.62}$$

cf. formula (V.33). As Fourier coefficient $a_{k,n}$ is $\mathrm{N}\big(0_{m\times 1}, \frac{h_n}{2\pi^2 k^2} I_m\big)$-distributed for all $k \in \mathbb{N}$, its characteristic function $\varphi_{a_{k,n}} \colon \mathbb{R}^m \to \mathbb{C}$ is given by

$$\varphi_{a_{k,n}}(v) = \mathrm{E}\Big[\mathrm{e}^{\mathrm{i} v^{\mathrm{T}} a_{k,n}}\Big] = \mathrm{e}^{-\frac{1}{2} v^{\mathrm{T}} \frac{h_n}{2\pi^2 k^2} I_m v} = \mathrm{e}^{-\frac{h_n}{4\pi^2 k^2} v^{\mathrm{T}} v}$$

for $v \in \mathbb{R}^m$, where i is the imaginary unit, see [67, Theorem 16.1]. Let $C \in \mathbb{R}^{(m+M)\times m}$ be an arbitrary matrix. Then, according to [67, Theorem 13.3], for all $k \in \mathbb{N}$, we have

$$\varphi_{Ca_{k,n}}(u) = \varphi_{a_{k,n}}(C^{\mathrm{T}} u) = \mathrm{e}^{-\frac{h_n}{4\pi^2 k^2} (C^{\mathrm{T}} u)^{\mathrm{T}} (C^{\mathrm{T}} u)} = \mathrm{e}^{-\frac{h_n}{4\pi^2 k^2} u^{\mathrm{T}} C C^{\mathrm{T}} u} \tag{V.63}$$

for all $u \in \mathbb{R}^{m+M}$, and the random variable $Ca_{k,n}$ is $\mathrm{N}\big(0_{(m+M)\times 1}, \frac{h_n}{2\pi^2 k^2} CC^{\mathrm{T}}\big)$-distributed by [67, Theorem 16.1]. Here, $CC^{\mathrm{T}} \in \mathbb{R}^{(m+M)\times(m+M)}$ is a positive semidefinite matrix because

$$x^{\mathrm{T}} C C^{\mathrm{T}} x = (C^{\mathrm{T}} x)^{\mathrm{T}} (C^{\mathrm{T}} x) = \|C^{\mathrm{T}} x\|^2 \geq 0$$

for all $x \in \mathbb{R}^{m+M}$. Now, we calculate, given Fourier coefficient $b_{k,n}$, the conditional characteristic function of random variable $B_k a_{k,n}$ from formula (V.62). For the definition of conditional characteristic functions, we refer to [92, p. 26]. Taking into account that Fourier coefficients $a_{k,n}$ and $b_{k,n}$ are independent and using [16, Corollary 4.38] or [43, Satz 5.3.22], it holds by equation (V.63) that

$$\mathrm{E}\Big[\mathrm{e}^{\mathrm{i}u^{\mathrm{T}} B_k a_{k,n}}\Big| b_{k,n}\Big] = \mathrm{E}\Big[\varphi_{a_{k,n}}(B_k^{\mathrm{T}} u)\Big| b_{k,n}\Big] = \mathrm{e}^{-\frac{h_n}{4\pi^2 k^2} u^{\mathrm{T}} B_k B_k^{\mathrm{T}} u} \tag{V.64}$$

P-almost surely for all $u \in \mathbb{R}^{m+M}$. That is, given Fourier coefficient $b_{k,n}$, the conditional distribution of random variable $B_k a_{k,n}$ is normal with conditional expectation

$$\mathrm{E}[B_k a_{k,n} | b_{k,n}] = 0_{(m+M)\times 1}$$

P-almost surely and conditional covariance

$$\mathrm{E}\big[B_k a_{k,n} (B_k a_{k,n})^{\mathrm{T}} \big| b_{k,n}\big] = \frac{h_n}{2\pi^2 k^2} B_k B_k^{\mathrm{T}}$$

P-almost surely. In the following, we consider the random variable

$$\sum_{k=K+1}^{\infty} B_k a_{k,n} = \begin{pmatrix} \displaystyle\sum_{k=K+1}^{\infty} a_{k,n} \\ \displaystyle\pi \sum_{k=K+1}^{\infty} k H_m (P_m - I_{m^2})(b_{k,n} \otimes a_{k,n}) \end{pmatrix},$$

see formula (V.33). Recall that the series converges in $L^p(\Omega; \mathbb{R}^{m+M})$ for every $p \in [2, \infty[$, cf. Theorem V.2. Since $a_{k,n}$ and $b_{k,n}$, $k \in \mathbb{N}$, are independent, the series also converge P-almost surely by [66, Theorem 3.1]. Next, we use the independence of Fourier coefficients $a_{k,n}$ and $b_{k,n}$ for all $k \in \mathbb{N}$ in order to show that, given $\{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\}$, the conditional distribution of $\sum_{k=K+1}^{\infty} B_k a_{k,n}$ is normal. More precisely, we show that

$$\mathrm{E}\Big[\mathrm{e}^{\mathrm{i}u^{\mathrm{T}} \sum_{k=K+1}^{\infty} B_k a_{k,n}}\Big| \{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\}\Big] = \mathrm{e}^{-\frac{1}{2} u^{\mathrm{T}} \left(\sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^2} B_k B_k^{\mathrm{T}}\right) u} \tag{V.65}$$

P-almost surely for all $u \in \mathbb{R}^{m+M}$. By the definition of the conditional expectation, we have for all $\mathcal{B} \in \sigma(\{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\})$ that

$$\mathrm{E}\Big[\mathbb{1}_{\mathcal{B}} \mathrm{E}\Big[\mathrm{e}^{\mathrm{i}u^{\mathrm{T}} \sum_{k=K+1}^{\infty} B_k a_{k,n}}\Big| \{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\}\Big]\Big] = \mathrm{E}\Big[\mathbb{1}_{\mathcal{B}} \mathrm{e}^{\mathrm{i}u^{\mathrm{T}} \sum_{k=K+1}^{\infty} B_k a_{k,n}}\Big]. \tag{V.66}$$

Consider the $\cap$-stable generator

$$\mathcal{E} := \Big\{ \bigcap_{k=K+1}^{\infty} A_k : A_k \in \sigma(b_{k,n}) \Big\}$$

of σ-algebra $\sigma(\{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\})$. By linearity of the expectation and the dominated convergence theorem, it is enough to consider $\mathcal{B} \in \mathcal{E}$ in equation (V.66) only, since for all $\mathcal{B} \in \sigma(\{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\})$, there exist $\mathcal{B}_l \in \mathcal{E}$, $l \in \mathbb{N}$, with $\mathcal{B}_l \cap \mathcal{B}_k = \emptyset$ for $l \neq k$ and $\mathcal{B} = \cup_{l=1}^{\infty} \mathcal{B}_l$ such that $\mathbb{1}_{\mathcal{B}} = \sum_{l=1}^{\infty} \mathbb{1}_{\mathcal{B}_l}$ holds P-almost surely, cf. the monotone class theorem [67, Theorem 6.2].

Consider an arbitrary set $\mathcal{B} \in \mathcal{E}$ with $\mathcal{B} = \cap_{k=K+1}^{\infty} \mathcal{B}_k$ where $\mathcal{B}_k \in \sigma(b_{k,n})$. Since $\mathbb{1}_{\mathcal{B}} = \prod_{k=K+1}^{\infty} \mathbb{1}_{\mathcal{B}_k}$, it holds

$$\mathrm{E}\Big[\mathbb{1}_{\mathcal{B}} \mathrm{e}^{\mathrm{i} u^{\mathrm{T}} \sum_{k=K+1}^{\infty} B_k a_{k,n}}\Big] = \mathrm{E}\Big[\prod_{k=K+1}^{\infty} \mathbb{1}_{\mathcal{B}_k} \mathrm{e}^{\mathrm{i} u^{\mathrm{T}} B_k a_{k,n}}\Big].$$

Using, for all $k \in \mathbb{N}$, the independence of Fourier coefficients $a_{k,n}$ and $b_{k,n}$, and that set $\mathcal{B}_k$ is $\sigma(b_{k,n})$-measurable, we obtain

$$\mathrm{E}\Big[\mathbb{1}_{\mathcal{B}} \mathrm{e}^{\mathrm{i} u^{\mathrm{T}} \sum_{k=K+1}^{\infty} B_k a_{k,n}}\Big] = \prod_{k=K+1}^{\infty} \mathrm{E}\Big[\mathbb{1}_{\mathcal{B}_k} \mathrm{e}^{\mathrm{i} u^{\mathrm{T}} B_k a_{k,n}}\Big] = \prod_{k=K+1}^{\infty} \mathrm{E}\Big[\mathbb{1}_{\mathcal{B}_k} \mathrm{E}\Big[\mathrm{e}^{\mathrm{i} u^{\mathrm{T}} B_k a_{k,n}} \Big| b_{k,n}\Big]\Big]$$

Then, equation (V.64) implies

$$\mathrm{E}\Big[\mathbb{1}_{\mathcal{B}} \mathrm{e}^{\mathrm{i} u^{\mathrm{T}} \sum_{k=K+1}^{\infty} B_k a_{k,n}}\Big] = \prod_{k=K+1}^{\infty} \mathrm{E}\Big[\mathbb{1}_{\mathcal{B}_k} \mathrm{e}^{-\frac{h_n}{4\pi^2 k^2} u^{\mathrm{T}} B_k B_k^{\mathrm{T}} u}\Big].$$

Using again the independence of Fourier coefficients $b_{k,n}$, $k \in \mathbb{N}$, it follows

$$\begin{aligned}\mathrm{E}\Big[\mathbb{1}_{\mathcal{B}} \mathrm{e}^{\mathrm{i} u^{\mathrm{T}} \sum_{k=K+1}^{\infty} B_k a_{k,n}}\Big] &= \mathrm{E}\Big[\prod_{k=K+1}^{\infty} \mathbb{1}_{\mathcal{B}_k} \prod_{k=K+1}^{\infty} \mathrm{e}^{-\frac{h_n}{4\pi^2 k^2} u^{\mathrm{T}} B_k B_k^{\mathrm{T}} u}\Big] \\ &= \mathrm{E}\Big[\mathbb{1}_{\mathcal{B}} \mathrm{e}^{-\frac{1}{2} u^{\mathrm{T}} \left(\sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^2} B_k B_k^{\mathrm{T}}\right) u}\Big],\end{aligned}$$

and formula (V.65) holds true by equation (V.66). Here, the series

$$\sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^2} B_k B_k^{\mathrm{T}}$$

converges absolutely in the $L^{\frac{p}{2}}(\Omega; L_{HS}(\mathbb{R}^{m+M}; \mathbb{R}^{m+M}))$-norm for every $p \in [1, \infty[$ and also P-almost surely due to the independence of the summands, see [66, Theorem 3.1]. The Hilbert-Schmidt norm $\|\cdot\|_{L_{HS}(\mathbb{R}^{m+M};\mathbb{R}^{m+M})}$ coincides with the Frobenius norm $\|\cdot\|_{\mathrm{F}}$ for matrices of size $(m+M) \times (m+M)$.

According to the conditional characteristic function of $\sum_{k=K+1}^{\infty} B_k a_{k,n}$ in formula (V.65), given $\{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\}$, the conditional distribution of series $\sum_{k=K+1}^{\infty} B_k a_{k,n}$ is normal with conditional expectation

$$\mathrm{E}\Big[\sum_{k=K+1}^{\infty} B_k a_{k,n} \Big| \{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\}\Big] = 0_{(m+M)\times 1}$$

P-almost surely and conditional covariance

$$\mathrm{E}\Big[\Big(\sum_{k=K+1}^{\infty} B_k a_{k,n}\Big)\Big(\sum_{k=K+1}^{\infty} B_k a_{k,n}\Big)^{\mathrm{T}} \Big| \{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\}\Big] = \sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^2} B_k B_k^{\mathrm{T}}$$

P-almost surely. Then, defining

$$\begin{pmatrix} \Sigma_{1,n}^K & (\Sigma_{2,n}^K)^{\mathrm{T}} \\ \Sigma_{2,n}^K & \Sigma_{3,n}^K \end{pmatrix} := \sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^2} B_k B_k^{\mathrm{T}},$$

the assertion of this lemma follows. □

Proof of Lemma V.6

Proof of Lemma V.6. At first, it holds by equation (V.37) that

$$\mathrm{E}\big[S_n^K\big] = \mathrm{E}\big[\varSigma_{3,n}^K - \varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}(\varSigma_{2,n}^K)^{\mathrm{T}}\big]. \tag{V.67}$$

Using equations (V.34), (V.35), and (V.36), we have according to the distribution of the Fourier coefficients that

$$\begin{aligned}\mathrm{E}\big[\varSigma_{3,n}^K\big] &= \frac{h_n}{2} H_m(P_m - I_{m^2})\left(\mathrm{E}\Bigg[\sum_{k=K+1}^{\infty} b_{k,n} b_{k,n}^{\mathrm{T}}\Bigg] \otimes I_m\right)(P_m - I_{m^2})H_m^{\mathrm{T}} \\ &= \left(\frac{h_n}{2\pi}\right)^2 \left(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\right) H_m(P_m - I_{m^2})(I_m \otimes I_m)(P_m - I_{m^2})H_m^{\mathrm{T}}\end{aligned} \tag{V.68}$$

and

$$\begin{aligned}&\mathrm{E}\big[\varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}(\varSigma_{2,n}^K)^{\mathrm{T}}\big] \\ &\quad= \frac{h_n}{2}\left(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\right)^{-1} \\ &\qquad \times \mathrm{E}\Bigg[H_m(P_m - I_{m^2})\left(\sum_{k=K+1}^{\infty} \frac{b_{k,n}}{k} \otimes I_m\right)\left(\sum_{k=K+1}^{\infty} \frac{b_{k,n}^{\mathrm{T}}}{k} \otimes I_m\right)(P_m - I_{m^2})H_m^{\mathrm{T}}\Bigg] \\ &\quad= \frac{h_n}{2}\left(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\right)^{-1} H_m(P_m - I_{m^2})\left(\mathrm{E}\Bigg[\sum_{k=K+1}^{\infty} \frac{b_{k,n}}{k} \sum_{k=K+1}^{\infty} \frac{b_{k,n}^{\mathrm{T}}}{k}\Bigg] \otimes I_m\right)(P_m - I_{m^2})H_m^{\mathrm{T}} \\ &\quad= \left(\frac{h_n}{2\pi}\right)^2 \left(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\right)^{-1}\left(\sum_{k=K+1}^{\infty} \frac{1}{k^4}\right) H_m(P_m - I_{m^2})(I_m \otimes I_m)(P_m - I_{m^2})H_m^{\mathrm{T}}.\end{aligned} \tag{V.69}$$

Here, we can interchange taking the limit and the expectation because the series converge componentwise in $L^p(\Omega;\mathbb{R})$ for every $p \in [1,\infty[$, cf. [43, Satz 5.4.11]. Since $(I_m \otimes I_m) = I_{m^2}$ and $(P_m - I_{m^2})(P_m - I_{m^2}) = -2(P_m - I_{m^2})$, it follows

$$\begin{aligned}H_m(P_m - I_{m^2})(I_m \otimes I_m)(P_m - I_{m^2})H_m^{\mathrm{T}} &= -2H_m(P_m - I_{m^2})H_m^{\mathrm{T}} \\ &= -2H_m P_m H_m^{\mathrm{T}} + 2H_m H_m^{\mathrm{T}} \\ &= 2I_M,\end{aligned}$$

where $H_m P_m H_m^{\mathrm{T}} = 0_{M\times M}$ and $H_m H_m^{\mathrm{T}} = I_M$, see [136, p. 479]. Inserting this into equations (V.68) and (V.69), we obtain

$$\mathrm{E}\big[S_n^K\big] = 2\left(\frac{h_n}{2\pi}\right)^2\left(\left(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\right) - \left(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\right)^{-1}\left(\sum_{k=K+1}^{\infty} \frac{1}{k^4}\right)\right) I_M$$

by equation (V.67). Hence, matrix $\mathrm{E}\big[S_n^K\big]$ is diagonal, and the assertion of this lemma simply follows by taking the square root. □

Proof of Theorem V.7

We denote by $\|\cdot\|_{\mathrm{F}}$ the Frobenius norm in the following. The proof is based on the subsequent lemma on square root matrices. This lemma is similar to [136, Lemma 4.1]. However, our lemma allows to estimate single rows of a matrix.

Lemma V.21
Let $A \in \mathbb{R}^{M\times M}$ be symmetric positive semidefinite matrix and $B \in \mathbb{R}^{M\times M}$ be symmetric positive definite matrix. Assume that their matrix square roots $\sqrt{A}$ and $\sqrt{B}$ commute, that is $\sqrt{A}\sqrt{B} - \sqrt{B}\sqrt{A} = 0_{M\times M}$. Further, denote the smallest eigenvalue of B by $\lambda_{\min}$. Then, it holds

$$\big\|e_i^{\mathrm{T}}\big(\sqrt{A}-\sqrt{B}\big)\big\|_{\mathrm{F}} \leq \frac{1}{\sqrt{\lambda_{\min}}}\|e_i^{\mathrm{T}}(A-B)\|_{\mathrm{F}}$$

for all $i \in \{1,\dots,M\}$, where e_i is the ith unit vector of $\mathbb{R}^M$.

Proof. As matrices $\sqrt{A}$ and $\sqrt{B}$ commute, they are both simultaneously diagonalizable [59, Theorem 4.5.15]. Thus, the smallest eigenvalue $\sqrt{\lambda}$ of the matrix $\sqrt{A}+\sqrt{B}$ fulfills $\sqrt{\lambda} \geq \sqrt{\lambda_{\min}} > 0$, and $\sqrt{A}+\sqrt{B}$ is symmetric positive definite.

Since $\sqrt{A}+\sqrt{B}$ is regular and since $\sqrt{A}\sqrt{B}-\sqrt{B}\sqrt{A} = 0_{n\times n}$, we have

$$\sqrt{A}-\sqrt{B} = (\sqrt{A}-\sqrt{B})(\sqrt{A}+\sqrt{B})(\sqrt{A}+\sqrt{B})^{-1} = (A-B)(\sqrt{A}+\sqrt{B})^{-1}.$$

Due to this, it holds by submultiplicativity of the norms, cf. [73, p. 141], that

$$\big\|e_i^{\mathrm{T}}\big(\sqrt{A}-\sqrt{B}\big)\big\|_{\mathrm{F}} \leq \|e_i^{\mathrm{T}}(A-B)\|_{\mathrm{F}}\big\|(\sqrt{A}+\sqrt{B})^{-1}\big\|_2,$$

where $\|\cdot\|_2$ denotes the spectral norm. Since

$$\big\|(\sqrt{A}+\sqrt{B})^{-1}\big\|_2 = \frac{1}{\sqrt{\lambda}} \leq \frac{1}{\sqrt{\lambda_{\min}}},$$

the assertion is proved. □

In addition to the lemma above, we need sophisticated lower and upper bounds of the series $\sum_{k=K+1}^{\infty}\frac{1}{k^p}$.

Lemma V.22
Let $p \in \{2,6\}$. It holds

$$\sum_{k=K+1}^{\infty}\frac{1}{k^p} \geq \frac{1}{(p-1)\big(K+\frac{3}{4}\big)^{p-1}}$$

for all $K \in \mathbb{N}$.

Proof. At first, we prove for all $k \in \mathbb{N}$ and $p \in \{2,6\}$ that

$$\frac{1}{k^p} = \int_{k-\frac{1}{4}}^{k+\frac{3}{4}}\frac{1}{k^p}\,\mathrm{d}x \geq \int_{k-\frac{1}{4}}^{k+\frac{3}{4}}\frac{1}{x^p}\,\mathrm{d}x. \tag{V.70}$$

It holds

$$\int_{k-\frac{1}{4}}^{k+\frac{3}{4}} \frac{1}{x^2}\,\mathrm{d}x = \frac{1}{k^2+\frac{k}{2}-\frac{3}{16}} \leq \frac{1}{k^2}$$

and

$$\begin{aligned}\int_{k-\frac{1}{4}}^{k+\frac{3}{4}} \frac{1}{x^6}\,\mathrm{d}x &= \left(1+\frac{1}{k}+\frac{7}{8k^2}+\frac{5}{16k^3}+\frac{61}{1280k^4}\right)\\ &\quad\times\Bigg(1+\frac{5}{2k}+\frac{25}{16k^2}-\frac{5}{8k^3}-\frac{95}{128k^4}+\frac{23}{256k^5}+\frac{285}{2048k^6}\\ &\qquad-\frac{45}{2048k^7}-\frac{675}{65536k^8}+\frac{405}{131072k^9}-\frac{243}{1048576k^{10}}\Bigg)^{-1}\cdot\frac{1}{k^6}\\ &\leq \frac{1+\frac{1}{k}+\frac{1581}{1280k^2}}{1+\frac{1}{k}+\frac{25}{16k^2}+\frac{3}{2k}-\frac{1467683}{1048576k}}\cdot\frac{1}{k^6}\\ &\leq \frac{1+\frac{1}{k}+\frac{1581}{1280k^2}}{1+\frac{1}{k}+\frac{25}{16k^2}}\cdot\frac{1}{k^6}\\ &= \frac{1+\frac{1}{k}+\frac{1581}{1280k^2}}{1+\frac{1}{k}+\frac{2000}{1280k^2}}\cdot\frac{1}{k^6}\\ &\leq \frac{1}{k^6}.\end{aligned}$$

Using inequality (V.70) in order to approximate the summands of series $\sum_{k=K+1}^{\infty}\frac{1}{k^p}$ from below, we obtain

$$\sum_{k=K+1}^{\infty}\frac{1}{k^p} \geq \sum_{k=K+1}^{\infty}\int_{k-\frac{1}{4}}^{k+\frac{3}{4}}\frac{1}{x^p}\,\mathrm{d}x = \int_{K+\frac{3}{4}}^{\infty}\frac{1}{x^p}\,\mathrm{d}x = \frac{1}{(p-1)\left(K+\frac{3}{4}\right)^{p-1}}.$$

□

Lemma V.23
Let $p\in]1,\infty[$. It holds

$$\sum_{k=K+1}^{\infty}\frac{1}{k^p} \leq \frac{1}{(p-1)\left(K+\frac{1}{2}\right)^{p-1}}$$

for all $K\in\mathbb{N}$.

Proof. To begin with, we prove for all $k\in\mathbb{N}$ and $p\in]1,\infty[$ that

$$\frac{1}{k^p} = \int_{k-\frac{1}{2}}^{k+\frac{1}{2}}\frac{1}{k^p}\,\mathrm{d}x \leq \int_{k-\frac{1}{2}}^{k+\frac{1}{2}}\frac{1}{x^p}\,\mathrm{d}x. \tag{V.71}$$

In order to show inequality (V.71), the convex function $x\mapsto\frac{1}{x^p}$, $x\in[k-\frac{1}{2},k+\frac{1}{2}]$, is bounded from below by its tangent in $x=k$, that is by $x\mapsto f(x):=-\frac{p}{k^{p+1}}x+\frac{p+1}{k^p}$. Thus, we have

$$\int_{k-\frac{1}{2}}^{k+\frac{1}{2}}\frac{1}{x^p}\,\mathrm{d}x \geq \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} f(x)\,\mathrm{d}x = -\frac{p}{k^{p+1}}\int_{k-\frac{1}{2}}^{k+\frac{1}{2}} x\,\mathrm{d}x + \frac{p+1}{k^p} = \frac{1}{k^p},$$

which proves inequality (V.71). Approximating series $\sum_{k=K+1}^{\infty} \frac{1}{k^p}$ from above by integrals, see inequality (V.71), yields

$$\sum_{k=K+1}^{\infty} \frac{1}{k^p} \leq \sum_{k=K+1}^{\infty} \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \frac{1}{x^p} \,\mathrm{d}x = \int_{K+\frac{1}{2}}^{\infty} \frac{1}{x^p} \,\mathrm{d}x = \frac{1}{(p-1)\big(K+\frac{1}{2}\big)^{p-1}}.$$

□

Proof of Theorem V.7. The proof follows similar considerations as in [136, Proof of Theorem 4.3]. At first, we have

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \big\|e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big)G_{1,n}\big\|_{L^2(\Omega;\mathbb{R})}^2 \\
&\quad= \max_{i\in\{1,\dots,M\}} \mathrm{E}\Big[\mathrm{E}\Big[\big|e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big)G_{1,n}\big|^2\Big| b_{k,n}, k\in\mathbb{N}\Big]\Big] \\
&\quad= \max_{i\in\{1,\dots,M\}} \mathrm{E}\big[\big\|e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big)\big\|_{\mathrm{F}}^2\big]. \qquad \text{(V.72)}
\end{aligned}$$

Using that $\sqrt{\mathrm{E}[S_n^K]}$ is diagonal matrix by Lemma V.6, $\sqrt{\mathrm{E}[S_n^K]}$ commutes with square root matrix $\sqrt{S_n^K}$. Hence, it follows by Lemma V.21 that

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \mathrm{E}\big[\big\|e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big)\big\|_{\mathrm{F}}^2\big] \\
&\quad\leq \frac{1}{2}\Big(\frac{h_n}{2\pi}\Big)^{-2}\bigg(\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big) - \Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big)\bigg)^{-1} \qquad \text{(V.73)} \\
&\qquad\times \max_{i\in\{1,\dots,M\}} \mathrm{E}\big[\big\|e_i^{\mathrm{T}}\big(S_n^K - \mathrm{E}[S_n^K]\big)\big\|_{\mathrm{F}}^2\big].
\end{aligned}$$

Further, we have

$$\mathrm{E}\big[\big\|e_i^{\mathrm{T}}\big(S_n^K - \mathrm{E}[S_n^K]\big)\big\|_{\mathrm{F}}^2\big] = \sum_{j=1}^{M} \mathrm{E}\big[\big|(S_n^K)_{i,j} - (\mathrm{E}[S_n^K])_{i,j}\big|^2\big] = \sum_{j=1}^{M} \mathrm{Var}\big[(S_n^K)_{i,j}\big], \qquad \text{(V.74)}$$

where $(S_n^K)_{i,j}$ denotes the entry in the ith row and jth column, $i,j\in\{1,\dots,M\}$, of Schur complement S_n^K. We now take a closer look at the Schur complement and its entries. For sake of simplicity, define

$$C := \frac{h_n}{2}\bigg(\sum_{k=K+1}^{\infty} b_{k,n}b_{k,n}^{\mathrm{T}} - \Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}}{k}\Big)\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^{\mathrm{T}}}{k}\Big)\bigg). \qquad \text{(V.75)}$$

Then, equations (V.38) reads as

$$S_n^K = H_m(P_m - I_{m^2})(C\otimes I_m)(P_m - I_{m^2})H_m^{\mathrm{T}} \qquad \text{(V.76)}$$

P-almost surely. Using the definition of the permutation matrix P_m in (V.29) and that $I_{m^2} = I_m \otimes I_m = \sum_{i,j=1}^{m} e_ie_i^{\mathrm{T}} \otimes e_je_j^{\mathrm{T}}$, it holds

$$\begin{aligned}
P_m - I_{m^2} &= \sum_{i,j=1}^{m} (e_ie_j^{\mathrm{T}} \otimes e_je_i^{\mathrm{T}}) - (e_ie_i^{\mathrm{T}} \otimes e_je_j^{\mathrm{T}}) \\
&= \sum_{i,j=1}^{m} (e_i\otimes e_j)\big((e_j^{\mathrm{T}}\otimes e_i^{\mathrm{T}}) - (e_i^{\mathrm{T}}\otimes e_j^{\mathrm{T}})\big). \qquad \text{(V.77)}
\end{aligned}$$

Here, $\{e_i \otimes e_j : i,j \in \{1,\dots,m\}\}$ is the canonical basis of $\mathbb{R}^{m^2}$. Next, we apply selection matrix H_m, defined in formula (V.27), from the left to equation (V.77). Using that

$$H_m(e_i \otimes e_j) = \begin{cases} H_m(e_i \otimes e_j) & \text{if } i < j \text{ and} \\ 0_{M\times 1} & \text{if } j \geq i, \end{cases}$$

we obtain

$$H_m(P_m - I_{m^2}) = \sum_{\substack{i,j=1\\ i<j}}^{m} H_m(e_i \otimes e_j)\big((e_j^{\mathrm{T}} \otimes e_i^{\mathrm{T}}) - (e_i^{\mathrm{T}} \otimes e_j^{\mathrm{T}})\big). \tag{V.78}$$

Inserting equation (V.78) into formula (V.76), it holds for the Schur complement

$$\begin{aligned} S_n^K = &\sum_{\substack{i,j=1\\ i<j}}^{m} H_m(e_i \otimes e_j)\big((e_j^{\mathrm{T}} \otimes e_i^{\mathrm{T}}) - (e_i^{\mathrm{T}} \otimes e_j^{\mathrm{T}})\big)(C \otimes I_m) \\ &\times \sum_{\substack{k,l=1\\ k<l}}^{m} \big((e_l \otimes e_k) - (e_k \otimes e_l)\big)(e_k^{\mathrm{T}} \otimes e_l^{\mathrm{T}})H_m^{\mathrm{T}} \end{aligned}$$

P-almost surely. Since

$$(e_i^{\mathrm{T}} \otimes e_j^{\mathrm{T}})(C \otimes I_m)(e_k \otimes e_l) = e_i^{\mathrm{T}} C e_k \otimes e_j^{\mathrm{T}} I_m e_l = C_{i,k} \cdot e_j^{\mathrm{T}} e_l,$$

it follows

$$\begin{aligned} S_n^K = &\sum_{\substack{i,j=1\\ i<j}}^{m} \sum_{\substack{k,l=1\\ k<l}}^{m} H_m(e_i \otimes e_j) \\ &\times \big(C_{j,l} \cdot e_i^{\mathrm{T}} e_k - C_{j,k} \cdot e_i^{\mathrm{T}} e_l - C_{i,l} \cdot e_j^{\mathrm{T}} e_k + C_{i,k} \cdot e_j^{\mathrm{T}} e_l\big)(e_k^{\mathrm{T}} \otimes e_l^{\mathrm{T}})H_m^{\mathrm{T}} \end{aligned}$$

P-almost surely. Taking into account that $\{H_m(e_i \otimes e_j) : i,j \in \{1,\dots,m\} \text{ with } i<j\}$ is the canonical basis of $\mathbb{R}^M$, we can write

$$\begin{aligned} &\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{M} \mathrm{Var}\big[(S_n^K)_{i,j}\big] \\ &= \max_{\substack{i,j\in\{1,\dots,m\}\\ i<j}} \sum_{\substack{k,l=1\\ k<l}}^{m} \mathrm{Var}\big[C_{j,l} \cdot e_i^{\mathrm{T}} e_k - C_{j,k} \cdot e_i^{\mathrm{T}} e_l - C_{i,l} \cdot e_j^{\mathrm{T}} e_k + C_{i,k} \cdot e_j^{\mathrm{T}} e_l\big], \end{aligned} \tag{V.79}$$

where

$$\begin{aligned} &C_{j,l} \cdot e_i^{\mathrm{T}} e_k - C_{j,k} \cdot e_i^{\mathrm{T}} e_l - C_{i,l} \cdot e_j^{\mathrm{T}} e_k + C_{i,k} \cdot e_j^{\mathrm{T}} e_l \\ &= \begin{cases} C_{j,j} + C_{i,i} & \text{if } k = i \text{ and } l = j, \\ C_{j,l} & \text{if } k = i \text{ and } l \neq j, \\ C_{i,k} & \text{if } l = j \text{ and } k \neq i, \\ -C_{i,l} & \text{if } k = j, \\ -C_{j,k} & \text{if } l = i, \\ 0 & \text{else} \end{cases} \end{aligned}$$

under the constraint of $i < j$ and $k < l$. Since Fourier coefficients $b^i_{k,n}$ and $b^j_{k,n}$ with $i, j \in \{1, \dots, m\}$ and $i \neq j$ are independent and identically distributed, we further have

$$\begin{aligned}
&\operatorname{Var}\bigl[C_{j,l} \cdot e_i^{\mathrm{T}} e_k - C_{j,k} \cdot e_i^{\mathrm{T}} e_l - C_{i,l} \cdot e_j^{\mathrm{T}} e_k + C_{i,k} \cdot e_j^{\mathrm{T}} e_l\bigr] \\
&\quad = \begin{cases} 2\operatorname{Var}[C_{1,1}] & \text{if } k = i \text{ and } l = j, \\ \operatorname{Var}[C_{1,2}] & \text{if } k = i \text{ and } l \neq j, \\ \operatorname{Var}[C_{1,2}] & \text{if } l = j \text{ and } k \neq i, \\ \operatorname{Var}[C_{1,2}] & \text{if } k = j, \\ \operatorname{Var}[C_{1,2}] & \text{if } l = i, \\ 0 & \text{else.} \end{cases}
\end{aligned}$$

In order to calculate expression (V.79), we have to count how often the variance takes the value $\operatorname{Var}[C_{1,2}]$. Counting the number of tuples (k, l) with $k < l$ in each case given a tuple (i, j) with $i < j$, we obtain

$$\begin{aligned}
&\#\bigl\{(k,l) : k, l \in \{1, \dots, m\} \text{ with } k < l\bigr\} \\
&\quad = \begin{cases} 1 & \text{if } k = i \text{ and } l = j, \\ m - i - 1 & \text{if } k = i \text{ and } l \neq j, \\ j - 2 & \text{if } l = j \text{ and } k \neq i, \\ m - j & \text{if } k = j, \\ i - 1 & \text{if } l = i, \\ \frac{1}{2}(m-2)(m-3) & \text{else.} \end{cases}
\end{aligned} \tag{V.80}$$

The number from the last case results from the fact that we have $\frac{1}{2}m(m-1)$ tuples (k, l) with $k < l$ in total. The variance $\operatorname{Var}[C_{1,2}]$ occurs

$$(m - i - 1) + (j - 2) + (m - j) + (i - 1) = 2(m - 2)$$

times, and thus, it holds $\frac{1}{2}m(m-1) - 1 - 2(m-2) = \frac{1}{2}(m-2)(m-3)$ in equation (V.80). These considerations finally lead to

$$\max_{i \in \{1,\dots,M\}} \sum_{j=1}^{M} \operatorname{Var}\bigl[(S_n^K)_{i,j}\bigr] = 2\operatorname{Var}[C_{1,1}] + 2(m-2)\operatorname{Var}[C_{1,2}]. \tag{V.81}$$

Next, we calculate both variances occurring on the right-hand side above equation (V.81). Before we start, let us note that we can interchange taking the limit and the expectation as well as the order of summation because the occurring series converge absolutely with respect to the $L^p(\Omega; \mathbb{R})$-norm for every $p \in [1, \infty[$. For the first variance on the right-hand side of equation (V.81), we have

$$\begin{aligned}
\operatorname{Var}[C_{1,1}] &= \frac{h_n^2}{4} \operatorname{Var}\Biggl[\sum_{k=K+1}^{\infty} (b^1_{k,n})^2 - \Biggl(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\Biggr)^{-1} \Biggl(\sum_{k=K+1}^{\infty} \frac{b^1_{k,n}}{k}\Biggr)^2\Biggr] \\
&= \frac{h_n^2}{4} \Biggl(\operatorname{Var}\Biggl[\sum_{k=K+1}^{\infty} (b^1_{k,n})^2\Biggr] + \Biggl(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\Biggr)^{-2} \operatorname{Var}\Biggl[\Biggl(\sum_{k=K+1}^{\infty} \frac{b^1_{k,n}}{k}\Biggr)^2\Biggr] \\
&\qquad - 2\Biggl(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\Biggr)^{-1} \operatorname{Cov}\Biggl[\sum_{k=K+1}^{\infty} (b^1_{k,n})^2, \Biggl(\sum_{k=K+1}^{\infty} \frac{b^1_{k,n}}{k}\Biggr)^2\Biggr]\Biggr),
\end{aligned}$$

where

$$\mathrm{Var}\Big[\sum_{k=K+1}^{\infty}(b_{k,n}^1)^2\Big] = \sum_{k=K+1}^{\infty}\mathrm{Var}\big[(b_{k,n}^1)^2\big] = 2\sum_{k=K+1}^{\infty}\frac{h_n^2}{4\pi^4k^4} = \frac{h_n^2}{2\pi^4}\sum_{k=K+1}^{\infty}\frac{1}{k^4}$$

and

$$\mathrm{Var}\Big[\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\Big)^2\Big] = 2\Big(\sum_{k=K+1}^{\infty}\frac{h_n}{2\pi^2k^4}\Big)^2 = \frac{h_n^2}{2\pi^4}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big)^2.$$

Further, considering the covariance and using that

$$\begin{aligned}
\mathrm{E}\Big[\sum_{k,l,r=K+1}^{\infty}(b_{k,n}^1)^2\frac{b_{l,n}^1}{l}\frac{b_{r,n}^1}{r}\Big] &= \mathrm{E}\Big[\sum_{k,l=K+1}^{\infty}(b_{k,n}^1)^2\frac{(b_{l,n}^1)^2}{l^2}\Big] \\
&= \mathrm{E}\Big[\sum_{k=K+1}^{\infty}\frac{(b_{l,n}^1)^4}{k^2} + \sum_{\substack{k,l=K+1\\k\neq l}}^{\infty}(b_{k,n}^1)^2\frac{(b_{l,n}^1)^2}{l^2}\Big] \\
&= 3\sum_{k=K+1}^{\infty}\frac{h_n^2}{4\pi^4k^6} + \sum_{\substack{k,l=K+1\\k\neq l}}^{\infty}\frac{h_n^2}{4\pi^4k^2l^4} \\
&= 2\sum_{k=K+1}^{\infty}\frac{h_n^2}{4\pi^4k^6} + \sum_{k,l=K+1}^{\infty}\frac{h_n^2}{4\pi^4k^2l^4} \\
&= \frac{h_n^2}{2\pi^4}\sum_{k=K+1}^{\infty}\frac{1}{k^6} + \frac{h_n^2}{2\pi^4}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big),
\end{aligned}$$

it holds

$$\begin{aligned}
&\mathrm{Cov}\Big[\sum_{k=K+1}^{\infty}(b_{k,n}^1)^2, \Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\Big)^2\Big] \\
&= \mathrm{E}\Big[\Big(\sum_{k=K+1}^{\infty}(b_{k,n}^1)^2 - \sum_{k=K+1}^{\infty}\frac{h_n}{2\pi^2k^2}\Big)\Big(\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\Big)^2 - \sum_{k=K+1}^{\infty}\frac{h_n}{2\pi^2k^4}\Big)\Big] \\
&= \mathrm{E}\Big[\Big(\sum_{k=K+1}^{\infty}(b_{k,n}^1)^2\Big)\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\Big)^2\Big] - \Big(\sum_{k=K+1}^{\infty}\frac{h_n}{2\pi^2k^2}\Big)\Big(\sum_{k=K+1}^{\infty}\frac{h_n}{2\pi^2k^4}\Big) \\
&= \mathrm{E}\Big[\sum_{k,l,r=K+1}^{\infty}(b_{k,n}^1)^2\frac{b_{l,n}^1}{l}\frac{b_{r,n}^1}{r}\Big] - \frac{h_n^2}{2\pi^4}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big) \\
&= \frac{h_n^2}{2\pi^4}\sum_{k=K+1}^{\infty}\frac{1}{k^6}.
\end{aligned}$$

In summary, we thus have

$$\begin{aligned}
&\mathrm{Var}[C_{1,1}] \\
&= \frac{h_n^4}{8\pi^4}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4} + \Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-2}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big)^2 - 2\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^6}\Big).
\end{aligned} \tag{V.82}$$

It analogously holds for the second variance on the right-hand side of equation (V.81) that

$$\begin{aligned}
\mathrm{Var}[C_{1,2}] &= \frac{h_n^2}{4}\mathrm{Var}\Big[\sum_{k=K+1}^{\infty} b_{k,n}^1 b_{k,n}^2 - \Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\Big)\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^2}{k}\Big)\Big] \\
&= \frac{h_n^2}{4}\Big(\mathrm{Var}\Big[\sum_{k=K+1}^{\infty} b_{k,n}^1 b_{k,n}^2\Big] \\
&\quad + \Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-2}\mathrm{Var}\Big[\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\Big)\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^2}{k}\Big)\Big] \\
&\quad - 2\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\mathrm{Cov}\Big[\sum_{k=K+1}^{\infty} b_{k,n}^1 b_{k,n}^2, \Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\Big)\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^2}{k}\Big)\Big]\Big) \\
&= \frac{h_n^2}{4}\Big(\sum_{k=K+1}^{\infty}\frac{h_n^2}{4\pi^4 k^4} + \Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-2}\Big(\sum_{k=K+1}^{\infty}\frac{h_n}{2\pi^2 k^4}\Big)^2 \\
&\quad - 2\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\mathrm{E}\Big[\Big(\sum_{k=K+1}^{\infty} b_{k,n}^1 b_{k,n}^2\Big)\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\Big)\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^2}{k}\Big)\Big]\Big).
\end{aligned}$$

Since

$$\begin{aligned}
\mathrm{E}\Big[\Big(\sum_{k=K+1}^{\infty} b_{k,n}^1 b_{k,n}^2\Big)\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\Big)\Big(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^2}{k}\Big)\Big] &= \mathrm{E}\Big[\sum_{k,l,r=K+1}^{\infty} b_{k,n}^1 b_{k,n}^2 \frac{b_{l,n}^1}{l}\frac{b_{r,n}^2}{r}\Big] \\
&= \mathrm{E}\Big[\sum_{k=K+1}^{\infty}\frac{1}{k^2}(b_{k,n}^1)^2(b_{k,n}^2)^2\Big] \\
&= \sum_{k=K+1}^{\infty}\frac{h_n^2}{4\pi^4 k^6},
\end{aligned}$$

we obtain

$$\begin{aligned}
&\mathrm{Var}[C_{1,2}] \\
&\quad = \frac{h_n^4}{16\pi^4}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4} + \Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-2}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big)^2 - 2\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^6}\Big).
\end{aligned} \tag{V.83}$$

Comparing the equations (V.82) and (V.83), we have $\mathrm{Var}[C_{1,1}] = 2\mathrm{Var}[C_{1,2}]$. Thus, inserting the results from equations (V.82) and (V.83) into equation (V.81) yields

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}}\sum_{j=1}^{M}\mathrm{Var}\big[(S_n^K)_{i,j}\big] \\
&\quad = 2m\mathrm{Var}[C_{1,2}] \\
&\quad = 2m\Big(\frac{h_n}{2\pi}\Big)^4\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4} + \Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-2}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big)^2 - 2\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^6}\Big).
\end{aligned} \tag{V.84}$$

Now, we take a closer look at the series in the formula above. Using Lemma V.22 and Lemma V.23, it follows for $K \in \mathbb{N}$ that

$$\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-2}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big)^{2} \leq \frac{\big(K+\frac{3}{4}\big)^2}{9\big(K+\frac{1}{2}\big)^6}$$

and

$$\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^6} \geq \frac{K+\frac{1}{2}}{5\big(K+\frac{3}{4}\big)^5}.$$

Considering the difference on the right-hand side in equation (V.84), we have with both previous inequalities that

$$\begin{aligned}&\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-2}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big)^{2} - 2\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^6}\\ &\quad\leq \frac{5\big(K+\frac{3}{4}\big)^7 - 18\big(K+\frac{1}{2}\big)^7}{45\big(K+\frac{1}{2}\big)^6\big(K+\frac{3}{4}\big)^5}.\end{aligned} \tag{V.85}$$

We continue showing that the right-hand side of inequality (V.85) above is less than zero. For this, we consider the numerator only. The Taylor expansion implies

$$5\Big(K+\frac{3}{4}\Big)^7 - 18\Big(K+\frac{1}{2}\Big)^7 = 5\sum_{i=1}^{7}\frac{7!}{(7-i)!i!4^i}\Big(K+\frac{1}{2}\Big)^{7-i} - 13\Big(K+\frac{1}{2}\Big)^7$$

for all $K \in \mathbb{N}$. Since $\big(K+\frac{1}{2}\big)^{-i}$ is monotonically decreasing in K for $i \in \{1,\dots,7\}$ and

$$5\sum_{i=1}^{7}\frac{7!}{(7-i)!i!4^i}\Big(1+\frac{1}{2}\Big)^{-i} - 13 = -\frac{921133}{279936} \leq 0,$$

we obtain

$$\frac{5\big(K+\frac{3}{4}\big)^7 - 18\big(K+\frac{1}{2}\big)^7}{\big(K+\frac{1}{2}\big)^7} = 5\sum_{i=1}^{7}\frac{7!}{(7-i)!i!4^i}\Big(K+\frac{1}{2}\Big)^{-i} - 13 \leq 0$$

for all $K \in \mathbb{N}$. Thus, we infer from inequality (V.85) that

$$\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-2}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Big)^{2} - 2\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^6} \leq 0,$$

and we estimate equation (V.84) to

$$\max_{i\in\{1,\dots,M\}}\sum_{j=1}^{M}\mathrm{Var}\big[(S_n^K)_{i,j}\big] \leq 2m\Big(\frac{h_n}{2\pi}\Big)^4\sum_{k=K+1}^{\infty}\frac{1}{k^4}. \tag{V.86}$$

Combining equation (V.72), inequality (V.73), equation (V.74), and above inequality (V.86), we obtain

$$\begin{aligned}&\max_{i\in\{1,\dots,M\}} \Big\| e_i^{\mathrm{T}} \Big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\Big) G_{1,n} \Big\|_{L^2(\Omega;\mathbb{R})}^2 \\ &\quad \leq m \Big(\frac{h_n}{2\pi}\Big)^2 \bigg(\Big(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\Big) - \Big(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\Big)^{-1} \Big(\sum_{k=K+1}^{\infty} \frac{1}{k^4}\Big)\bigg)^{-1} \sum_{k=K+1}^{\infty} \frac{1}{k^4}.\end{aligned}$$

Using Lemma V.22 and Lemma V.23, it holds

$$\begin{aligned}&\bigg(\Big(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\Big) - \Big(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\Big)^{-1} \Big(\sum_{k=K+1}^{\infty} \frac{1}{k^4}\Big)\bigg)^{-1} \sum_{k=K+1}^{\infty} \frac{1}{k^4} \\ &\quad \leq \bigg(\frac{1}{K+\frac{3}{4}} - \frac{K+\frac{3}{4}}{3\big(K+\frac{1}{2}\big)^3} \bigg)^{-1} \frac{1}{3\big(K+\frac{1}{2}\big)^3} \\ &\quad = \frac{3K+\frac{9}{4}}{3K+\frac{7}{2}+\frac{3}{4K}-\frac{3}{16K^2}} \cdot \frac{1}{3K^2} \\ &\quad \leq \frac{1}{3K^2} \end{aligned} \tag{V.87}$$

for $K \in \mathbb{N}$, and finally, we obtain

$$\max_{i\in\{1,\dots,M\}} \big\| e_i^{\mathrm{T}} \big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big) G_{1,n} \big\|_{L^2(\Omega;\mathbb{R})}^2 \leq \Big(\frac{h_n}{2\pi}\Big)^2 \frac{m}{3K^2},$$

which completes the proof. □

Proof of Theorem V.9

Proof of Theorem V.9. Since $I_{(j,j),n} = I_{(j,j),n}^{K+}$ for $j \in \{1,\dots,m\}$, we have

$$\max_{j\in\{1,\dots,m\}} \| I_{(j,j),n} - I_{(j,j),n}^{K+} \|_{L^p(\Omega;\mathbb{R})} = 0,$$

and it holds

$$\max_{i,j\in\{1,\dots,m\}} \| I_{(i,j),n} - I_{(i,j),n}^{K+} \|_{L^p(\Omega;\mathbb{R})} = \max_{i\in\{1,\dots,M\}} \big\| e_i^{\mathrm{T}} \big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big) G_{1,n} \big\|_{L^p(\Omega;\mathbb{R})}, \tag{V.88}$$

cf. equation (V.42). Similar considerations as in the proof of Lemma V.5 show that, given $\{b_{k,n} : k \in \mathbb{N} \text{ with } k > K\}$, the conditional distribution of

$$e_i^{\mathrm{T}} \big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big) G_{1,n} = \sum_{j=1}^{M} \big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big)_{i,j} G_{1,n}^j$$

is normal with conditional expectation zero P-almost surely and conditional variance

$$\sum_{j=1}^{M} \Big(\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big)_{i,j} \Big)^2 = \big\| e_i^{\mathrm{T}} \big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big) \big\|_{\mathrm{F}}^2$$

P-almost surely. Since $\mathrm{N}(0_{M\times 1}, I_M)$-distributed random variable $G_{1,n}$ is independent of Fourier coefficients $b_{k,n}$, $k \in \mathbb{N}$, we obtain by using [16, Corollary 4.38] or [43, Satz 5.3.22], and by using Lemma V.20 that

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \big\| e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big) G_{1,n}\big\|^2_{L^p(\Omega;\mathbb{R})} \\
&= \max_{i\in\{1,\dots,M\}} \Big(\mathrm{E}\Big[\mathrm{E}\big[\big|e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big) G_{1,n}\big|^p \big| b_{k,n}, k\in\mathbb{N}\big]\Big]\Big)^{\frac{2}{p}} \\
&= \frac{2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \max_{i\in\{1,\dots,M\}} \big(\mathrm{E}\big[\big\|e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big)\big\|_{\mathrm{F}}^p\big]\big)^{\frac{2}{p}}
\end{aligned} \tag{V.89}$$

cf. equation (V.72). Then, similarly to inequality (V.73), it follows by Lemma V.21 that

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \big\| e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big) G_{1,n}\big\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \Big(\frac{h_n}{2\pi}\Big)^{-2} \Bigg(\Big(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\Big) - \Big(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\Big)^{-1} \Big(\sum_{k=K+1}^{\infty} \frac{1}{k^4}\Big)\Bigg)^{-1} \\
&\quad\times \max_{i\in\{1,\dots,M\}} \big(\mathrm{E}\big[\big\|e_i^{\mathrm{T}}\big(S_n^K - \mathrm{E}[S_n^K]\big)\big\|_{\mathrm{F}}^p\big]\big)^{\frac{2}{p}}.
\end{aligned} \tag{V.90}$$

Considering the last factor on the right-hand side of above inequality (V.90) and applying the triangle inequality, it holds

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \big(\mathrm{E}\big[\big\|e_i^{\mathrm{T}}\big(S_n^K - \mathrm{E}[S_n^K]\big)\big\|_{\mathrm{F}}^p\big]\big)^{\frac{2}{p}} \\
&= \max_{i\in\{1,\dots,M\}} \Bigg\| \sum_{j=1}^{M} \big((S_n^K)_{i,j} - (\mathrm{E}[S_n^K])_{i,j}\big)^2 \Bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \\
&\leq \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{M} \big\|(S_n^K)_{i,j} - (\mathrm{E}[S_n^K])_{i,j}\big\|^2_{L^p(\Omega;\mathbb{R})}.
\end{aligned} \tag{V.91}$$

Since $\mathrm{E}[S_n^K]$ is a diagonal matrix by Lemma V.6, we further have

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{M} \big\|(S_n^K)_{i,j} - (\mathrm{E}[S_n^K])_{i,j}\big\|^2_{L^p(\Omega;\mathbb{R})} \\
&= \max_{i\in\{1,\dots,M\}} \Bigg(\big\|(S_n^K)_{i,i} - (\mathrm{E}[S_n^K])_{i,i}\big\|^2_{L^p(\Omega;\mathbb{R})} + \sum_{\substack{j=1\\ j\neq i}}^{M} \big\|(S_n^K)_{i,j}\big\|^2_{L^p(\Omega;\mathbb{R})}\Bigg).
\end{aligned} \tag{V.92}$$

In the following, we use the same notations as in the proof of Theorem V.7. Similar considerations that lead to inequality (V.81), imply together with equation (V.92) that

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{M} \big\|(S_n^K)_{i,j} - (\mathrm{E}[S_n^K])_{i,j}\big\|^2_{L^p(\Omega;\mathbb{R})} \\
&= \max_{\substack{i,j\in\{1,\dots,m\}\\ i<j}} \|C_{i,i} - \mathrm{E}[C_{i,i}] + C_{j,j} - \mathrm{E}[C_{j,j}]\|^2_{L^p(\Omega;\mathbb{R})} + 2(m-2)\|C_{i,j}\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq 4\|C_{1,1} - \mathrm{E}[C_{1,1}]\|^2_{L^p(\Omega;\mathbb{R})} + 2(m-2)\|C_{1,2}\|^2_{L^p(\Omega;\mathbb{R})}
\end{aligned} \tag{V.93}$$

because

$$\begin{aligned}
&\|C_{i,i} - \mathrm{E}[C_{i,i}] + C_{j,j} - \mathrm{E}[C_{j,j}]\|^2_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq \big(\|C_{i,i} - \mathrm{E}[C_{i,i}]\|_{L^p(\Omega;\mathbb{R})} + \|C_{j,j} - \mathrm{E}[C_{j,j}]\|_{L^p(\Omega;\mathbb{R})} \big)^2 \\
&\quad = \big(2\|C_{1,1} - \mathrm{E}[C_{1,1}]\|_{L^p(\Omega;\mathbb{R})} \big)^2.
\end{aligned}$$

Now, we proceed with the estimation of the $L^p(\Omega;\mathbb{R})$-norms in inequality (V.93). Considering the first $L^p(\Omega;\mathbb{R})$-norm on the right-hand side of inequality (V.93), we obtain by equation (V.75) and by the proof of Lemma V.6 that

$$\begin{aligned}
&C_{1,1} - \mathrm{E}[C_{1,1}] \\
&\quad = \frac{h_n}{2} \bigg(\sum_{k=K+1}^{\infty} \big(b^1_{k,n}\big)^2 - \frac{h_n}{2\pi^2 k^2} - \bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2} \bigg)^{-1} \bigg(\bigg(\sum_{k=K+1}^{\infty} \frac{b^1_{k,n}}{k} \bigg)^2 - \sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^4} \bigg) \bigg).
\end{aligned}$$

Further, using the triangle inequality and Lemma V.20, we have

$$\begin{aligned}
&\|C_{1,1} - \mathrm{E}[C_{1,1}]\|^2_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq \frac{h_n^2}{4} \bigg(\bigg\| \sum_{k=K+1}^{\infty} \big(b^1_{k,n}\big)^2 - \frac{h_n}{2\pi^2 k^2} \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\qquad + \bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2} \bigg)^{-1} \bigg(\bigg\| \sum_{k=K+1}^{\infty} \frac{b^1_{k,n}}{k} \bigg\|^2_{L^{2p}(\Omega;\mathbb{R})} + \sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^4} \bigg) \bigg)^2 \\
&\quad = \frac{h_n^2}{4} \bigg(\bigg\| \sum_{k=K+1}^{\infty} \big(b^1_{k,n}\big)^2 - \frac{h_n}{2\pi^2 k^2} \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\qquad + \bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2} \bigg)^{-1} \bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} + 1 \bigg) \sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^4} \bigg)^2.
\end{aligned} \tag{V.94}$$

Consider the $L^p(\Omega;\mathbb{R})$-norm on the right-hand side of inequality (V.94) above. Since the series is convergent in $L^p(\Omega;\mathbb{R})$ and a discrete martingale due to the independence of $b^1_{k,n}$, $k \in \mathbb{N}$, Theorem II.5 implies

$$\begin{aligned}
\bigg\| \sum_{k=K+1}^{\infty} \big(b^1_{k,n}\big)^2 - \frac{h_n}{2\pi^2 k^2} \bigg\|_{L^p(\Omega;\mathbb{R})} &= \lim_{N\to\infty} \bigg\| \sum_{k=K+1}^{N} \big(b^1_{k,n}\big)^2 - \frac{h_n}{2\pi^2 k^2} \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\leq \sqrt{p-1} \lim_{N\to\infty} \bigg(\sum_{k=K+1}^{N} \Big\| \big(b^1_{k,n}\big)^2 - \frac{h_n}{2\pi^2 k^2} \Big\|^2_{L^p(\Omega;\mathbb{R})} \bigg)^{\frac{1}{2}} \\
&\leq \sqrt{p-1} \bigg(\sum_{k=K+1}^{\infty} \Big\| \big(b^1_{k,n}\big)^2 - \frac{h_n}{2\pi^2 k^2} \Big\|^2_{L^p(\Omega;\mathbb{R})} \bigg)^{\frac{1}{2}}.
\end{aligned}$$

Then, the application of the triangle inequality and Lemma V.20 leads to

$$\begin{aligned}\Big\| \sum_{k=K+1}^{\infty} \big(b_{k,n}^1\big)^2 - \frac{h_n}{2\pi^2 k^2}\Big\|_{L^p(\Omega;\mathbb{R})} &\leq \sqrt{p-1}\bigg(\sum_{k=K+1}^{\infty} \Big(\|b_{k,n}^1\|_{L^{2p}(\Omega;\mathbb{R})}^2 + \frac{h_n}{2\pi^2 k^2}\Big)^2\bigg)^{\frac{1}{2}} \\ &= \sqrt{p-1}\bigg(\sum_{k=K+1}^{\infty} \bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} \frac{h_n}{2\pi^2 k^2} + \frac{h_n}{2\pi^2 k^2}\bigg)^2\bigg)^{\frac{1}{2}} \\ &= \sqrt{p-1}\bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} + 1\bigg) \frac{h_n}{2\pi^2}\bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^4}\bigg)^{\frac{1}{2}}.\end{aligned}$$

Inserting this into inequality (V.94), we obtain

$$\begin{aligned}&\|C_{1,1} - \mathrm{E}[C_{1,1}]\|_{L^p(\Omega;\mathbb{R})}^2 \\ &\quad\leq \Big(\frac{h_n}{2\pi}\Big)^4 \bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} + 1\bigg)^2 \bigg(\sqrt{p-1}\bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^4}\bigg)^{\frac{1}{2}} + \bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\bigg)^{-1} \sum_{k=K+1}^{\infty} \frac{1}{k^4}\bigg)^2.\end{aligned} \tag{V.95}$$

We use similar considerations in order to estimate the second $L^p(\Omega;\mathbb{R})$-norm on the right-hand side of inequality (V.93). Using the triangle inequality, it first holds

$$\begin{aligned}&\|C_{1,2}\|_{L^p(\Omega;\mathbb{R})}^2 \\ &\quad= \frac{h_n^2}{4}\bigg\| \sum_{k=K+1}^{\infty} b_{k,n}^1 b_{k,n}^2 - \bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\bigg)^{-1} \bigg(\sum_{k=K+1}^{\infty} \frac{b_{k,n}^1}{k}\bigg)\bigg(\sum_{k=K+1}^{\infty} \frac{b_{k,n}^2}{k}\bigg)\bigg\|_{L^p(\Omega;\mathbb{R})}^2 \\ &\quad\leq \frac{h_n^2}{4}\bigg(\Big\| \sum_{k=K+1}^{\infty} b_{k,n}^1 b_{k,n}^2 \Big\|_{L^p(\Omega;\mathbb{R})} \\ &\qquad + \bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\bigg)^{-1} \bigg\| \bigg(\sum_{k=K+1}^{\infty} \frac{b_{k,n}^1}{k}\bigg)\bigg(\sum_{k=K+1}^{\infty} \frac{b_{k,n}^2}{k}\bigg)\bigg\|_{L^p(\Omega;\mathbb{R})}\bigg)^2.\end{aligned}$$

Then, the independence of $b_{k,n}^1$ and $b_{k,n}^2$ for $k \in \mathbb{N}$, Theorem II.5, and Lemma V.20 imply

$$\begin{aligned}&\|C_{1,2}\|_{L^p(\Omega;\mathbb{R})}^2 \\ &\quad\leq \frac{h_n^2}{4}\bigg(\sqrt{p-1}\bigg(\sum_{k=K+1}^{\infty} \|b_{k,n}^1\|_{L^p(\Omega;\mathbb{R})}^2 \|b_{k,n}^2\|_{L^p(\Omega;\mathbb{R})}^2\bigg)^{\frac{1}{2}} \\ &\qquad + \bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\bigg)^{-1} \bigg\| \sum_{k=K+1}^{\infty} \frac{b_{k,n}^1}{k}\bigg\|_{L^p(\Omega;\mathbb{R})} \bigg\| \sum_{k=K+1}^{\infty} \frac{b_{k,n}^2}{k}\bigg\|_{L^p(\Omega;\mathbb{R})}\bigg)^2 \\ &\quad= \frac{h_n^2}{4}\bigg(\sqrt{p-1}\bigg(\sum_{k=K+1}^{\infty} \frac{4\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}} \frac{h_n^2}{4\pi^4 k^4}\bigg)^{\frac{1}{2}} \\ &\qquad + \bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\bigg)^{-1} \frac{2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \sum_{k=K+1}^{\infty} \frac{h_n}{2\pi^2 k^4}\bigg)^2 \\ &\quad= \Big(\frac{h_n}{2\pi}\Big)^4 \frac{4\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}} \bigg(\sqrt{p-1}\bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^4}\bigg)^{\frac{1}{2}} + \bigg(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\bigg)^{-1} \sum_{k=K+1}^{\infty} \frac{1}{k^4}\bigg)^2.\end{aligned} \tag{V.96}$$

Inserting inequalities (V.95) and (V.96) into inequality (V.93), we finally obtain

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{M} \big\|(S_n^K)_{i,j} - (\mathrm{E}[S_n^K])_{i,j}\big\|^2_{L^p(\Omega;\mathbb{R})} \\
&\le \Big(\frac{h_n}{2\pi}\Big)^4 \Bigg(4\bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} + 1\bigg)^2 + 2(m-2)\frac{4\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\Bigg) \\
&\quad \times \Bigg(\sqrt{p-1}\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\bigg)^{\frac{1}{2}} + \bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Bigg)^2,
\end{aligned}$$

and together with the inequalities (V.91) and (V.90), we in total have

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \Big\| e_i^{\mathrm{T}}\big(\sqrt{S_n^K} - \sqrt{\mathrm{E}[S_n^K]}\big)G_{1,n}\Big\|^2_{L^p(\Omega;\mathbb{R})} \\
&\le \Big(\frac{h_n}{2\pi}\Big)^2 \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \Bigg(4\bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} + 1\bigg)^2 + 2(m-2)\frac{4\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\Bigg) \\
&\quad \times \Bigg(\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg) - \bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)^{-1}\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\bigg)\Bigg)^{-1} \\
&\quad \times \Bigg(\sqrt{p-1}\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\bigg)^{\frac{1}{2}} + \bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Bigg)^2.
\end{aligned} \tag{V.97}$$

Next, we estimate the series on the right-hand side of above inequality (V.97) using Lemma V.22 and Lemma V.23. It holds for $K \in \mathbb{N}$ that

$$\begin{aligned}
&\Bigg(\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg) - \bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)^{-1}\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\bigg)\Bigg)^{-1} \\
&\quad \times \Bigg(\sqrt{p-1}\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\bigg)^{\frac{1}{2}} + \bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Bigg)^2 \\
&\quad \le \bigg(\frac{1}{K+\frac{3}{4}} - \frac{K+\frac{3}{4}}{3\big(K+\frac{1}{2}\big)^3}\bigg)^{-1}\bigg(\frac{\sqrt{p-1}}{\sqrt{3}\big(K+\frac{1}{2}\big)^{\frac{3}{2}}} + \frac{K+\frac{3}{4}}{3\big(K+\frac{1}{2}\big)^3}\bigg)^2 \\
&\quad = \bigg(\frac{1}{K+\frac{3}{4}} - \frac{K+\frac{3}{4}}{3\big(K+\frac{1}{2}\big)^3}\bigg)^{-1}\bigg(\frac{p-1}{3\big(K+\frac{1}{2}\big)^3} + \frac{2\sqrt{p-1}\big(K+\frac{3}{4}\big)}{3\sqrt{3}\big(K+\frac{1}{2}\big)^{\frac{9}{2}}} + \frac{\big(K+\frac{3}{4}\big)^2}{9\big(K+\frac{1}{2}\big)^6}\bigg).
\end{aligned}$$

Since

$$\frac{K+\frac{3}{4}}{\big(K+\frac{1}{2}\big)^{\frac{3}{2}}} \le 1$$

and thus also

$$\frac{\big(K+\frac{3}{4}\big)^2}{\big(K+\frac{1}{2}\big)^3} \le 1$$

for all $K \in \mathbb{N}$, we obtain by inequality (V.87) that

$$\begin{aligned}
&\Bigg(\Bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Bigg)-\Bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Bigg)^{-1}\Bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Bigg)\Bigg)^{-1}\\
&\quad\times\Bigg(\sqrt{p-1}\Bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Bigg)^{\frac{1}{2}}+\Bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Bigg)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^4}\Bigg)^2\\
&\leq\Bigg(\frac{1}{K+\frac{3}{4}}-\frac{K+\frac{3}{4}}{3\big(K+\frac{1}{2}\big)^3}\Bigg)^{-1}\frac{1}{3\big(K+\frac{1}{2}\big)^3}\Bigg(p-1+\frac{2\sqrt{p-1}}{\sqrt{3}}+\frac{1}{3}\Bigg)\\
&\leq\frac{1}{3K^2}\Bigg(\sqrt{p-1}+\frac{1}{\sqrt{3}}\Bigg)^2\\
&=\frac{\big(\sqrt{3}\sqrt{p-1}+1\big)^2}{9K^2}.
\end{aligned}$$

Inserting this into inequality (V.97), it holds together with equation (V.88) that

$$\begin{aligned}
&\max_{i,j\in\{1,\ldots,m\}}\|I_{(i,j),n}-I_{(i,j),n}^{K+}\|_{L^p(\Omega;\mathbb{R})}^2\\
&\leq\Big(\frac{h_n}{2\pi}\Big)^2\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}}\Bigg(4\Bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}+1\Bigg)^2+2(m-2)\frac{4\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\Bigg)\frac{\big(\sqrt{3}\sqrt{p-1}+1\big)^2}{9K^2},
\end{aligned}$$

which completes the proof. □

Proof of Theorem V.11

The proof follows the same ideas that are used in the proof of Theorem V.7.

Proof of Theorem V.11. At first, we have

$$\begin{aligned}
&\max_{i\in\{1,\ldots,M\}}\big\|e_i^{\mathrm{T}}\Sigma_{2,n}^K(\Sigma_{1,n}^K)^{-1}\sqrt{\Sigma_{1,n}^K}G_{0,n}\big\|_{L^2(\Omega;\mathbb{R})}^2\\
&=\max_{i\in\{1,\ldots,M\}}\mathrm{E}\Big[\mathrm{E}\big[\big|e_i^{\mathrm{T}}\Sigma_{2,n}^K(\Sigma_{1,n}^K)^{-1}\sqrt{\Sigma_{1,n}^K}G_{0,n}\big|^2\big|b_{k,n},k\in\mathbb{N}\big]\Big]\\
&=\max_{i\in\{1,\ldots,M\}}\mathrm{E}\big[\big\|e_i^{\mathrm{T}}\Sigma_{2,n}^K(\Sigma_{1,n}^K)^{-1}\sqrt{\Sigma_{1,n}^K}\big\|_{\mathrm{F}}^2\big]\\
&=\Bigg(\frac{h_n}{2\pi^2}\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Bigg)^{-1}\max_{i\in\{1,\ldots,M\}}\mathrm{E}\big[\|e_i^{\mathrm{T}}\Sigma_{2,n}^K\|_{\mathrm{F}}^2\big],
\end{aligned}\tag{V.98}$$

where

$$(\Sigma_{1,n}^K)^{-1}\sqrt{\Sigma_{1,n}^K}=\Bigg(\frac{h_n}{2\pi^2}\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Bigg)^{-\frac{1}{2}}I_m$$

is used. We now take a closer look at covariance matrix $\Sigma_{2,n}^K$. Using equations (V.35) and (V.78), it holds

$$\Sigma_{2,n}^K=\frac{h_n}{2\pi}\sum_{\substack{i,j=1\\i<j}}^{m}H_m(e_i\otimes e_j)\Bigg(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^j}{k}e_i^{\mathrm{T}}-\sum_{k=K+1}^{\infty}\frac{b_{k,n}^i}{k}e_j^{\mathrm{T}}\Bigg).$$

Taking into account that $\big\{H_m(e_i \otimes e_j) : i,j \in \{1,\dots,m\} \text{ with } i<j\big\}$ is the canonical basis of $\mathbb{R}^M$ and that $\mathrm{E}[\varSigma_{2,n}^K] = 0_{M\times m}$, we obtain

$$\begin{aligned}
\max_{i\in\{1,\dots,M\}} \mathrm{E}\big[\|e_i^{\mathrm{T}}\varSigma_{2,n}^K\|_{\mathrm{F}}^2\big] &= \max_{i\in\{1,\dots,M\}} \sum_{l=1}^m \mathrm{E}\big[|(\varSigma_{2,n}^K)_{i,l}|^2\big] \\
&= \max_{i\in\{1,\dots,M\}} \sum_{l=1}^m \mathrm{Var}\big[(\varSigma_{2,n}^K)_{i,l}\big] \\
&= \max_{\substack{i,j\in\{1,\dots,m\}\\ i<j}} \sum_{l=1}^m \mathrm{Var}\bigg[\frac{h_n}{2\pi}\bigg(\sum_{k=K+1}^{\infty}\frac{b_{k,n}^j}{k}e_i^l - \sum_{k=K+1}^{\infty}\frac{b_{k,n}^i}{k}e_j^l\bigg)\bigg] \\
&= \max_{\substack{i,j\in\{1,\dots,m\}\\ i<j}} \mathrm{Var}\bigg[\frac{h_n}{2\pi}\sum_{k=K+1}^{\infty}\frac{b_{k,n}^j}{k}\bigg] + \mathrm{Var}\bigg[\frac{h_n}{2\pi}\sum_{k=K+1}^{\infty}\frac{b_{k,n}^i}{k}\bigg] \\
&= 2\mathrm{Var}\bigg[\frac{h_n}{2\pi}\sum_{k=K+1}^{\infty}\frac{b_{k,n}^1}{k}\bigg] \\
&= 2\Big(\frac{h_n}{2\pi}\Big)^2 \sum_{k=K+1}^{\infty}\frac{h_n}{2\pi^2k^4}. \qquad \text{(V.99)}
\end{aligned}$$

Inserting this into equation (V.98), it holds

$$\max_{i\in\{1,\dots,M\}} \big\|e_i^{\mathrm{T}}\varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}\sqrt{\varSigma_{1,n}^K}G_{0,n}\big\|_{L^2(\Omega;\mathbb{R})}^2 = 2\Big(\frac{h_n}{2\pi}\Big)^2\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^4},$$

where

$$\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^4} \le \frac{K+\frac{3}{4}}{3\big(K+\frac{1}{2}\big)^3} = \frac{1}{3K^2}\frac{K+\frac{3}{4}}{K+\frac{3}{2}+\frac{3}{4K}+\frac{1}{4K^2}} \le \frac{1}{3K^2} \qquad \text{(V.100)}$$

by Lemma V.22 and Lemma V.23. Thus, we have

$$\max_{i\in\{1,\dots,M\}} \big\|e_i^{\mathrm{T}}\varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}\sqrt{\varSigma_{1,n}^K}G_{0,n}\big\|_{L^2(\Omega;\mathbb{R})}^2 \le \Big(\frac{h_n}{2\pi}\Big)^2\frac{2}{3K^2}. \qquad \text{(V.101)}$$

Now, we consider the error of approximation $\tilde{I}_{(i,j),n}^{K+}$. According to equations (V.41) and (V.44), it holds

$$\mathrm{vec}\big[(\tilde{I}_n^{K+})^{\mathrm{T}}\big] = \mathrm{vec}\big[(I_n^{K+})^{\mathrm{T}}\big] - \varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}\sqrt{\varSigma_{1,n}^K}G_{0,n},$$

and hence, we obtain

$$\begin{aligned}
&\max_{i,j\in\{1,\dots,m\}} \|I_{(i,j),n} - \tilde{I}_{(i,j),n}^{K+}\|_{L^2(\Omega;\mathbb{R})} \\
&= \max_{i\in\{1,\dots,M\}} \big\|e_i^{\mathrm{T}}\big(H_m\mathrm{vec}\big[(I_n)^{\mathrm{T}}\big] - H_m\mathrm{vec}\big[(\tilde{I}_n^{K+})^{\mathrm{T}}\big]\big)\big\|_{L^2(\Omega;\mathbb{R})} \\
&= \max_{i\in\{1,\dots,M\}} \big\|e_i^{\mathrm{T}}\big(H_m\mathrm{vec}\big[(I_n)^{\mathrm{T}}\big] - H_m\mathrm{vec}\big[(I_n^{K+})^{\mathrm{T}}\big] + \varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}\sqrt{\varSigma_{1,n}^K}G_{0,n}\big)\big\|_{L^2(\Omega;\mathbb{R})} \\
&\le \max_{i\in\{1,\dots,M\}} \big\|e_i^{\mathrm{T}}\big(H_m\mathrm{vec}\big[(I_n)^{\mathrm{T}}\big] - H_m\mathrm{vec}\big[(I_n^{K+})^{\mathrm{T}}\big]\big)\big\|_{L^2(\Omega;\mathbb{R})} \\
&\quad + \max_{i\in\{1,\dots,M\}} \big\|e_i^{\mathrm{T}}\big(\varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}\sqrt{\varSigma_{1,n}^K}G_{0,n}\big)\big\|_{L^2(\Omega;\mathbb{R})}. \qquad \text{(V.102)}
\end{aligned}$$

Then, Theorem V.8 and inequality (V.101) imply

$$\max_{i,j\in\{1,\dots,m\}} \|I_{(i,j),n} - \tilde{I}^{K+}_{(i,j),n}\|_{L^2(\Omega;\mathbb{R})} \leq \frac{\sqrt{m}h_n}{\sqrt{12}\pi K} + \frac{h_n}{\sqrt{6}\pi K} = \frac{\big(\sqrt{m}+\sqrt{2}\big)h_n}{\sqrt{12}\pi K}.$$

□

Proof of Theorem V.12

The proof follows the same ideas that are used in the proofs of Theorem V.11 and Theorem V.9.

Proof of Theorem V.12. Similarly to equations (V.89) and (V.98), it holds

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \big\|e_i^{\mathrm{T}} \Sigma^K_{2,n}(\Sigma^K_{1,n})^{-1}\sqrt{\Sigma^K_{1,n}}G_{0,n}\big\|^2_{L^p(\Omega;\mathbb{R})} \\
&= \max_{i\in\{1,\dots,M\}} \Big(\mathrm{E}\Big[\mathrm{E}\big[\big|e_i^{\mathrm{T}} \Sigma^K_{2,n}(\Sigma^K_{1,n})^{-1}\sqrt{\Sigma^K_{1,n}}G_{0,n}\big|^p\big|b_{k,n}, k\in\mathbb{N}\big]\Big]\Big)^{\frac{2}{p}} \\
&= \frac{2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \max_{i\in\{1,\dots,M\}} \big(\mathrm{E}\big[\|e_i^{\mathrm{T}} \Sigma^K_{2,n}(\Sigma^K_{1,n})^{-1}\sqrt{\Sigma^K_{1,n}}\|^p_{\mathrm{F}}\big]\big)^{\frac{2}{p}} \\
&= \frac{2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \bigg(\frac{h_n}{2\pi^2}\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)^{-1} \max_{i\in\{1,\dots,M\}} \big(\mathrm{E}\big[\|e_i^{\mathrm{T}} \Sigma^K_{2,n}\|^p_{\mathrm{F}}\big]\big)^{\frac{2}{p}}.
\end{aligned} \tag{V.103}$$

Considering the last factor on the right-hand side of equation (V.103) above, it follows, analogously to equations (V.99) and (V.91), that

$$\begin{aligned}
&\max_{i\in\{1,\dots,M\}} \Big(\mathrm{E}\big[\|e_i^{\mathrm{T}} \Sigma^K_{2,n}\|^p_{\mathrm{F}}\big]\Big)^{\frac{2}{p}} \\
&= \max_{i\in\{1,\dots,M\}} \bigg\|\sum_{l=1}^{m}|(\Sigma^K_{2,n})_{i,l}|^2\bigg\|_{L^{\frac{p}{2}}(\Omega;\mathbb{R})} \\
&\leq \max_{i\in\{1,\dots,M\}} \sum_{l=1}^{m}\|(\Sigma^K_{2,n})_{i,l}\|^2_{L^p(\Omega;\mathbb{R})} \\
&= \max_{\substack{i,j\in\{1,\dots,m\}\\ i<j}} \sum_{l=1}^{m}\bigg\|\frac{h_n}{2\pi}\bigg(\sum_{k=K+1}^{\infty}\frac{b^j_{k,n}}{k}e^l_i - \sum_{k=K+1}^{\infty}\frac{b^i_{k,n}}{k}e^l_j\bigg)\bigg\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq \Big(\frac{h_n}{2\pi}\Big)^2 \max_{\substack{i,j\in\{1,\dots,m\}\\ i<j}} \bigg(\bigg\|\sum_{k=K+1}^{\infty}\frac{b^j_{k,n}}{k}\bigg\|_{L^p(\Omega;\mathbb{R})} + \bigg\|\sum_{k=K+1}^{\infty}\frac{b^i_{k,n}}{k}\bigg\|_{L^p(\Omega;\mathbb{R})}\bigg)^2 \\
&= 4\Big(\frac{h_n}{2\pi}\Big)^2\bigg\|\sum_{k=K+1}^{\infty}\frac{b^1_{k,n}}{k}\bigg\|^2_{L^p(\Omega;\mathbb{R})}.
\end{aligned}$$

Then, Lemma V.20 implies

$$\max_{i\in\{1,\dots,M\}} \Big(\mathrm{E}\big[\|e_i^{\mathrm{T}} \Sigma^K_{2,n}\|^p_{\mathrm{F}}\big]\Big)^{\frac{2}{p}} \leq 4\Big(\frac{h_n}{2\pi}\Big)^2 \frac{2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \sum_{k=K+1}^{\infty}\frac{h_n}{2\pi^2 k^4}.$$

Inserting this into equation (V.103), we obtain

$$\begin{aligned}&\max_{i\in\{1,\ldots,M\}}\big\|e_i^{\mathrm{T}}\varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}\sqrt{\varSigma_{1,n}^K}G_{0,n}\big\|_{L^p(\Omega;\mathbb{R})}^2\\&\leq 4\Big(\frac{h_n}{2\pi}\Big)^2\frac{4\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\Big(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\Big)^{-1}\sum_{k=K+1}^{\infty}\frac{1}{k^4}.\end{aligned}$$

Further, using inequality (V.100), it holds

$$\max_{i\in\{1,\ldots,M\}}\big\|e_i^{\mathrm{T}}\varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}\sqrt{\varSigma_{1,n}^K}G_{0,n}\big\|_{L^p(\Omega;\mathbb{R})}^2\leq\Big(\frac{h_n}{2\pi}\Big)^2\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\frac{16}{3K^2}, \tag{V.104}$$

and similarly to inequality (V.102), we have

$$\begin{aligned}&\max_{i,j\in\{1,\ldots,m\}}\|I_{(i,j),n}-\tilde{I}_{(i,j),n}^{K+}\|_{L^p(\Omega;\mathbb{R})}\\&\leq\max_{i\in\{1,\ldots,M\}}\Big\|e_i^{\mathrm{T}}\Big(H_m\mathrm{vec}\big[(I_n)^{\mathrm{T}}\big]-H_m\mathrm{vec}\big[(I_n^{K+})^{\mathrm{T}}\big]\Big)\Big\|_{L^p(\Omega;\mathbb{R})}\\&\quad+\max_{i\in\{1,\ldots,M\}}\Big\|e_i^{\mathrm{T}}\Big(\varSigma_{2,n}^K(\varSigma_{1,n}^K)^{-1}\sqrt{\varSigma_{1,n}^K}G_{0,n}\Big)\Big\|_{L^p(\Omega;\mathbb{R})}.\end{aligned}$$

Then, Theorem V.9 and inequality (V.104) imply

$$\begin{aligned}&\max_{i,j\in\{1,\ldots,m\}}\|I_{(i,j),n}-\tilde{I}_{(i,j),n}^{K+}\|_{L^p(\Omega;\mathbb{R})}\\&\leq\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}\bigg(\bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}+1\bigg)^2+2(m-2)\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\bigg)^{\frac{1}{2}}\frac{\big(\sqrt{3}\sqrt{p-1}+1\big)h_n}{3\pi K}\\&\quad+\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{1}{p}}}\frac{2h_n}{\sqrt{3}\pi K}\\&=\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}\Bigg(\bigg(\bigg(\frac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}+1\bigg)^2+2(m-2)\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\bigg)^{\frac{1}{2}}\big(\sqrt{3}\sqrt{p-1}+1\big)\\&\quad+\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}\sqrt{12}\Bigg)\frac{h_n}{3\pi K}.\end{aligned}$$

□

Proof of Lemma V.16

The proof is divided into three parts. At first, we show that the absolute moments of the Milstein with approximated iterated stochastic integrals, defined in formula (V.56), are bounded. After that, we show the strong convergence of Milstein scheme (V.56) to Milstein approximation Y defined in formula (IV.33). From this, we conclude the assertion of Lemma V.16. The proofs below involve the discrete version of Gronwall's Lemma II.7.

Lemma V.24 (Discrete Gronwall, cf. [22])
Let $N\in\mathbb{N}$, and for all $n\in\{0,1,\ldots,N\}$, let $x_n\leq c+\sum_{k=0}^n y_kx_k$, where $c,x_n,y_n>0$. Then, it holds $x_n\leq c\,\mathrm{e}^{\sum_{k=0}^n y_k}$ for all $n\in\{0,1,\ldots,N\}$.

Proof. Cf. [22], and use that $1 + y_n \leq \mathrm{e}^{y_n}$, which implies $\Pi_{k=0}^{n}(1 + y_n) \leq \mathrm{e}^{\sum_{k=0}^{n} y_n}$. □

Lemma V.25

Let the Borel-measurable coefficients of SDDE (II.1) *fulfill Assumption IV.8 ii) and Assumption IV.8 iv), where* $b^j(t, t - \tau_1, \ldots, t - \tau_D, \cdot, \ldots, \cdot) \in C^1(\mathbb{R}^{d\times(D+1)}; \mathbb{R}^d)$ *for all* $t \in [t_0, T]$ *and* $j \in \{1, \ldots, m\}$*. Further, let initial condition* ξ *belong to* $S^p([t_0 - \tau, t_0] \times \Omega; \mathbb{R}^d)$ *for some* $p \in [2, \infty[$.

For $i, j \in \{1, \ldots, m\}$, $l \in \{0, 1, \ldots, D\}$, *and* $n \in \{0, 1, \ldots, N - 1\}$, *let approximation* $\bar{I}^{K_n}_{(i,j),n,\tau_l}$ *fulfill assumption* (V.54)*, be* $\mathscr{F}_{t_{n+1}}/\mathscr{B}(\mathbb{R})$*-measurable, and be independent of* $\mathscr{F}_{(t_n-\tau_l)\vee t_0}$.

Then, it holds for Milstein scheme $\overline{Y}$ *with approximated iterated stochastic integrals, defined in* (V.56)*, that*

$$\begin{aligned}
1 + &\left\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|\overline{Y}_t\| \right\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq \left(1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\right) \\
&\quad \times \mathrm{e}^{2\left(K_a\sqrt{T-t_0}+\frac{p}{\sqrt{p-1}}K_b m\sqrt{2}\left(\Gamma(\frac{p+1}{2})\right)^{\frac{1}{p}}\pi^{-\frac{1}{2p}}+L_b\sqrt{d}K_b C_{\bar{I},p}(D+1)m^2\sqrt{T-t_0}\right)^2(T-t_0)}.
\end{aligned}$$

Proof. Since $\xi \in S^p([t_0 - \tau, t_0] \times \Omega; \mathbb{R}^d)$, it holds

$$\left\| \sup_{t\in[t_0-\tau,t_0]} \|\overline{Y}_t\| \right\|_{L^p(\Omega;\mathbb{R})} = \|\xi\|_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} < \infty.$$

We assume that

$$\left\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}} \|\overline{Y}_t\| \right\|_{L^p(\Omega;\mathbb{R})} < \infty$$

has been proven for all $\nu \in \{0, 1, \ldots, n - 1\}$, where $n \in \{1, \ldots, N\}$. For all $n \in \{1, \ldots, N\}$, inequality (II.6) and the triangle inequality imply

$$\begin{aligned}
1 + &\left\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_n\}} \|\overline{Y}_t\| \right\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq 1 + 2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} + 2\Bigg(\left\| \sup_{\nu\in\{1,\ldots,n\}} \left\| \sum_{\mu=0}^{\nu-1} a(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu}))h_\mu \right\| \right\|_{L^p(\Omega;\mathbb{R})} \\
&\qquad + \left\| \sup_{\nu\in\{1,\ldots,n\}} \left\| \sum_{\mu=0}^{\nu-1} \sum_{j=1}^{m} b^j(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu}))\Delta W^j_\mu \right\| \right\|_{L^p(\Omega;\mathbb{R})} \qquad (\text{V.105}) \\
&\qquad + \Bigg\| \sup_{\nu\in\{1,\ldots,n\}} \Bigg\| \sum_{\mu=0}^{\nu-1} \sum_{l=0}^{D} \sum_{j_1,j_2=1}^{m} \sum_{i=1}^{d} \partial_{x^i_l} b^{j_1}(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu})) \\
&\qquad\qquad \times b^{i,j_2}(\mathcal{T}((t_\mu - \tau_l) \vee t_0, \overline{Y}_{(t_\mu-\tau_l)\vee t_0}))\bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l} \Bigg\| \Bigg\|_{L^p(\Omega;\mathbb{R})} \Bigg)^2.
\end{aligned}$$

In the following, we estimate the three $L^p(\Omega;\mathbb{R})$-norms on the right-hand side of previous inequality (V.105) separately. Using the triangle inequality and Assumption IV.8 *iv)*, we have

$$\begin{aligned}
&\Big\| \sup_{\nu\in\{1,\ldots,n\}} \Big\| \sum_{\mu=0}^{\nu-1} a(\mathcal{T}(t_\mu,\overline{Y}_{t_\mu}))h_\mu \Big\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\leq \Big\| \sup_{\nu\in\{1,\ldots,n\}} \sum_{\mu=0}^{\nu-1} \|a(\mathcal{T}(t_\mu,\overline{Y}_{t_\mu}))\| h_\mu \Big\|_{L^p(\Omega;\mathbb{R})} \\
&= \Big\| \sum_{\mu=0}^{n-1} \|a(\mathcal{T}(t_\mu,\overline{Y}_{t_\mu}))\| h_\mu \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\leq \Big\| \sum_{\mu=0}^{n-1} K_a \sup_{l\in\{0,1,\ldots,D\}} \big(1+\|\overline{Y}_{t_\mu-\tau_l}\|^2\big)^{\frac{1}{2}} h_\mu \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\leq K_a \sum_{\mu=0}^{n-1} \bigg(1+ \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\mu\}} \|\overline{Y}_t\| \Big\|^2_{L^p(\Omega;\mathbb{R})} \bigg)^{\frac{1}{2}} h_\mu \\
&\leq K_a \sqrt{t_n-t_0} \bigg(\sum_{\mu=0}^{n-1} \bigg(1+ \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\mu\}} \|\overline{Y}_t\| \Big\|^2_{L^p(\Omega;\mathbb{R})} \bigg) h_\mu \bigg)^{\frac{1}{2}},
\end{aligned} \tag{V.106}$$

where the Cauchy-Schwarz inequality is used in the last step. Next, we consider the second $L^p(\Omega;\mathbb{R})$-norm on the right-hand side of inequality (V.105). Here, the time-discrete process

$$\bigg(\sum_{\mu=0}^{n-1} \sum_{j=1}^{m} b^j(\mathcal{T}(t_\mu,\overline{Y}_{t_\mu}))\Delta W_\mu^j \bigg)_{n\in\{1,\ldots,N\}}$$

is a discrete martingales in $L^p(\Omega;\mathbb{R}^d)$ with respect to the filtration $(\mathscr{F}_{t_n})_{n\in\{1,\ldots,N\}}$. Using the discrete Burkholder-type inequality from Theorem II.5, the triangle inequality as well as the independence of $b^j(\mathcal{T}(t_\mu,\overline{Y}_{t_\mu}))$ and ΔW_μ^j, we obtain

$$\begin{aligned}
&\Big\| \sup_{\nu\in\{1,\ldots,n\}} \Big\| \sum_{\mu=0}^{\nu-1} \sum_{j=1}^{m} b^j(\mathcal{T}(t_\mu,\overline{Y}_{t_\mu}))\Delta W_\mu^j \Big\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\leq \frac{p}{\sqrt{p-1}} \bigg(\sum_{\mu=0}^{n-1} \Big\| \sum_{j=1}^{m} b^j(\mathcal{T}(t_\mu,\overline{Y}_{t_\mu}))\Delta W_\mu^j \Big\|^2_{L^p(\Omega;\mathbb{R}^d)} \bigg)^{\frac{1}{2}} \\
&\leq \frac{p}{\sqrt{p-1}} \bigg(\sum_{\mu=0}^{n-1} \bigg(\sum_{j=1}^{m} \|b^j(\mathcal{T}(t_\mu,\overline{Y}_{t_\mu}))\Delta W_\mu^j\|_{L^p(\Omega;\mathbb{R}^d)} \bigg)^2 \bigg)^{\frac{1}{2}} \\
&= \frac{p}{\sqrt{p-1}} \bigg(\sum_{\mu=0}^{n-1} \bigg(\sum_{j=1}^{m} \|b^j(\mathcal{T}(t_\mu,\overline{Y}_{t_\mu}))\|_{L^p(\Omega;\mathbb{R}^d)} \|\Delta W_\mu^j\|_{L^p(\Omega;\mathbb{R})} \bigg)^2 \bigg)^{\frac{1}{2}}.
\end{aligned} \tag{V.107}$$

Considering the two norms in term (V.107), Lemma V.20 and the linear growth condition from Assumption IV.8 *iv)* imply

$$\|\Delta W_\mu^j\|_{L^p(\Omega;\mathbb{R})} = \frac{\sqrt{2}\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} \sqrt{h_\mu}$$

and

$$\|b^j(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu}))\|_{L^p(\Omega;\mathbb{R}^d)} \leq K_b\bigg(1 + \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\mu\}} \|\overline{Y}_t\| \Big\|^2_{L^p(\Omega;\mathbb{R})}\bigg)^{\frac{1}{2}}. \tag{V.108}$$

Substituting this into inequality (V.107) leads to

$$\begin{aligned}
&\bigg\| \sup_{\nu\in\{1,\dots,n\}} \Big\| \sum_{\mu=0}^{\nu-1}\sum_{j=1}^{m} b^j(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu}))\Delta W_\mu^j \Big\| \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \frac{p}{\sqrt{p-1}} \frac{\sqrt{2}\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} \bigg(\sum_{\mu=0}^{n-1} \bigg(\sum_{j=1}^{m} \|b^j(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu}))\|_{L^p(\Omega;\mathbb{R}^d)} \bigg)^2 h_\mu \bigg)^{\frac{1}{2}} \\
&\quad\leq \frac{p\sqrt{2}\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\sqrt{p-1}\pi^{\frac{1}{2p}}} K_b m \bigg(\sum_{\mu=0}^{n-1} \bigg(1 + \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\mu\}} \|\overline{Y}_t\| \Big\|^2_{L^p(\Omega;\mathbb{R})} \bigg) h_\mu \bigg)^{\frac{1}{2}}.
\end{aligned} \tag{V.109}$$

We continue with the third $L^p(\Omega;\mathbb{R})$-norm on the right-hand side of inequality (V.105). Similarly to inequality (V.106), it holds by Assumption IV.8 *ii)* and inequality (IV.67) that

$$\begin{aligned}
&\bigg\| \sup_{\nu\in\{1,\dots,n\}} \Big\| \sum_{\mu=0}^{\nu-1}\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu})) \\
&\qquad\qquad \times b^{i,j_2}(\mathcal{T}((t_\mu-\tau_l)\vee t_0, \overline{Y}_{(t_\mu-\tau_l)\vee t_0}))\bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l} \Big\| \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \bigg\| \sum_{\mu=0}^{n-1} \Big\| \sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu})) \\
&\qquad\qquad \times b^{i,j_2}(\mathcal{T}((t_\mu-\tau_l)\vee t_0, \overline{Y}_{(t_\mu-\tau_l)\vee t_0}))\bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l} \Big\| \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \bigg\| \sum_{\mu=0}^{n-1}\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\sum_{i=1}^{d} \|\partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu}))\| \\
&\qquad\qquad \times |b^{i,j_2}(\mathcal{T}((t_\mu-\tau_l)\vee t_0, \overline{Y}_{(t_\mu-\tau_l)\vee t_0}))| \big|\bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l}\big| \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq L_b\sqrt{d} \bigg\| \sum_{\mu=0}^{n-1}\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m} \|b^{j_2}(\mathcal{T}((t_\mu-\tau_l)\vee t_0, \overline{Y}_{(t_\mu-\tau_l)\vee t_0}))\| \big|\bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l}\big| \bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq L_b\sqrt{d} \sum_{\mu=0}^{n-1}\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m} \Big\| \|b^{j_2}(\mathcal{T}((t_\mu-\tau_l)\vee t_0, \overline{Y}_{(t_\mu-\tau_l)\vee t_0}))\| \big|\bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l}\big| \Big\|_{L^p(\Omega;\mathbb{R})}.
\end{aligned} \tag{V.110}$$

Using further that $\bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l}$ is independent of $\mathscr{F}_{(t_\mu-\tau_l)\vee t_0}$ and that, similarly to inequality (V.108),

$$\begin{aligned}
&\|b^{j_2}(\mathcal{T}((t_\mu-\tau_l)\vee t_0, \overline{Y}_{(t_\mu-\tau_l)\vee t_0}))\|_{L^p(\Omega;\mathbb{R}^d)} \\
&\quad\leq K_b\bigg(1 + \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\mu\}} \|\overline{Y}_t\| \Big\|^2_{L^p(\Omega;\mathbb{R})}\bigg)^{\frac{1}{2}},
\end{aligned} \tag{V.111}$$

we obtain by assumption (V.54) from inequality (V.110) that

$$\begin{aligned}
&\Bigg\| \sup_{\nu\in\{1,\ldots,n\}} \Bigg\| \sum_{\mu=0}^{\nu-1} \sum_{l=0}^{D} \sum_{j_1,j_2=1}^{m} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\mu, \overline{Y}_{t_\mu})) \\
&\qquad \times b^{i,j_2}(\mathcal{T}((t_\mu-\tau_l)\vee t_0, \overline{Y}_{(t_\mu-\tau_l)\vee t_0})) \bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l} \Bigg\|\Bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\leq L_b\sqrt{d} \sum_{\mu=0}^{n-1} \sum_{l=0}^{D} \sum_{j_1,j_2=1}^{m} \|b^{j_2}(\mathcal{T}((t_\mu-\tau_l)\vee t_0, \overline{Y}_{(t_\mu-\tau_l)\vee t_0}))\|_{L^p(\Omega;\mathbb{R}^d)} \big\| \bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l} \big\|_{L^p(\Omega;\mathbb{R})} \\
&\leq L_b\sqrt{d}K_b \sum_{\mu=0}^{n-1} \Bigg(1+ \Bigg\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\mu\}} \|\overline{Y}_t\| \Bigg\|^2_{L^p(\Omega;\mathbb{R})} \Bigg)^{\frac{1}{2}} \sum_{l=0}^{D} \sum_{j_1,j_2=1}^{m} \big\| \bar{I}^{K_\mu}_{(j_2,j_1),\mu,\tau_l} \big\|_{L^p(\Omega;\mathbb{R})} \\
&\leq L_b\sqrt{d}K_b C_{\bar{I},p}(D+1)m^2 \sum_{\mu=0}^{n-1} \Bigg(1+ \Bigg\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\mu\}} \|\overline{Y}_t\| \Bigg\|^2_{L^p(\Omega;\mathbb{R})} \Bigg)^{\frac{1}{2}} h_\mu \\
&\leq L_b\sqrt{d}K_b C_{\bar{I},p}(D+1)m^2 \sqrt{t_n-t_0} \Bigg(\sum_{\mu=0}^{n-1} \Bigg(1+ \Bigg\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\mu\}} \|\overline{Y}_t\| \Bigg\|^2_{L^p(\Omega;\mathbb{R})} \Bigg) h_\mu \Bigg)^{\frac{1}{2}},
\end{aligned} \tag{V.112}$$

where again the Cauchy-Schwarz inequality was used in the last step. Inserting inequalities (V.106), (V.109), and (V.112) into inequality (V.105) yields

$$\begin{aligned}
&1+ \Bigg\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_n\}} \|\overline{Y}_t\| \Bigg\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq 1+2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)} + 2\Bigg(K_a\sqrt{t_n-t_0} + \frac{p\sqrt{2}\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\sqrt{p-1}\,\pi^{\frac{1}{2p}}} K_b m \\
&\qquad + L_b\sqrt{d}K_b C_{\bar{I},p}(D+1)m^2\sqrt{t_n-t_0} \Bigg)^2 \sum_{\mu=0}^{n-1} \Bigg(1+ \Bigg\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\mu\}} \|\overline{Y}_t\| \Bigg\|^2_{L^p(\Omega;\mathbb{R})} \Bigg) h_\mu.
\end{aligned}$$

Then, by discrete Gronwall's Lemma V.24, we have

$$\begin{aligned}
&1+ \Bigg\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_n\}} \|\overline{Y}_t\| \Bigg\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq \big(1+2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\big) \\
&\qquad \times \mathrm{e}^{2\left(K_a\sqrt{t_n-t_0}+\frac{p}{\sqrt{p-1}}\sqrt{2}\left(\Gamma(\frac{p+1}{2})\right)^{\frac{1}{p}}\pi^{-\frac{1}{2p}}K_b m + L_b\sqrt{d}K_b C_{\bar{I},p}(D+1)m^2\sqrt{t_n-t_0}\right)^2 (t_n-t_0)}.
\end{aligned}$$

Moreover, taking the maximum over $n \in \{1,\ldots,N\}$ on both sides of the inequality above implies that

$$\begin{aligned}
&1+ \Bigg\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_N\}} \|\overline{Y}_t\| \Bigg\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq \big(1+2\|\xi\|^2_{S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R}^d)}\big) \\
&\qquad \times \mathrm{e}^{2\left(K_a\sqrt{T-t_0}+\frac{p}{\sqrt{p-1}}\sqrt{2}\left(\Gamma(\frac{p+1}{2})\right)^{\frac{1}{p}}\pi^{-\frac{1}{2p}}K_b m + L_b\sqrt{d}K_b C_{\bar{I},p}(D+1)m^2\sqrt{T-t_0}\right)^2 (T-t_0)},
\end{aligned}$$

which proves the assertion. □

Lemma V.26
Let the assumptions from Lemma V.16 be fulfilled for some $p \in [2, \infty[$. *Consider Milstein approximation* Y *and Milstein scheme* $\overline{Y}$ *with approximated iterated stochastic integrals defined in formulas* (IV.33) *and* (V.56), *where both have maximum step size* h. *It holds*

$$\begin{aligned}
&\Big\| \sup_{t \in [t_0-\tau, t_0] \cup \{t_1, \dots, t_N\}} \|Y_t - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq \frac{\sqrt{2}p(D+1)m^2}{\sqrt{p-1}} \Bigg(\frac{pL_{\partial b}^2 (D+1)m^2 (4(\beta+1)p-1) \big(\Gamma(\frac{4(\beta+1)p+1}{2})\big)^{\frac{1}{2(\beta+1)p}}}{\sqrt{2}\sqrt{p-1}\pi^{\frac{1}{4(\beta+1)p}}} \\
&\qquad \times \Big\| \sup_{t \in [t_0-\tau, t_0[\cup \{t_0, t_1, \dots, t_\nu\}} \big(1 + \|Y_t\|^2 + \|\overline{Y}_t\|^2\big) \Big\|_{L^{(\beta+1)p}(\Omega;\mathbb{R})}^{\frac{2\beta+1}{2}} \sqrt{T-t_0}h \\
&\qquad + L_b \sqrt{d} K_b \Big(1 + \Big\| \sup_{t \in [t_0-\tau, t_0] \cup \{t_1, \dots, t_N\}} \|\overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}^2 \Big)^{\frac{1}{2}} \Big(\sum_{\nu=0}^{N-1} \big(\mathcal{E}_{\bar{I},p}(h_\nu, K_\nu) \big)^2 \Big)^{\frac{1}{2}} \Bigg) \\
&\quad \times \mathrm{e}^{\left(L_a \sqrt{T-t_0} + \frac{p\sqrt{2}(\Gamma(\frac{p+1}{2}))^{\frac{1}{p}}}{\sqrt{p-1}} \pi^{-\frac{1}{2p}} L_b m + \frac{1}{2} \right)^2 (T-t_0)}.
\end{aligned}$$

Proof. We introduce the auxiliary scheme $Y^{\overline{Y}}$ defined by

$$\begin{aligned}
&Y_t^{\overline{Y}} = \xi_t \quad \text{for } t \in [t_0-\tau, t_0] \text{ and} \\
&Y_{t_{n+1}}^{\overline{Y}} = Y_{t_n}^{\overline{Y}} + a(\mathcal{T}(t_n, \overline{Y}_{t_n}))h_n + \sum_{j=1}^m b^j(\mathcal{T}(t_n, \overline{Y}_{t_n}))\Delta W_n^j \\
&\qquad + \sum_{l=0}^D \sum_{j_1,j_2=1}^m \sum_{i=1}^d \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_n, \overline{Y}_{t_n})) b^{i,j_2}(\mathcal{T}((t_n - \tau_l) \vee t_0, \overline{Y}_{(t_n-\tau_l)\vee t_0})) I_{(j_2,j_1),n,\tau_l} \\
&\text{for } n = 0, 1, \dots, N-1
\end{aligned} \tag{V.113}$$

and estimate

$$\begin{aligned}
&\Big\| \sup_{t \in [t_0-\tau, t_0] \cup \{t_1, \dots, t_N\}} \|Y_t - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq \Big\| \sup_{t \in [t_0-\tau, t_0] \cup \{t_1, \dots, t_N\}} \|Y_t - Y_t^{\overline{Y}}\| \Big\|_{L^p(\Omega;\mathbb{R})} + \Big\| \sup_{t \in [t_0-\tau, t_0] \cup \{t_1, \dots, t_N\}} \|Y_t^{\overline{Y}} - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}.
\end{aligned} \tag{V.114}$$

Considering the first term on the right-hand side of inequality (V.114) above, the triangle inequality implies for schemes (IV.33) and (V.113) that

$$
\begin{aligned}
&\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|Y_t - Y_t^{\overline{Y}}\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&= \Big\| \sup_{n\in\{1,\dots,N\}} \|Y_{t_n} - Y_{t_n}^{\overline{Y}}\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\leq \Big\| \sup_{n\in\{1,\dots,N\}} \Big\| \sum_{\nu=0}^{n-1} (a(\mathcal{T}(t_\nu, Y_{t_\nu})) - a(\mathcal{T}(t_\nu, \overline{Y}_{t_\nu})))h_\nu \Big\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad + \Big\| \sup_{n\in\{1,\dots,N\}} \Big\| \sum_{\nu=0}^{n-1} \sum_{j=1}^{m} (b^j(\mathcal{T}(t_\nu, Y_{t_\nu})) - b^j(\mathcal{T}(t_\nu, \overline{Y}_{t_\nu})))\Delta W_\nu^j \Big\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad + \Big\| \sup_{n\in\{1,\dots,N\}} \Big\| \sum_{\nu=0}^{n-1} \sum_{l=0}^{D} \sum_{j_1,j_2=1}^{m} \sum_{i=1}^{d} \Big(\partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu, Y_{t_\nu})) b^{i,j_2}(\mathcal{T}((t_\nu - \tau_l)\vee t_0, Y_{(t_\nu-\tau_l)\vee t_0})) \\
&\qquad - \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu, \overline{Y}_{t_\nu})) b^{i,j_2}(\mathcal{T}((t_\nu - \tau_l)\vee t_0, \overline{Y}_{(t_\nu-\tau_l)\vee t_0})) \Big) I_{(j_2,j_1),\nu,\tau_l} \Big\| \Big\|_{L^p(\Omega;\mathbb{R})} .
\end{aligned}
\tag{V.115}
$$

In the following, we estimate the three terms on the right-hand side of inequality (V.115) above separately. Similarly to inequality (V.106), it holds, using the Lipschitz condition from Assumption IV.8 *ii)* and the Cauchy-Schwarz inequality, that

$$
\begin{aligned}
&\Big\| \sup_{n\in\{1,\dots,N\}} \Big\| \sum_{\nu=0}^{n-1} (a(\mathcal{T}(t_\nu, Y_{t_\nu})) - a(\mathcal{T}(t_\nu, \overline{Y}_{t_\nu})))h_\nu \Big\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\leq \sum_{\nu=0}^{N-1} \|a(\mathcal{T}(t_\nu, Y_{t_\nu})) - a(\mathcal{T}(t_\nu, \overline{Y}_{t_\nu}))\|_{L^p(\Omega;\mathbb{R}^d)} h_\nu \\
&\leq L_a \sum_{\nu=0}^{N-1} \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}} \|Y_t - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})} h_\nu \\
&\leq L_a \sqrt{T-t_0} \Big(\sum_{\nu=0}^{N-1} \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}} \|Y_t - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}^2 h_\nu \Big)^{\frac{1}{2}} .
\end{aligned}
\tag{V.116}
$$

Analogously to estimates (V.107), (V.109), and (V.116), we have

$$
\begin{aligned}
&\Big\| \sup_{n\in\{1,\dots,N\}} \Big\| \sum_{\nu=0}^{n-1} \sum_{j=1}^{m} (b^j(\mathcal{T}(t_\nu, Y_{t_\nu})) - b^j(\mathcal{T}(t_\nu, \overline{Y}_{t_\nu})))\Delta W_\nu^j \Big\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\leq \frac{p}{\sqrt{p-1}} \Big(\sum_{\nu=0}^{N-1} \Big(\sum_{j=1}^{m} \|b^j(\mathcal{T}(t_\nu, Y_{t_\nu})) - b^j(\mathcal{T}(t_\nu, \overline{Y}_{t_\nu}))\|_{L^p(\Omega;\mathbb{R}^d)} \|\Delta W_\nu^j\|_{L^p(\Omega;\mathbb{R})} \Big)^2 \Big)^{\frac{1}{2}} \\
&= \frac{p}{\sqrt{p-1}} \frac{\sqrt{2}\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}} \Big(\sum_{\nu=0}^{N-1} \Big(\sum_{j=1}^{m} \|b^j(\mathcal{T}(t_\nu, Y_{t_\nu})) - b^j(\mathcal{T}(t_\nu, \overline{Y}_{t_\nu}))\|_{L^p(\Omega;\mathbb{R}^d)} \sqrt{h_\nu} \Big)^2 \Big)^{\frac{1}{2}} \\
&\leq \frac{p\sqrt{2}\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\sqrt{p-1}\pi^{\frac{1}{2p}}} L_b m \Big(\sum_{\nu=0}^{N-1} \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}} \|Y_t - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}^2 h_\nu \Big)^{\frac{1}{2}} .
\end{aligned}
\tag{V.117}
$$

We proceed with the third term on the right-hand side of inequality (V.115) and introduce the notation

$$\mathit{III} := \Bigg\| \sup_{n\in\{1,\dots,N\}} \Bigg\| \sum_{\nu=0}^{n-1}\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\sum_{i=1}^{d}\Big(\partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu,Y_{t_\nu}))b^{i,j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, Y_{(t_\nu-\tau_l)\vee t_0})) \\ - \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu,\overline{Y}_{t_\nu}))b^{i,j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, \overline{Y}_{(t_\nu-\tau_l)\vee t_0}))\Big) I_{(j_2,j_1),\nu,\tau_l}\Bigg\|\Bigg\|_{L^p(\Omega;\mathbb{R})}.$$

Since approximation $\bar{I}^{K_\nu}_{(j_2,j_1),\nu,\tau_l}$ is $\mathscr{F}_{t_{\nu+1}}/\mathscr{B}(\mathbb{R})$-measurable and satisfies

$$\mathrm{E}\big[\bar{I}^{K_\nu}_{(j_2,j_1),\nu,\tau_l}\big|\mathscr{F}_{t_\nu}\big] = 0$$

P-almost surely for all $j_1, j_2 \in \{1,\dots,m\}$, $l \in \{0,1,\dots,D\}$, and $\nu \in \{0,1,\dots,N-1\}$, the time-discrete process inside the Euclidean norm is a martingale in $L^p(\Omega;\mathbb{R}^d)$. The discrete Burkholder-type inequality from Theorem II.5, the triangle inequality, and Assumption IV.8 *iii)* imply that

$$\mathit{III} \le \frac{p}{\sqrt{p-1}}\Bigg(\sum_{\nu=0}^{N-1}\Bigg\|\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\sum_{i=1}^{d}\Big(\partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu,Y_{t_\nu}))b^{i,j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, Y_{(t_\nu-\tau_l)\vee t_0})) \\ - \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu,\overline{Y}_{t_\nu}))b^{i,j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, \overline{Y}_{(t_\nu-\tau_l)\vee t_0}))\Big) I_{(j_2,j_1),\nu,\tau_l}\Bigg\|^2_{L^p(\Omega;\mathbb{R}^d)}\Bigg)^{\frac{1}{2}} \\ \le \frac{p}{\sqrt{p-1}}\Bigg(\sum_{\nu=0}^{N-1}\Bigg(\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\Bigg\|\Bigg\|\sum_{i=1}^{d}\Big(\partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu,Y_{t_\nu}))b^{i,j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, Y_{(t_\nu-\tau_l)\vee t_0})) \\ - \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu,\overline{Y}_{t_\nu}))b^{i,j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, \overline{Y}_{(t_\nu-\tau_l)\vee t_0}))\Big)\Bigg\| |I_{(j_2,j_1),\nu,\tau_l}|\Bigg\|_{L^p(\Omega;\mathbb{R})}\Bigg)^2\Bigg)^{\frac{1}{2}} \\ \le \frac{pL_{\partial b}}{\sqrt{p-1}}\Bigg(\sum_{\nu=0}^{N-1}\Bigg(\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\Bigg\|\Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}}\big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)^{\frac{\beta}{2}}\Bigg) \\ \Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}}\|Y_t-\overline{Y}_t\|\Bigg)|I_{(j_2,j_1),\nu,\tau_l}|\Bigg\|_{L^p(\Omega;\mathbb{R})}\Bigg)^2\Bigg)^{\frac{1}{2}}.$$

The following estimates are similar to those appearing in the considerations of term $\mathcal{R}_{11}$ in the proof of Theorem IV.9. Inequality (IV.161), where X and Y are replaced by Y and $\overline{Y}$, respectively, implies

$$\mathit{III} \le \frac{pL_{\partial b}2^{\frac{1}{4}}}{\sqrt{p-1}}\Bigg(\sum_{\nu=0}^{N-1}\Bigg(\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\Bigg\|\Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}}\big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)^{\frac{2\beta+1}{4}}\Bigg) \\ \Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}}\|Y_t-\overline{Y}_t\|\Bigg)^{\frac{1}{2}}|I_{(j_2,j_1),\nu,\tau_l}|\Bigg\|_{L^p(\Omega;\mathbb{R})}\Bigg)^2\Bigg)^{\frac{1}{2}}.$$

Using inequality (IV.163) with $\gamma = \Big(\frac{pL_{\partial b}(D+1)m^2 2^{\frac{1}{4}}}{\sqrt{p-1}\sqrt{h_\nu}}\Big)^{\frac{1}{2}}$ and the triangle inequality, we obtain

$$\begin{aligned}
I\!I\!I \leq \frac{pL_{\partial b}2^{\frac{1}{4}}}{\sqrt{p-1}}\Bigg(\sum_{\nu=0}^{N-1}\Bigg(\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\Bigg\|\frac{1}{2}\gamma^2\Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}}\big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)^{\frac{2\beta+1}{2}}\Bigg) \\
\times|I_{(j_2,j_1),\nu,\tau_l}|^2+\frac{1}{2}\gamma^{-2}\Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}}\|Y_t-\overline{Y}_t\|\Bigg)\Bigg\|_{L^p(\Omega;\mathbb{R})}\Bigg)^2\Bigg)^{\frac{1}{2}} \\
\leq \frac{pL_{\partial b}2^{\frac{1}{4}}}{\sqrt{p-1}}\Bigg(\sum_{\nu=0}^{N-1}\Bigg(\frac{1}{2}\gamma^2\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\Bigg\|\Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}}\big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)^{\frac{2\beta+1}{2}}\Bigg) \\
\times|I_{(j_2,j_1),\nu,\tau_l}|^2\Bigg\|_{L^p(\Omega;\mathbb{R})}+\frac{(D+1)m^2}{2\gamma^2}\Bigg\|\Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}}\|Y_t-\overline{Y}_t\|\Bigg)\Bigg\|_{L^p(\Omega;\mathbb{R})}\Bigg)^2\Bigg)^{\frac{1}{2}}.
\end{aligned}$$

Considering the term $\big(\sum_{\nu=0}^{N-1}x_i^2\big)^{\frac{1}{2}}$ on the right-hand side of the inequality above as the Euclidean norm and replacing γ by its definition, the triangle inequality leads to

$$\begin{aligned}
I\!I\!I \leq \frac{p^2L_{\partial b}^2(D+1)m^2}{\sqrt{2}(p-1)}\Bigg(\sum_{\nu=0}^{N-1}\Bigg(\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\Bigg\|\Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}}\big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)^{\frac{2\beta+1}{2}}\Bigg) \\
\times\frac{|I_{(j_2,j_1),\nu,\tau_l}|^2}{\sqrt{h_\nu}}\Bigg\|_{L^p(\Omega;\mathbb{R})}\Bigg)^2\Bigg)^{\frac{1}{2}}+\frac{1}{2}\Bigg(\sum_{\nu=0}^{N-1}\Bigg\|\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}}\|Y_t-\overline{Y}_t\|\Bigg\|^2_{L^p(\Omega;\mathbb{R})}h_\nu\Bigg)^{\frac{1}{2}}.
\end{aligned} \tag{V.118}$$

Hölder's inequality with $1=\frac{2\beta+1}{2(\beta+1)}+\frac{1}{2(\beta+1)}$ provides

$$\begin{aligned}
&\Bigg\|\Bigg(\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}}\big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)^{\frac{2\beta+1}{2}}\Bigg)\frac{|I_{(j_2,j_1),\nu,\tau_l}|^2}{\sqrt{h_\nu}}\Bigg\|_{L^p(\Omega;\mathbb{R})} \\
&\leq \Bigg\|\sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}}\big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)\Bigg\|^{\frac{2\beta+1}{2}}_{L^{(\beta+1)p}(\Omega;\mathbb{R})}\frac{1}{\sqrt{h_\nu}}\|I_{(j_2,j_1),\nu,\tau_l}\|^2_{L^{4(\beta+1)p}(\Omega;\mathbb{R})}
\end{aligned}$$

for the first $L^p(\Omega;\mathbb{R})$-norm in inequality (V.118) above. Further, Theorem II.6 and Lemma V.20 yield

$$\begin{aligned}
\|I_{(j_2,j_1),\nu,\tau_l}\|^2_{L^{4(\beta+1)p}(\Omega;\mathbb{R})} &= \Bigg\|\int_{t_\nu}^{t_{\nu+1}}\int_{(t_\nu-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}\mathrm{d}W_u^{j_2}\,\mathrm{d}W_s^{j_1}\Bigg\|^2_{L^{4(\beta+1)p}(\Omega;\mathbb{R})} \\
&\leq (4(\beta+1)p-1)\int_{t_\nu}^{t_{\nu+1}}\Bigg\|\int_{(t_\nu-\tau_l)\vee t_0}^{(s-\tau_l)\vee t_0}\mathrm{d}W_u^{j_2}\Bigg\|^2_{L^{4(\beta+1)p}(\Omega;\mathbb{R})}\mathrm{d}s \\
&\leq \frac{2(4(\beta+1)p-1)\big(\Gamma(\frac{4(\beta+1)p+1}{2})\big)^{\frac{1}{2(\beta+1)p}}}{\pi^{\frac{1}{4(\beta+1)p}}}\int_{t_\nu}^{t_{\nu+1}}s-t_\nu\,\mathrm{d}s \\
&= \frac{(4(\beta+1)p-1)\big(\Gamma(\frac{4(\beta+1)p+1}{2})\big)^{\frac{1}{2(\beta+1)p}}}{\pi^{\frac{1}{4(\beta+1)p}}}h_\nu^2.
\end{aligned}$$

Thus, by inserting these estimates into inequality (V.118) and using $\sum_{\nu=0}^{N-1} h_\nu^3 \le (T-t_0)h^2$, we have

$$\begin{aligned}
\mathit{III} \le{}& \frac{p^2 L_{\partial b}^2 (D+1)^2 m^4 (4(\beta+1)p-1)\big(\Gamma(\frac{4(\beta+1)p+1}{2})\big)^{\frac{1}{2(\beta+1)p}}}{\sqrt{2}(p-1)\pi^{\frac{1}{4(\beta+1)p}}} \\
&\times \bigg(\sum_{\nu=0}^{N-1} \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}} \big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)\Big\|_{L^{(\beta+1)p}(\Omega;\mathbb{R})}^{2\beta+1} h_\nu^3\bigg)^{\frac{1}{2}} \\
&+ \frac{1}{2}\bigg(\sum_{\nu=0}^{N-1} \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}} \|Y_t-\overline{Y}_t\|\Big\|_{L^p(\Omega;\mathbb{R})}^2 h_\nu\bigg)^{\frac{1}{2}} \\
\le{}& \frac{p^2 L_{\partial b}^2 (D+1)^2 m^4 (4(\beta+1)p-1)\big(\Gamma(\frac{4(\beta+1)p+1}{2})\big)^{\frac{1}{2(\beta+1)p}}}{\sqrt{2}(p-1)\pi^{\frac{1}{4(\beta+1)p}}} \\
&\times \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_N\}} \big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)\Big\|_{L^{(\beta+1)p}(\Omega;\mathbb{R})}^{\frac{2\beta+1}{2}} \sqrt{T-t_0}\,h \qquad \text{(V.119)} \\
&+ \frac{1}{2}\bigg(\sum_{\nu=0}^{N-1} \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}} \|Y_t-\overline{Y}_t\|\Big\|_{L^p(\Omega;\mathbb{R})}^2 h_\nu\bigg)^{\frac{1}{2}}.
\end{aligned}$$

Combining inequalities (V.115), (V.116), (V.117), and (V.119) results in

$$\begin{aligned}
&\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|Y_t - Y_t^{\overline{Y}}\|\Big\|_{L^p(\Omega;\mathbb{R})} \\
&\le \bigg(L_a\sqrt{T-t_0} + \frac{p\sqrt{2}\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\sqrt{p-1}\pi^{\frac{1}{2p}}} L_b m + \frac{1}{2}\bigg) \\
&\quad\times \bigg(\sum_{\nu=0}^{N-1} \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_\nu\}} \|Y_t-\overline{Y}_t\|\Big\|_{L^p(\Omega;\mathbb{R})}^2 h_\nu\bigg)^{\frac{1}{2}} \qquad \text{(V.120)} \\
&\quad+ \frac{p^2 L_{\partial b}^2 (D+1)^2 m^4 (4(\beta+1)p-1)\big(\Gamma(\frac{4(\beta+1)p+1}{2})\big)^{\frac{1}{2(\beta+1)p}}}{\sqrt{2}(p-1)\pi^{\frac{1}{4(\beta+1)p}}} \\
&\quad\times \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\dots,t_N\}} \big(1+\|Y_t\|^2+\|\overline{Y}_t\|^2\big)\Big\|_{L^{(\beta+1)p}(\Omega;\mathbb{R})}^{\frac{2\beta+1}{2}} \sqrt{T-t_0}\,h.
\end{aligned}$$

In the following, the second term of the right-hand side of the inequality (V.114) is considered. By rewriting and inserting schemes (V.56) and (V.113), we obtain

$$\begin{aligned}
&\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|Y_t^{\overline{Y}} - \overline{Y}_t\|\Big\|_{L^p(\Omega;\mathbb{R})} \\
&= \Big\| \sup_{n\in\{1,\dots,N\}} \|Y_{t_n}^{\overline{Y}} - \overline{Y}_{t_n}\|\Big\|_{L^p(\Omega;\mathbb{R})} \\
&= \bigg\| \sup_{n\in\{1,\dots,N\}} \bigg\| \sum_{\nu=0}^{n-1}\sum_{l=0}^{D}\sum_{j_1,j_2=1}^{m}\sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu, Z_{t_\nu}))\, b^{i,j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, Z_{(t_\nu-\tau_l)\vee t_0})) \\
&\qquad \times \big(I_{(j_2,j_1),\nu,\tau_l} - \bar{I}_{(j_2,j_1),\nu,\tau_l}^{K_\nu}\big)\bigg\|\bigg\|_{L^p(\Omega;\mathbb{R})}.
\end{aligned}$$

The time-discrete process inside the Euclidean norm is a discrete martingale in $L^p(\Omega;\mathbb{R}^d)$ because iterated stochastic integral $I_{(j_2,j_1),\nu,\tau_l}$ and its approximation $\bar{I}^{K_\nu}_{(j_2,j_1),\nu,\tau_l}$ are $\mathscr{F}_{t_{\nu+1}}/\mathscr{B}(\mathbb{R})$-measurable and satisfy

$$\mathrm{E}\big[I_{(j_2,j_1),\nu,\tau_l}\big|\mathscr{F}_{t_\nu}\big] = \mathrm{E}\big[\bar{I}^{K_\nu}_{(j_2,j_1),\nu,\tau_l}\big|\mathscr{F}_{t_\nu}\big] = 0$$

P-almost surely for all $j_1, j_2 \in \{1,\ldots,m\}$, $l \in \{0,1,\ldots,D\}$, and $\nu \in \{0,1,\ldots,N-1\}$. The discrete Burkholder-type inequality from Theorem II.5 implies

$$\begin{aligned}
&\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|Y_t^{\overline{Y}} - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \frac{p}{\sqrt{p-1}} \Bigg(\sum_{\nu=0}^{N-1} \Big\| \sum_{l=0}^{D} \sum_{j_1,j_2=1}^{m} \sum_{i=1}^{d} \partial_{x_l^i} b^{j_1}(\mathcal{T}(t_\nu,\overline{Y}_{t_\nu})) b^{i,j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, \overline{Y}_{(t_\nu-\tau_l)\vee t_0})) \\
&\qquad \times \big(I_{(j_2,j_1),\nu,\tau_l} - \bar{I}^{K_\nu}_{(j_2,j_1),\nu,\tau_l}\big)\Big\|^2_{L^p(\Omega;\mathbb{R}^d)} \Bigg)^{\frac{1}{2}}.
\end{aligned}$$

Then, using the triangle inequality, Assumption IV.8 *iii)*, and inequality (IV.67), we obtain

$$\begin{aligned}
&\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|Y_t^{\overline{Y}} - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \frac{pL_b\sqrt{d}}{\sqrt{p-1}} \Bigg(\sum_{\nu=0}^{N-1} \bigg(\sum_{l=0}^{D} \sum_{j_1,j_2=1}^{m} \Big\| \|b^{j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, \overline{Y}_{(t_\nu-\tau_l)\vee t_0}))\| \\
&\qquad \times \big|I_{(j_2,j_1),\nu,\tau_l} - \bar{I}^{K_\nu}_{(j_2,j_1),\nu,\tau_l}\big| \Big\|_{L^p(\Omega;\mathbb{R})} \bigg)^2 \Bigg)^{\frac{1}{2}} \\
&\quad= \frac{pL_b\sqrt{d}}{\sqrt{p-1}} \Bigg(\sum_{\nu=0}^{N-1} \bigg(\sum_{l=0}^{D} \sum_{j_1,j_2=1}^{m} \|b^{j_2}(\mathcal{T}((t_\nu-\tau_l)\vee t_0, \overline{Y}_{(t_\nu-\tau_l)\vee t_0}))\|_{L^p(\Omega;\mathbb{R}^d)} \\
&\qquad \times \big\|I_{(j_2,j_1),\nu,\tau_l} - \bar{I}^{K_\nu}_{(j_2,j_1),\nu,\tau_l}\big\|_{L^p(\Omega;\mathbb{R})} \bigg)^2 \Bigg)^{\frac{1}{2}},
\end{aligned}$$

where the equality holds by independence of $I_{(j_2,j_1),\nu,\tau_l}$ and $\bar{I}^{K_\nu}_{(j_2,j_1),\nu,\tau_l}$ from $\mathscr{F}_{(t_\nu-\tau_l)\vee t_0}$. Further, using inequality (V.111) and assumption (V.55), it follows

$$\begin{aligned}
&\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|Y_t^{\overline{Y}} - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad\leq \frac{pL_b\sqrt{d}K_b(D+1)m^2}{\sqrt{p-1}} \Bigg(1+ \Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|\overline{Y}_t\| \Big\|^2_{L^p(\Omega;\mathbb{R})}\Bigg)^{\frac{1}{2}} \Bigg(\sum_{\nu=0}^{N-1} \big(\mathcal{E}_{\bar{I},p}(h_\nu,K_\nu)\big)^2 \Bigg)^{\frac{1}{2}}.
\end{aligned} \tag{V.121}$$

Inserting inequalities (V.120) and (V.121) into inequality (V.114), we have by inequality (II.6) that

$$
\begin{aligned}
&\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|Y_t - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}^2 \\
&\quad = \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_N\}} \|Y_t - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}^2 \\
&\quad \leq 2\bigg(L_a\sqrt{T-t_0} + \frac{p\sqrt{2}\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\sqrt{p-1}\pi^{\frac{1}{2p}}} L_b m + \frac{1}{2}\bigg)^2 \\
&\qquad\quad \times \bigg(\sum_{\nu=0}^{N-1} \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}} \|Y_t - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}^2 h_\nu \bigg) \\
&\qquad + \frac{2p^2(D+1)^2m^4}{p-1} \Bigg(\frac{pL_{\partial b}^2(D+1)m^2(4(\beta+1)p-1)\big(\Gamma(\frac{4(\beta+1)p+1}{2})\big)^{\frac{1}{2(\beta+1)p}}}{\sqrt{2}\sqrt{p-1}\pi^{\frac{1}{4(\beta+1)p}}} \\
&\qquad\quad \times \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}} \big(1 + \|Y_t\|^2 + \|\overline{Y}_t\|^2\big) \Big\|_{L^{(\beta+1)p}(\Omega;\mathbb{R})}^{\frac{2\beta+1}{2}} \sqrt{T-t_0}h \\
&\qquad\quad + L_b\sqrt{d}K_b\bigg(1 + \Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|\overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}^2\bigg)^{\frac{1}{2}} \bigg(\sum_{\nu=0}^{N-1} \big(\mathcal{E}_{\overline{I},p}(h_\nu, K_\nu)\big)^2 \bigg)^{\frac{1}{2}} \Bigg)^2 .
\end{aligned}
$$

Then, the discrete Gronwall Lemma V.24 implies

$$
\begin{aligned}
&\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|Y_t - \overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}^2 \\
&\quad \leq \frac{2p^2(D+1)^2m^4}{p-1} \Bigg(\frac{pL_{\partial b}^2(D+1)m^2(4(\beta+1)p-1)\big(\Gamma(\frac{4(\beta+1)p+1}{2})\big)^{\frac{1}{2(\beta+1)p}}}{\sqrt{2}\sqrt{p-1}\pi^{\frac{1}{4(\beta+1)p}}} \\
&\qquad\quad \times \Big\| \sup_{t\in[t_0-\tau,t_0[\cup\{t_0,t_1,\ldots,t_\nu\}} \big(1 + \|Y_t\|^2 + \|\overline{Y}_t\|^2\big) \Big\|_{L^{(\beta+1)p}(\Omega;\mathbb{R})}^{\frac{2\beta+1}{2}} \sqrt{T-t_0}h \\
&\qquad\quad + L_b\sqrt{d}K_b\bigg(1 + \Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|\overline{Y}_t\| \Big\|_{L^p(\Omega;\mathbb{R})}^2\bigg)^{\frac{1}{2}} \bigg(\sum_{\nu=0}^{N-1} \big(\mathcal{E}_{\overline{I},p}(h_\nu, K_\nu)\big)^2 \bigg)^{\frac{1}{2}} \Bigg)^2 \\
&\qquad \times e^{2\big(L_a\sqrt{T-t_0} + \frac{p\sqrt{2}(\Gamma(\frac{p+1}{2}))^{\frac{1}{p}}}{\sqrt{p-1}}\pi^{-\frac{1}{2p}} L_b m + \frac{1}{2}\big)^2 (T-t_0)} .
\end{aligned}
$$

Finally, the assertion of Lemma V.26 follows by taking the square root. □

Proof of Lemma V.16. The triangle inequality implies

$$
\begin{aligned}
&\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|X_t - \overline{Y}_t^h\| \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq \Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|X_t - Y_t^h\| \Big\|_{L^p(\Omega;\mathbb{R})} + \Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\ldots,t_N\}} \|Y_t^h - \overline{Y}_t^h\| \Big\|_{L^p(\Omega;\mathbb{R})} .
\end{aligned}
$$

Then, due to Lemma V.26 and assumption (V.57), there exist constants $\overline{C}_1, \overline{C}_2 > 0$, independent of h and N, such that

$$\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - \overline{Y}_t^h\| \Big\|_{L^p(\Omega;\mathbb{R})} \leq \overline{C}_1 h + \overline{C}_2 \bigg(\sum_{\nu=0}^{N-1} \big(\mathcal{E}_{\bar{I},p}(h_\nu, K_\nu)\big)^2 \bigg)^{\frac{1}{2}}$$

for all $h \in\,]0, T-t_0]$. If further $\sum_{\nu=0}^{N-1} \big(\mathcal{E}_{\bar{I},p}(h_\nu, K_\nu)\big)^2 \in \mathcal{O}(h^2)$, the existence of constant $\overline{C}_{\text{Milstein}} > 0$, independent of h and N, such that

$$\Big\| \sup_{t\in[t_0-\tau,t_0]\cup\{t_1,\dots,t_N\}} \|X_t - \overline{Y}_t^h\| \Big\|_{L^p(\Omega;\mathbb{R})} \leq \overline{C}_{\text{Milstein}} h$$

for all $h \in\,]0, T-t_0]$ is evident. □

Proof of Theorem V.18

At first, we show that the approximations obtained by Algorithm V.4 satisfy assumption (V.55).

Lemma V.27
Let $p \in [2, \infty[$ and $n \in \{0, 1, \dots, N-1\}$. Consider approximation $I^K_{(i,j),n,\tau_l}$ defined by equations (V.20) *and* (V.21), *where $K \in \mathbb{N}$, $i, j \in \{1, \dots, m\}$, and $l \in \{0, 1, \dots, D\}$. It holds*

$$\max_{\substack{i,j\in\{1,\dots,m\}\\ l\in\{0,1,\dots,D\}}} \|I^K_{(i,j),n,\tau_l}\|_{L^p(\Omega;\mathbb{R})} \leq \frac{(p-1)\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}} h_n}{\sqrt{6}\pi^{\frac{1}{2p}}}.$$

Proof. Similarly to inequalities (V.60) and (V.61), it holds

$$\begin{aligned} \|I^K_{(i,j),n,\tau_l}\|^2_{L^p(\Omega;\mathbb{R})} &\leq (p-1)^2 \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}} h_n^2}{\pi^{\frac{2p+1}{p}}} \sum_{k=1}^K \frac{1}{k^2} \\ &\leq (p-1)^2 \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}} h_n^2}{\pi^{\frac{2p+1}{p}}} \sum_{k=1}^\infty \frac{1}{k^2} \\ &= (p-1)^2 \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}} h_n^2}{6\pi^{\frac{1}{p}}}. \end{aligned}$$

Since the upper bound is independent of $i, j \in \{1, \dots, m\}$ and $l \in \{0, 1, \dots, D\}$, the assertion follows by taking the maximum and the square root. □

Proof of Theorem V.18. From the construction of approximation $I^{K_n}_{(i,j),n,\tau_l}$, see Section V.1, it is evident that $I^{K_n}_{(i,j),n,\tau_l}$ is $\mathscr{F}_{t_{n+1}}/\mathscr{B}(\mathbb{R})$-measurable, independent of σ-algebra $\mathscr{F}_{(t_n-\tau_l)\vee t_0}$ and satisfies $\mathrm{E}\big[I^{K_n}_{(i,j),n,\tau_l} \big| \mathscr{F}_{t_n}\big] = 0$ P-almost surely for all $i, j \in \{1, \dots, m\}$, $l \in \{0, 1, \dots, D\}$, and $n \in \{0, 1, \dots, N-1\}$. Further, assumption (V.54) is fulfilled by Lemma V.27, and assumption (V.55) holds by Theorem V.2 with

$$\mathcal{E}_{\bar{I},p}(h_n, K_n) = \frac{(p-1)\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}} h_n}{\pi^{\frac{2p+1}{2p}} \sqrt{K_n}}.$$

Since $K_n \geq Ch^{-1}$ for all $n \in \{0, 1, \ldots, N-1\}$, we obtain

$$\sum_{n=0}^{N-1} \big(\mathcal{E}_{\bar{I},p}(h_n, K_n)\big)^2 \leq \frac{(p-1)^2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{2p+1}{p}}} \sum_{n=0}^{N-1} \frac{h_n^2}{K_n} \leq \frac{(p-1)^2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}}{\pi^{\frac{2p+1}{p}}} \frac{T-t_0}{C} h^2,$$

and the assertion of Theorem V.18 follows from Lemma V.16 and Corollary V.17. □

Proof of Theorem V.19

The proof is similar to the one of Theorem V.18. At first, we show that the approximations from Algorithm V.10 satisfy assumption (V.55).

Lemma V.28
Let $p \in [2, \infty[$ and $n \in \{0, 1, \ldots, N-1\}$. Consider approximation $I^K_{(i,j),n,\tau_l}$ defined by equation (V.41)*, where $K \in \mathbb{N}$ and $i, j \in \{1, \ldots, m\}$. It holds*

$$\max_{i,j \in \{1,\ldots,m\}} \|I^{K+}_{(i,j),n}\|_{L^p(\Omega;\mathbb{R})} \leq \left(\frac{p-1}{\sqrt{6}} + \frac{2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\sqrt{3}\pi^{\frac{2p+1}{2p}}} + \frac{1}{\pi}\right) \frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}} h_n}{\pi^{\frac{1}{2p}}}.$$

Proof. At first, it holds by the triangle inequality that

$$\begin{aligned}
&\max_{i,j \in \{1,\ldots,m\}} \|I^{K+}_{(i,j),n}\|_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq \max_{i,j \in \{1,\ldots,m\}} \|I^{K}_{(i,j),n}\|_{L^p(\Omega;\mathbb{R})} + \max_{i \in \{1,\ldots,M\}} \Big\|e_i^{\mathrm{T}} \Sigma^K_{2,n} (\Sigma^K_{1,n})^{-1} \sqrt{\Sigma^K_{1,n}} G_{0,n}\Big\|_{L^p(\Omega;\mathbb{R})} \\
&\qquad + \max_{i \in \{1,\ldots,M\}} \Big\|e_i^{\mathrm{T}} \sqrt{\mathrm{E}[S^K_n]} G_{1,n}\Big\|_{L^p(\Omega;\mathbb{R})},
\end{aligned}$$

where

$$\max_{i,j \in \{1,\ldots,m\}} \|I^{K}_{(i,j),n}\|_{L^p(\Omega;\mathbb{R})} \leq \frac{(p-1)\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}} h_n}{\sqrt{6}\pi^{\frac{1}{2p}}}$$

by Lemma V.27 and

$$\max_{i \in \{1,\ldots,M\}} \Big\|e_i^{\mathrm{T}} \Sigma^K_{2,n} (\Sigma^K_{1,n})^{-1} \sqrt{\Sigma^K_{1,n}} G_{0,n}\Big\|_{L^p(\Omega;\mathbb{R})} \leq \frac{2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}} h_n}{\sqrt{3}\pi^{\frac{p+1}{p}} K}$$

by inequality (V.104). Further, Lemma V.6 implies

$$\begin{aligned}
&\max_{i \in \{1,\ldots,M\}} \Big\|e_i^{\mathrm{T}} \sqrt{\mathrm{E}[S^K_n]} G_{1,n}\Big\|_{L^p(\Omega;\mathbb{R})} \\
&\quad \leq \frac{h_n}{\sqrt{2}\pi} \left(\left(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\right) - \left(\sum_{k=K+1}^{\infty} \frac{1}{k^2}\right)^{-1} \left(\sum_{k=K+1}^{\infty} \frac{1}{k^4}\right)\right)^{\frac{1}{2}} \max_{i \in \{1,\ldots,M\}} \|G^i_{1,n}\|_{L^p(\Omega;\mathbb{R})},
\end{aligned}$$

and using Lemma V.22 and Lemma V.23, we obtain

$$\begin{aligned}
&\Bigg(\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)-\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^2}\bigg)^{-1}\bigg(\sum_{k=K+1}^{\infty}\frac{1}{k^4}\bigg)\Bigg)^{\frac{1}{2}}\\
&\quad\leq\bigg(\frac{1}{K+\frac{1}{2}}-\frac{K+\frac{1}{2}}{3\big(K+\frac{3}{4}\big)^3}\bigg)^{\frac{1}{2}}\\
&\quad=\left(\frac{1}{K}\cdot\frac{3+\frac{23}{4K}+\frac{65}{16K^2}+\frac{65}{64K^3}}{3+\frac{33}{4K}+\frac{135}{16K^2}+\frac{243}{64K^3}+\frac{81}{128K^4}}\right)^{\frac{1}{2}}\\
&\quad\leq\frac{1}{\sqrt{K}}.
\end{aligned}$$

Thus, we have by Lemma V.20 that

$$\max_{i\in\{1,\dots,M\}}\big\|e_i^{\mathrm{T}}\sqrt{\mathrm{E}[S_n^K]}G_{1,n}\big\|_{L^p(\Omega;\mathbb{R})}\leq\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}h_n}{\pi^{\frac{2p+1}{2p}}\sqrt{K}},$$

and finally, we obtain

$$\begin{aligned}
\max_{i,j\in\{1,\dots,m\}}\|I_{(i,j),n}^{K+}\|_{L^p(\Omega;\mathbb{R})}&\leq\frac{(p-1)\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}h_n}{\sqrt{6}\pi^{\frac{1}{2p}}}+\frac{2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{2}{p}}h_n}{\sqrt{3}\pi^{\frac{p+1}{p}}K}+\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}h_n}{\pi^{\frac{2p+1}{2p}}\sqrt{K}}\\
&=\left(\frac{p-1}{\sqrt{6}}+\frac{2\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\sqrt{3}\pi^{\frac{2p+1}{2p}}K}+\frac{1}{\pi\sqrt{K}}\right)\frac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}h_n}{\pi^{\frac{1}{2p}}}.
\end{aligned}$$

Since $K\in\mathbb{N}$, we can neglect K and $\sqrt{K}$ in the denominator and obtain the proposed upper bound. □

Proof of Theorem V.19. From the construction of approximations $I_{(i,j),n,\tau_l}^{K_n+}$, see Section V.2, it is evident that $I_{(i,j),n,\tau_l}^{K_n+}$ is $\mathscr{F}_{t_{n+1}}/\mathscr{B}(\mathbb{R})$-measurable, independent of σ-algebra $\mathscr{F}_{(t_n-\tau_l)\vee t_0}$ and satisfies $\mathrm{E}\big[I_{(i,j),n,\tau_l}^{K_n+}\big|\mathscr{F}_{t_n}\big]=0$ P-almost surely for all $i,j\in\{1,\dots,m\}$ and $n\in\{0,1,\dots,N-1\}$. Further, assumption (V.54) is fulfilled by Lemma V.28, and assumption (V.55) holds by Theorem V.8 and Theorem V.9 with

$$\mathcal{E}_{\bar{I},p}(h_n,K_n)=\begin{cases}\dfrac{\sqrt{m}h_n}{\sqrt{12}\pi K_n} & \text{if } p=2,\\[2ex]
\dfrac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}\left(\left(\dfrac{2\big(\Gamma(\frac{2p+1}{2})\big)^{\frac{1}{p}}}{\pi^{\frac{1}{2p}}}+1\right)^2+2(m-2)\dfrac{\big(\Gamma(\frac{p+1}{2})\big)^{\frac{4}{p}}}{\pi^{\frac{2}{p}}}\right)^{\frac{1}{2}}\\
\qquad\times\dfrac{\big(\sqrt{3}\sqrt{p-1}+1\big)h_n}{3\pi K_n} & \text{if } p\in\,]2,\infty[.\end{cases}$$

Since $K_n\geq Ch^{-\frac{1}{2}}$ for all $n\in\{0,1,\dots,N-1\}$, we obtain

$$\sum_{n=0}^{N-1}\big(\mathcal{E}_{\bar{I},p}(h_n,K_n)\big)^2\in\mathcal{O}(h^2),$$

and the assertion of Theorem V.19 follows from Lemma V.16 and Corollary V.17. □

VI
Numerical Simulations

In this chapter, we illustrate our theoretical results from Chapter IV on the strong and pathwise convergence of the Euler-Maruyama and the Milstein scheme for SDDEs. In order to confirm the order of convergence of the approximations numerically, there is a great interest in analytical solutions of SDDEs.

In Section VI.1, we develop analytical solutions of linear as well as nonlinear SDDEs that can be simulated exactly. Here, we consider SDDEs with multidimensional noise that satisfy commutativity condition (V.1). These solutions are then used in Section VI.2 in order to provide meaningful simulation studies on the convergence of the Euler-Maruyama and the Milstein scheme. To the best of our knowledge, the presented numerical examples are the first that compare the Milstein approximations to analytical solutions of SDDEs, which are simulated error-free.

VI.1. Exact Simulation of Analytical Solutions of Stochastic Delay Differential Equations

Küchler and Platen derived analytical solutions of linear SDDEs in case of $d = m = D = 1$ in [82, Section 8]. In this section, we first recall their results and then address the problem of the exact simulation of this analytical solutions in case of multidimensional additive noise. Finally, we derive some analytical solutions of (nonlinear) SDDEs with more general noises that can be simulated correctly without approximations. For sake of simplicity, let $d = D = 1$ in the following. Moreover, let $m \in \mathbb{N}$ and $\tau = \tau_1 > 0$.

Consider the linear SDDE

$$X_t = \begin{cases} \xi_t & \text{if } t \in [t_0 - \tau, t_0], \\ \xi_{t_0} + \int_{t_0}^t a_1(s)X_s + a_2(s)X_{s-\tau} + a_3(s)\,\mathrm{d}s & \\ \qquad + \sum_{j=1}^m \int_{t_0}^t b_1^j(s)X_s + b_2^j(s)X_{s-\tau} + b_3^j(s)\,\mathrm{d}W_s^j & \text{if } t \in \,]t_0, T], \end{cases} \tag{VI.1}$$

where initial condition ξ belongs to $S^p([t_0-\tau,t_0]\times\Omega;\mathbb{R})$ for some $p\in[2,\infty[$ and has P-almost surely continuous realizations, and where coefficients $a_i, b_i^j\colon \mathbb{R}\to\mathbb{R}$, $i\in\{1,2,3\}$ and $j\in\{1,\dots,m\}$, are Borel-measurable and bounded functions. Linear SDDE (VI.1) can be transformed in a system of linear SODEs with random coefficients, see [82, Section 7]. The resulting linear SODEs can be analytically solved, see e. g. [7, Section 8.4], [46, Example 2.5.3] or [75, Section 5.6]. Let $s\in[t_0,T[$, and consider the fundamental solution $(\Phi_{s,t})_{t\in[s,T]}$ with

$$\Phi_{s,t} = \exp\bigg(\int_s^t a_1(u) - \frac{1}{2}\sum_{j=1}^m (b_1^j(u))^2\,\mathrm{d}u + \sum_{j=1}^m \int_s^t b_1^j(u)\,\mathrm{d}W_u^j\bigg)$$

for $t\in[s,T]$ P-almost surely, which is the unique strong solution of the linear and homogeneous SODE

$$\Phi_{s,t} = 1 + \int_s^t a_1(u)\Phi_{s,u}\,\mathrm{d}u + \sum_{j=1}^m \int_s^t b_1^j(u)\Phi_{s,u}\,\mathrm{d}W_u^j,$$

where $t\in[s,T]$. According to [82, Equations (8.3) and (8.4)], linear SDDE (VI.1) can be solved sequentially. First, we have

$$X_t = \xi_t$$

for $t\in[t_0-\tau,t_0]$ P-almost surely. Then, we sequentially obtain for $l\in\mathbb{N}_0$, as long as $t_0+l\tau<T$, that

$$\begin{aligned} X_t = \Phi_{t_0+l\tau,t}\bigg(X_{t_0+l\tau} &+ \int_{t_0+l\tau}^t \Phi_{t_0+l\tau,s}^{-1}\Big(a_2(s)X_{s-\tau} + a_3(s) \\ &\qquad - \sum_{j=1}^m b_1^j(s)\big(b_2^j(s)X_{s-\tau} + b_3^j(s)\big)\Big)\,\mathrm{d}s \\ &+ \sum_{j=1}^m \int_{t_0+l\tau}^t \Phi_{t_0+l\tau,s}^{-1}\big(b_2^j(s)X_{s-\tau}+b_3^j(s)\big)\,\mathrm{d}W_s^j\bigg) \end{aligned} \tag{VI.2}$$

for $t\in\,]t_0+l\tau,(t_0+(l+1)\tau)\wedge T]$ P-almost surely. Especially for $l=0$, we have

$$\begin{aligned} X_t = \Phi_{t_0,t}\bigg(\xi_{t_0} &+ \int_{t_0}^t \Phi_{t_0,s}^{-1}\Big(a_2(s)\xi_{s-\tau} + a_3(s) - \sum_{j=1}^m b_1^j(s)\big(b_2^j(s)\xi_{s-\tau}+b_3^j(s)\big)\Big)\,\mathrm{d}s \\ &+ \sum_{j=1}^m \int_{t_0}^t \Phi_{t_0,s}^{-1}\big(b_2^j(s)\xi_{s-\tau}+b_3^j(s)\big)\,\mathrm{d}W_s^j\bigg) \end{aligned} \tag{VI.3}$$

for all $t\in[t_0,(t_0+\tau)\wedge T]$ P-almost surely. We remark that [82, Equations (8.3) and (8.4)] contain a typo because term $\Phi_{t_0+l\tau,s}^{-1}$ is missing inside the Itô integral. We see from equation (VI.3) that it might be hard to simulate the solution on the interval $[t_0,t_0+\tau]$ already when not $b_2^j(s)=b_3^j(s)=0$ for all $s\in[t_0,T]$ and $j\in\{1,\dots,m\}$. However, using Itô's formula, it might be possible to solve the Itô integral for more general functions b_2^j and b_3^j and to simulate the resulting random variables exactly, cf. [84, Example 8.1.5].

Next, we consider the following linear SDDE with multiplicative noise from [82, Equation (9.1)] and address some problems on its simulation, which is claimed to be done in [82, Section 9].

Let $a_1, a_2, b_1^1 \in \mathbb{R}$ be constants and

$$X_t = \begin{cases} 1 & \text{if } t \in [-1,0] \text{ and} \\ \xi_{t_0} + \int_{t_0}^t a_1 X_s + a_2 X_{s-1} \, \mathrm{d}s + \int_{t_0}^t b_1^1 X_s \, \mathrm{d}W_s^1 & \text{if } t \in \,]0,2], \end{cases} \tag{VI.4}$$

where W^1 is a one-dimensional Wiener process with $W_0^1 = 0$ P-almost surely. Using equation (VI.2), we can state the analytical solution of linear SDDE with multiplicative noise (VI.4), see also [82, Equation (9.2)]. It holds

$$X_t = \begin{cases} 1 & \text{if } t \in [-1,0], \\ \mathrm{e}^{(a_1 - \frac{1}{2}(b_1^1)^2)t + b_1^1 W_t^1} \left(1 + a_2 \int_0^t \mathrm{e}^{-(a_1 - \frac{1}{2}(b_1^1)^2)s - b_1^1 W_s^1} \, \mathrm{d}s \right) & \text{if } t \in \,]0,1], \\ \mathrm{e}^{(a_1 - \frac{1}{2}(b_1^1)^2)(t-1) + b_1^1 (W_t^1 - W_1^1)} & \\ \quad \times \left(X_1 + a_2 \int_1^t X_{s-1} \mathrm{e}^{-(a_1 - \frac{1}{2}(b_1^1)^2)(s-1) - b_1^1 (W_s^1 - W_1^1)} \, \mathrm{d}s \right) & \text{if } t \in \,]1,2] \end{cases} \tag{VI.5}$$

for all $t \in [-1,2]$ P-almost surely. Considering the integral over time in case of $t \in [1,2]$ and inserting the solution $(X_{s-1})_{s \in [1,t]}$, we obtain

$$\begin{aligned} &\int_1^t X_{s-1} \mathrm{e}^{-(a_1 - \frac{1}{2}(b_1^1)^2)(s-1) - b_1^1 (W_s^1 - W_1^1)} \, \mathrm{d}s \\ &\quad = \mathrm{e}^{b_1^1 W_1^1} \int_1^t \mathrm{e}^{-b_1^1 (W_s^1 - W_{s-1}^1)} \left(1 + a_2 \int_0^{s-1} \mathrm{e}^{-(a_1 - \frac{1}{2}(b_1^1)^2)u - b_1^1 W_u^1} \, \mathrm{d}u \right) \mathrm{d}s \end{aligned} \tag{VI.6}$$

for all $t \in [1,2]$ P-almost surely. Küchler and Platen claim that they plot one realization of analytic solution (VI.5), see [82, Fig. 1 and Fig. 2]. Unfortunately, they do not provide any information how they simulate analytic solution (VI.5) with random variable

$$\int_0^t \mathrm{e}^{-(a_1 - \frac{1}{2}(b_1^1)^2)s - b_1^1 W_s^1} \, \mathrm{d}s$$

for some $t \in [0,1]$ and random variable in equation (VI.6) for some $t \in [1,2]$ exactly and error-free. In [82, Fig. 1 and Fig. 2], Küchler and Platen also plot a realization of the Euler-Maruyama and the Milstein scheme, respectively. Hence, they even have to generate these random variables conditionally given the increments of the Wiener process, which are involved in the numerical schemes. We remark that these random variables do not appear in analytical solutions of linear SODEs because there we have $a_2 = 0$.

In the following, we focus on linear SDDEs with additive noise, whose analytical solutions can be simulated exactly. More specifically, we consider the SDDEs

$$X_t = \begin{cases} \xi_t & \text{if } t \in [t_0 - \tau, t_0], \\ \xi_{t_0} + \int_{t_0}^t a_1(s) X_s + a_2(s) X_{s-\tau} + a_3(s) \, \mathrm{d}s + \sum_{j=1}^m \int_{t_0}^t b_3^j(s) \, \mathrm{d}W_s^j & \text{if } t \in \,]t_0, T], \end{cases} \tag{VI.7}$$

where $\xi \in C([t_0-\tau,t_0];\mathbb{R})$ and $a_1,a_2,a_3,b_3^j\colon\ \mathbb{R}\to\mathbb{R}$, $j\in\{1,\dots,m\}$, are Borel-measurable and bounded functions. By equation (VI.2), we sequentially obtain for $l\in\mathbb{N}_0$, as long as $t_0+l\tau<T$, the analytical solution

$$\begin{aligned} X_t = \mathrm{e}^{\int_{t_0+l\tau}^{t} a_1(u)\,\mathrm{d}u}\Bigg(X_{t_0+l\tau} + \int_{t_0+l\tau}^{t} \mathrm{e}^{-\int_{t_0+l\tau}^{s} a_1(u)\,\mathrm{d}u}\big(a_2(s)X_{s-\tau}+a_3(s)\big)\,\mathrm{d}s \\ + \sum_{j=1}^{m}\int_{t_0+l\tau}^{t} \mathrm{e}^{-\int_{t_0+l\tau}^{s} a_1(u)\,\mathrm{d}u} b_3^j(s)\,\mathrm{d}W_s^j\Bigg) \end{aligned} \tag{VI.8}$$

for all $t\in\,]t_0+l\tau,(t_0+(l+1)\tau)\wedge T]$ P-almost surely. Here, the Itô integral in formula (VI.8) above is a Wiener integral, and hence, it is normally distributed with expectation 0 and variance

$$\int_{t_0+l\tau}^{t}\sum_{j=1}^{m}\left(\mathrm{e}^{-\int_{t_0+l\tau}^{s} a_1(u)\,\mathrm{d}u} b_3^j(s)\right)^2\mathrm{d}s.$$

Thus, we can exactly simulate formula (VI.8) in case of $l=0$ and have

$$\begin{aligned} X_t = \mathrm{e}^{\int_{t_0}^{t} a_1(u)\,\mathrm{d}u}\Bigg(\xi_{t_0} + \int_{t_0}^{t} \mathrm{e}^{-\int_{t_0}^{s} a_1(u)\,\mathrm{d}u}\big(a_2(s)\xi_{s-\tau}+a_3(s)\big)\,\mathrm{d}s \\ + \sum_{j=1}^{m}\int_{t_0}^{t} \mathrm{e}^{-\int_{t_0}^{s} a_1(u)\,\mathrm{d}u} b_3^j(s)\,\mathrm{d}W_s^j\Bigg) \end{aligned} \tag{VI.9}$$

for $t\in\,]t_0,t_0+\tau]$ P-almost surely, where the integral over time is deterministic. However, for $l\in\mathbb{N}$, the integral over time is not deterministic anymore. Considering the case $l=1$, it P-almost surely holds

$$\begin{aligned} X_t = \mathrm{e}^{\int_{t_0+\tau}^{t} a_1(u)\,\mathrm{d}u}\Bigg(X_{t_0+\tau} + \int_{t_0+\tau}^{t} \mathrm{e}^{-\int_{t_0+\tau}^{s} a_1(u)\,\mathrm{d}u}\big(a_2(s)X_{s-\tau}+a_3(s)\big)\,\mathrm{d}s \\ + \sum_{j=1}^{m}\int_{t_0+\tau}^{t} \mathrm{e}^{-\int_{t_0+\tau}^{s} a_1(u)\,\mathrm{d}u} b_3^j(s)\,\mathrm{d}W_s^j\Bigg) \end{aligned}$$

$t\in\,]t_0+\tau,t_0+2\tau]$. Using equation (VI.9) with regard to term $X_{s-\tau}$, we obtain the random variable

$$\sum_{j=1}^{m}\int_{t_0+\tau}^{t}\mathrm{e}^{-\int_{t_0+\tau}^{s} a_1(u)\,\mathrm{d}u} a_2(s)\,\mathrm{e}^{\int_{t_0}^{s-\tau} a_1(u)\,\mathrm{d}u}\int_{t_0}^{s-\tau}\mathrm{e}^{-\int_{t_0}^{u} a_1(r)\,\mathrm{d}r} b_3^j(u)\,\mathrm{d}W_u^j\,\mathrm{d}s, \tag{VI.10}$$

which needs to be simulated. Using the substitution $s=v+\tau$ and stochastic integration by parts formula based on Itô's formula, see e. g. [64] or [75, p. 155], it P-almost surely holds for all $t \in]t_0, t_0+\tau]$ and $j \in \{1, \dots, m\}$ that

$$\begin{aligned}
&\sum_{j=1}^{m} \int_{t_0+\tau}^{t} \mathrm{e}^{-\int_{t_0+\tau}^{s} a_1(u)\,\mathrm{d}u} a_2(s)\, \mathrm{e}^{\int_{t_0}^{s-\tau} a_1(u)\,\mathrm{d}u} \int_{t_0}^{s-\tau} \mathrm{e}^{-\int_{t_0}^{u} a_1(r)\,\mathrm{d}r} b_3^j(u)\,\mathrm{d}W_u^j\,\mathrm{d}s \\
&\quad= \sum_{j=1}^{m} \int_{t_0}^{t-\tau} \mathrm{e}^{-\int_{t_0+\tau}^{v+\tau} a_1(u)\,\mathrm{d}u} a_2(s)\, \mathrm{e}^{\int_{t_0}^{v} a_1(u)\,\mathrm{d}u} \int_{t_0}^{v} \mathrm{e}^{-\int_{t_0}^{u} a_1(r)\,\mathrm{d}r} b_3^j(u)\,\mathrm{d}W_u^j\,\mathrm{d}v \\
&\quad= \sum_{j=1}^{m} \int_{t_0}^{t-\tau} \mathrm{e}^{-\int_{t_0+\tau}^{v+\tau} a_1(u)\,\mathrm{d}u} a_2(s)\, \mathrm{e}^{\int_{t_0}^{v} a_1(u)\,\mathrm{d}u}\,\mathrm{d}v \int_{t_0}^{t-\tau} \mathrm{e}^{-\int_{t_0}^{u} a_1(r)\,\mathrm{d}r} b_3^j(u)\,\mathrm{d}W_u^j \\
&\qquad - \sum_{j=1}^{m} \int_{t_0}^{t-\tau} \int_{t_0}^{u} \mathrm{e}^{-\int_{t_0+\tau}^{v+\tau} a_1(r)\,\mathrm{d}r} a_2(s)\, \mathrm{e}^{\int_{t_0}^{v} a_1(u)\,\mathrm{d}u}\,\mathrm{d}v\, \mathrm{e}^{-\int_{t_0}^{u} a_1(r)\,\mathrm{d}r} b_3^j(u)\,\mathrm{d}W_u^j,
\end{aligned} \tag{VI.11}$$

which is again normally distributed. According to equation (VI.11), the normally distributed random variables in (VI.9) and (VI.10) are not independent. Using similar considerations as in equation (VI.11), we obtain for all $l \in \mathbb{N}_0$ that all random variables occurring in equation (VI.8) are normally distributed. Thus, all these random variables can be generated exactly by taking their covariances into account.

Next, we go more into detail how to simulate the analytical solution of the linear SDDE with additive noise (VI.8). For sake of simplicity, let coefficients $a_1, a_2, a_3, b_3^j \in \mathbb{R}$, $j \in \{1, \dots, m\}$, be constant in the following.

Consider the points in time $t_n, t_{n+1} \in [t_0+l\tau, t_0+(l+1)\tau]$ with $t_n < t_{n+1}$, where $l \in \mathbb{N}_0$. Assume that we have simulated X_{t_n} already and that we are now interested in simulating $X_{t_{n+1}}$. At first, using equation (VI.2), we have for the analytical solution X of linear SDDE (VI.8) at the point in time t_{n+1} that

$$\begin{aligned}
X_{t_{n+1}} = \mathrm{e}^{a_1(t_{n+1}-(t_0+l\tau))} \Bigg(& X_{t_0+l\tau} + \int_{t_0+l\tau}^{t_{n+1}} \mathrm{e}^{-a_1(s-(t_0+l\tau))} (a_2 X_{s-\tau} + a_3)\,\mathrm{d}s \\
& + \sum_{j=1}^{m} \int_{t_0+l\tau}^{t_{n+1}} \mathrm{e}^{-a_1(s-(t_0+l\tau))} b_3^j\,\mathrm{d}W_s^j \Bigg)
\end{aligned} \tag{VI.12}$$

P-almost surely. Using a similar expression for X_{t_n} as in previous equation (VI.12) for $X_{t_{n+1}}$, we rewrite equation (VI.12) to

$$
\begin{aligned}
X_{t_{n+1}} = {}& \mathrm{e}^{a_1(t_{n+1}-t_n)}\mathrm{e}^{a_1(t_n-(t_0+l\tau))}\bigg(X_{t_0+l\tau} + \int_{t_0+l\tau}^{t_n} \mathrm{e}^{-a_1(s-(t_0+l\tau))}(a_2X_{s-\tau}+a_3)\,\mathrm{d}s \\
&+ \sum_{j=1}^{m}\int_{t_0+l\tau}^{t_n} \mathrm{e}^{-a_1(s-(t_0+l\tau))}b_3^j\,\mathrm{d}W_s^j \\
&+ \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-(t_0+l\tau))}(a_2X_{s-\tau}+a_3)\,\mathrm{d}s + \sum_{j=1}^{m}\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-(t_0+l\tau))}b_3^j\,\mathrm{d}W_s^j\bigg) \\
= {}& \mathrm{e}^{a_1(t_{n+1}-t_n)}\bigg(X_{t_n} + \mathrm{e}^{a_1(t_n-(t_0+l\tau))}\bigg(\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-(t_0+l\tau))}(a_2X_{s-\tau}+a_3)\,\mathrm{d}s \\
&+ \sum_{j=1}^{m}\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-(t_0+l\tau))}b_3^j\,\mathrm{d}W_s^j\bigg)\bigg) \\
= {}& \mathrm{e}^{a_1(t_{n+1}-t_n)}\bigg(X_{t_n} + \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}(a_2X_{s-\tau}+a_3)\,\mathrm{d}s \\
&+ \sum_{j=1}^{m} b_3^j\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}W_s^j\bigg)
\end{aligned} \tag{VI.13}
$$

for all $t_n, t_{n+1} \in [t_0+l\tau, t_0+(l+1)\tau]$ with $t_n < t_{n+1}$ P-almost surely, where $l \in \mathbb{N}_0$. Similarly to equations (VI.12) and (VI.13), it P-almost surely holds for $X_{s-\tau}$ in previous equation (VI.13), whenever $t_n, s \in [t_0+l\tau, t_0+(l+1)\tau]$ with $t_n \leq s$, where $l \in \mathbb{N}$, that

$$
\begin{aligned}
X_{s-\tau} = {}& \mathrm{e}^{a_1(s-\tau-(t_0+(l-1)\tau))}\bigg(X_{t_0+(l-1)\tau} + \int_{t_0+(l-1)\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_0+(l-1)\tau))}(a_2X_{u-\tau}+a_3)\,\mathrm{d}u \\
&+ \sum_{j=1}^{m}\int_{t_0+(l-1)\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_0+(l-1)\tau))}b_3^j\,\mathrm{d}W_u^j\bigg) \\
= {}& \mathrm{e}^{a_1(s-\tau-(t_n-\tau))}\mathrm{e}^{a_1(t_n-\tau-(t_0+(l-1)\tau))}\bigg(X_{t_0+(l-1)\tau} \\
&+ \int_{t_0+(l-1)\tau}^{t_n-\tau} \mathrm{e}^{-a_1(u-(t_0+(l-1)\tau))}(a_2X_{u-\tau}+a_3)\,\mathrm{d}u \\
&+ \sum_{j=1}^{m}\int_{t_0+(l-1)\tau}^{t_n-\tau} \mathrm{e}^{-a_1(u-(t_0+(l-1)\tau))}b_3^j\,\mathrm{d}W_u^j \\
&+ \int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_0+(l-1)\tau))}(a_2X_{u-\tau}+a_3)\,\mathrm{d}u \\
&+ \sum_{j=1}^{m}\int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_0+(l-1)\tau))}b_3^j\,\mathrm{d}W_u^j\bigg) \\
= {}& \mathrm{e}^{a_1(s-t_n)}\bigg(X_{t_n-\tau} + \int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_n-\tau))}(a_2X_{u-\tau}+a_3)\,\mathrm{d}u \\
&+ \sum_{j=1}^{m} b_3^j\int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_n-\tau))}\,\mathrm{d}W_u^j\bigg).
\end{aligned} \tag{VI.14}
$$

Inserting this into equation (VI.13), we obtain P-almost surely for all $t_n, t_{n+1} \in [t_0 + l\tau, t_0 +$

$(l+1)\tau]$ with $t_n < t_{n+1}$, where $l \in \mathbb{N}$, that

$$\begin{aligned} X_{t_{n+1}} = \mathrm{e}^{a_1(t_{n+1}-t_n)} \Big(& X_{t_n} + a_2(t_{n+1}-t_n) X_{t_n-\tau} + a_3 \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)} \, \mathrm{d}s \\ & + a_2 \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_n-\tau))} (a_2 X_{u-\tau} + a_3) \, \mathrm{d}u \, \mathrm{d}s \\ & + a_2 \sum_{j=1}^{m} b_3^j \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_n-\tau))} \, \mathrm{d}W_u^j \, \mathrm{d}s \\ & + \sum_{j=1}^{m} b_3^j \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)} \, \mathrm{d}W_s^j \Big). \end{aligned} \tag{VI.15}$$

Using the substitution $s = v + \tau$ and stochastic integration by parts formula based on Itô's formula, see e. g. [64] or [75, p. 155], it P-almost surely holds

$$\begin{aligned} & \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_n-\tau))} \, \mathrm{d}W_u^j \, \mathrm{d}s \\ & = \int_{t_n-\tau}^{t_{n+1}-\tau} \int_{t_n-\tau}^{v} \mathrm{e}^{-a_1(u-(t_n-\tau))} \, \mathrm{d}W_u^j \, \mathrm{d}v \\ & = (t_{n+1}-t_n) \int_{t_n-\tau}^{t_{n+1}-\tau} \mathrm{e}^{-a_1(s-(t_n-\tau))} \, \mathrm{d}W_s^j - \int_{t_n-\tau}^{t_{n+1}-\tau} (s-(t_n-\tau)) \, \mathrm{e}^{-a_1(s-(t_n-\tau))} \, \mathrm{d}W_s^j \end{aligned} \tag{VI.16}$$

for the iterated integral in equation (VI.15), and thus, we have

$$\begin{aligned} X_{t_{n+1}} = \mathrm{e}^{a_1(t_{n+1}-t_n)} \Big(& X_{t_n} + a_2(t_{n+1}-t_n) X_{t_n-\tau} + a_3 \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)} \, \mathrm{d}s \\ & + a_2 \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_n-\tau))} (a_2 X_{u-\tau} + a_3) \, \mathrm{d}u \, \mathrm{d}s \\ & + a_2(t_{n+1}-t_n) \sum_{j=1}^{m} b_3^j \int_{t_n-\tau}^{t_{n+1}-\tau} \mathrm{e}^{-a_1(s-(t_n-\tau))} \, \mathrm{d}W_s^j \\ & - a_2 \sum_{j=1}^{m} b_3^j \int_{t_n-\tau}^{t_{n+1}-\tau} (s-(t_n-\tau)) \, \mathrm{e}^{-a_1(s-(t_n-\tau))} \, \mathrm{d}W_s^j \\ & + \sum_{j=1}^{m} b_3^j \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)} \, \mathrm{d}W_s^j \Big) \end{aligned} \tag{VI.17}$$

for all $t_n, t_{n+1} \in [t_0 + l\tau, t_0 + (l+1)\tau]$ with $t_n < t_{n+1}$ P-almost surely, where $l \in \mathbb{N}$. Further, similarly to equation (VI.14), we have

$$\begin{aligned} X_{u-\tau} = \mathrm{e}^{a_1(u-(t_n-\tau))} \Big(& X_{t_n-2\tau} + \int_{t_n-2\tau}^{u-\tau} \mathrm{e}^{-a_1(r-(t_n-2\tau))} (a_2 X_{r-\tau} + a_3) \, \mathrm{d}r \\ & + \sum_{j=1}^{m} b_3^j \int_{t_n-2\tau}^{u-\tau} \mathrm{e}^{-a_1(r-(t_n-2\tau))} \, \mathrm{d}W_r^j \Big) \end{aligned}$$

for all $u \in [t_0 + (l-1)\tau, t_0 + l\tau]$ and $t_n \in [t_0 + l\tau, t_0 + (l+1)\tau]$ with $t_n - \tau \leq u$ P-almost surely, where $l \in \mathbb{N} \setminus \{1\}$. Inserting this into equation (VI.17) implies, analogously to equation (VI.15), that

$$\begin{aligned} X_{t_{n+1}} = \mathrm{e}^{a_1(t_{n+1}-t_n)}\Bigg(& X_{t_n} + a_2(t_{n+1}-t_n)X_{t_n-\tau} + \frac{1}{2}a_2^2(t_{n+1}-t_n)^2 X_{t_n-2\tau} \\ & + a_3 \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}s + a_2 a_3 \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_n-\tau))}\,\mathrm{d}u\,\mathrm{d}s \\ & + a_2^2 \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{s-\tau} \int_{t_n-2\tau}^{u-\tau} \mathrm{e}^{-a_1(r-(t_n-2\tau))}(a_2 X_{r-\tau} + a_3)\,\mathrm{d}r\,\mathrm{d}u\,\mathrm{d}s \\ & + a_2^2 \sum_{j=1}^m b_3^j \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{s-\tau} \int_{t_n-2\tau}^{u-\tau} \mathrm{e}^{-a_1(r-(t_n-2\tau))}\,\mathrm{d}W_r^j\,\mathrm{d}u\,\mathrm{d}s \\ & + a_2(t_{n+1}-t_n) \sum_{j=1}^m b_3^j \int_{t_n-\tau}^{t_{n+1}-\tau} \mathrm{e}^{-a_1(s-(t_n-\tau))}\,\mathrm{d}W_s^j \\ & - a_2 \sum_{j=1}^m b_3^j \int_{t_n-\tau}^{t_{n+1}-\tau} (s-(t_n-\tau))\,\mathrm{e}^{-a_1(s-(t_n-\tau))}\,\mathrm{d}W_s^j \\ & + \sum_{j=1}^m b_3^j \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}W_s^j \Bigg). \end{aligned} \tag{VI.18}$$

for all $t_n, t_{n+1} \in [t_0 + l\tau, t_0 + (l+1)\tau]$ with $t_n < t_{n+1}$ P-almost surely, where $l \in \mathbb{N} \setminus \{1\}$. We now consider the iterated integral, which has the Wiener integral as the integrand. Similarly to equation (VI.16), it holds by applying a substitution and the stochastic integration by parts formula twice that

$$\begin{aligned} & \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{s-\tau} \int_{t_n-2\tau}^{u-\tau} \mathrm{e}^{-a_1(r-(t_n-2\tau))}\,\mathrm{d}W_r^j\,\mathrm{d}u\,\mathrm{d}s \\ & = \int_{t_n-2\tau}^{t_{n+1}-2\tau} (s-(t_n-2\tau)) \int_{t_n-2\tau}^{s} \mathrm{e}^{-a_1(r-(t_n-2\tau))}\,\mathrm{d}W_r^j\,\mathrm{d}s \\ & \quad - \int_{t_n-2\tau}^{t_{n+1}-2\tau} \int_{t_n-2\tau}^{s} (r-(t_n-2\tau))\mathrm{e}^{-a_1(r-(t_n-2\tau))}\,\mathrm{d}W_r^j\,\mathrm{d}s \\ & = \frac{1}{2}(t_{n+1}-t_n)^2 \int_{t_n-2\tau}^{t_{n+1}-2\tau} \mathrm{e}^{-a_1(s-(t_n-2\tau))}\,\mathrm{d}W_s^j \\ & \quad - (t_{n+1}-t_n) \int_{t_n-2\tau}^{t_{n+1}-2\tau} (s-(t_n-2\tau))\,\mathrm{e}^{-a_1(s-(t_n-2\tau))}\,\mathrm{d}W_s^j \\ & \quad + \frac{1}{2} \int_{t_n-2\tau}^{t_{n+1}-2\tau} (s-(t_n-2\tau))^2\,\mathrm{e}^{-a_1(s-(t_n-2\tau))}\,\mathrm{d}W_s^j \end{aligned}$$

P-almost surely. Finally, inserting this into equation (VI.18), we obtain

$$\begin{aligned}
X_{t_{n+1}} = \mathrm{e}^{a_1(t_{n+1}-t_n)}\bigg(& X_{t_n} + a_2(t_{n+1}-t_n)X_{t_n-\tau} + \frac{1}{2}a_2^2(t_{n+1}-t_n)^2 X_{t_n-2\tau} \\
& + a_3\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}s + a_2a_3\int_{t_n}^{t_{n+1}}\int_{t_n-\tau}^{s-\tau} \mathrm{e}^{-a_1(u-(t_n-\tau))}\,\mathrm{d}u\,\mathrm{d}s \\
& + a_2^2\int_{t_n}^{t_{n+1}}\int_{t_n-\tau}^{s-\tau}\int_{t_n-2\tau}^{u-\tau} \mathrm{e}^{-a_1(r-(t_n-2\tau))}(a_2X_{r-\tau}+a_3)\,\mathrm{d}r\,\mathrm{d}u\,\mathrm{d}s \\
& + \frac{1}{2}a_2^2(t_{n+1}-t_n)^2\sum_{j=1}^m b_3^j\int_{t_n-2\tau}^{t_{n+1}-2\tau} \mathrm{e}^{-a_1(s-(t_n-2\tau))}\,\mathrm{d}W_s^j \\
& - a_2^2(t_{n+1}-t_n)\sum_{j=1}^m b_3^j\int_{t_n-2\tau}^{t_{n+1}-2\tau} (s-(t_n-2\tau))\,\mathrm{e}^{-a_1(s-(t_n-2\tau))}\,\mathrm{d}W_s^j \\
& + \frac{1}{2}a_2^2\sum_{j=1}^m b_3^j\int_{t_n-2\tau}^{t_{n+1}-2\tau} (s-(t_n-2\tau))^2\,\mathrm{e}^{-a_1(s-(t_n-2\tau))}\,\mathrm{d}W_s^j \\
& + a_2(t_{n+1}-t_n)\sum_{j=1}^m b_3^j\int_{t_n-\tau}^{t_{n+1}-\tau} \mathrm{e}^{-a_1(s-(t_n-\tau))}\,\mathrm{d}W_s^j \\
& - a_2\sum_{j=1}^m b_3^j\int_{t_n-\tau}^{t_{n+1}-\tau} (s-(t_n-\tau))\,\mathrm{e}^{-a_1(s-(t_n-\tau))}\,\mathrm{d}W_s^j \\
& + \sum_{j=1}^m b_3^j\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}W_s^j\bigg).
\end{aligned} \tag{VI.19}$$

for all $t_n, t_{n+1} \in [t_0 + l\tau, t_0 + (l+1)\tau]$ with $t_n < t_{n+1}$ P-almost surely, where $l \in \mathbb{N} \setminus \{1\}$. One can of course proceed similarly and replace $X_{r-\tau}$ in equation (VI.19). However, the formulas become rather long and lose their clarity more and more.

In the following, we go into detail how to generate the occurring random variables correctly. For sake of simplicity, we set $T = 3\tau$ and consider the points in time

$$\Big\{t_n = t_0 + n\hbar_M : n \in \{0,1,\ldots,3M\},\ \hbar_M := \frac{\tau}{M}\Big\} \subset [t_0, t_0+3\tau] \tag{VI.20}$$

for some $M \in \mathbb{N}$. In the following, we use that

$$t_n - \tau = t_0 + n\frac{\tau}{M} - \tau = t_0 + (n-M)\frac{\tau}{M} = t_{n-M}.$$

and $t_{n+1} - t_n = \hbar_M$. We take the equations (VI.13), (VI.17), and (VI.19) into account. It P-almost surely holds for the analytical solution of linear SDDE (VI.7) with additive noise and constant coefficients $a_1, a_2, a_3, b_3^j \in \mathbb{R}$, $j \in \{1,\ldots,m\}$, that

$$\begin{aligned}
X_{t_{n+1}} = \mathrm{e}^{a_1\hbar_M}\bigg(& X_{t_n} + a_2\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\xi_{s-\tau}\,\mathrm{d}s \\
& + a_3\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}s + \sum_{j=1}^m b_3^j\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}W_s^j\bigg)
\end{aligned} \tag{VI.21}$$

for $n \in \{0, 1, \ldots, M-1\}$,

$$
\begin{aligned}
X_{t_{n+1}} = \mathrm{e}^{a_1 \hbar_M} \Bigg(& X_{t_n} + a_2 \hbar_M X_{t_{n-M}} + a_2^2 \int_{t_{n-M}}^{t_{n-M+1}} \int_{t_{n-M}}^{s} \mathrm{e}^{-a_1(u-t_{n-M})} \xi_{u-\tau} \, \mathrm{d}u \, \mathrm{d}s \\
& + a_3 \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)} \, \mathrm{d}s + a_2 a_3 \int_{t_{n-M}}^{t_{n-M+1}} \int_{t_{n-M}}^{s} \mathrm{e}^{-a_1(u-t_{n-M})} \, \mathrm{d}u \, \mathrm{d}s \\
& + a_2 \hbar_M \sum_{j=1}^{m} b_3^j \int_{t_{n-M}}^{t_{n-M+1}} \mathrm{e}^{-a_1(s-t_{n-M})} \, \mathrm{d}W_s^j \\
& - a_2 \sum_{j=1}^{m} b_3^j \int_{t_{n-M}}^{t_{n-M+1}} (s - t_{n-M}) \, \mathrm{e}^{-a_1(s-t_{n-M})} \, \mathrm{d}W_s^j \\
& + \sum_{j=1}^{m} b_3^j \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)} \, \mathrm{d}W_s^j \Bigg)
\end{aligned}
\tag{VI.22}
$$

for $n \in \{M, M+1, \ldots, 2M-1\}$, and

$$
\begin{aligned}
X_{t_{n+1}} = \mathrm{e}^{a_1 \hbar_M} \Bigg(& X_{t_n} + a_2 \hbar_M X_{t_{n-M}} + \frac{1}{2} a_2^2 \hbar_M^2 X_{t_{n-2M}} \\
& + a_2^3 \int_{t_{n-2M}}^{t_{n-2M+1}} \int_{t_{n-2M}}^{s} \int_{t_{n-2M}}^{u} \mathrm{e}^{-a_1(r-t_{n-2M})} \xi_{r-\tau} \, \mathrm{d}r \, \mathrm{d}u \, \mathrm{d}s \\
& + a_3 \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)} \, \mathrm{d}s + a_2 a_3 \int_{t_{n-M}}^{t_{n-M+1}} \int_{t_{n-M}}^{s} \mathrm{e}^{-a_1(u-t_{n-M})} \, \mathrm{d}u \, \mathrm{d}s \\
& + a_2^2 a_3 \int_{t_{n-2M}}^{t_{n-2M+1}} \int_{t_{n-2M}}^{s} \int_{t_{n-2M}}^{u} \mathrm{e}^{-a_1(r-t_{n-2M})} \, \mathrm{d}r \, \mathrm{d}u \, \mathrm{d}s \\
& + \frac{1}{2} a_2^2 \hbar_M^2 \sum_{j=1}^{m} b_3^j \int_{t_{n-2M}}^{t_{n-2M+1}} \mathrm{e}^{-a_1(s-t_{n-2M})} \, \mathrm{d}W_s^j \\
& - a_2^2 \hbar_M \sum_{j=1}^{m} b_3^j \int_{t_{n-2M}}^{t_{n-2M+1}} (s - t_{n-2M}) \, \mathrm{e}^{-a_1(s-t_{n-2M})} \, \mathrm{d}W_s^j \\
& + \frac{1}{2} a_2^2 \sum_{j=1}^{m} b_3^j \int_{t_{n-2M}}^{t_{n-2M+1}} (s - t_{n-2M})^2 \, \mathrm{e}^{-a_1(s-t_{n-2M})} \, \mathrm{d}W_s^j \\
& + a_2 \hbar_M \sum_{j=1}^{m} b_3^j \int_{t_{n-M}}^{t_{n-M+1}} \mathrm{e}^{-a_1(s-t_{n-M})} \, \mathrm{d}W_s^j \\
& - a_2 \sum_{j=1}^{m} b_3^j \int_{t_{n-M}}^{t_{n-M+1}} (s - t_{n-M}) \, \mathrm{e}^{-a_1(s-t_{n-M})} \, \mathrm{d}W_s^j \\
& + \sum_{j=1}^{m} b_3^j \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)} \, \mathrm{d}W_s^j \Bigg).
\end{aligned}
\tag{VI.23}
$$

for $n \in \{2M, 2M+1, \ldots, 3M-1\}$. Here, we see for example that the random variable

$$
\sum_{j=1}^{m} b_3^j \int_{t_{n-M}}^{t_{n-M+1}} \mathrm{e}^{-a_1(s-t_{n-M})} \, \mathrm{d}W_s^j
\tag{VI.24}
$$

from equation (VI.21) is reused in equation (VI.22) and that the random variable

$$\sum_{j=1}^{m} b_3^j \int_{t_{n-M}}^{t_{n-M+1}} (s - t_{n-M})\, \mathrm{e}^{-a_1(s-t_{n-M})}\, \mathrm{d}W_s^j$$

in equation (VI.22) is not independent of the one in (VI.24).

Next, we focus on the distributional properties of the random variables, which are needed in order to simulate the formulas (VI.21), (VI.22), and (VI.23). Since we are interested in comparing the analytical solutions with some numerical approximations, we generate the corresponding random variables and the increments of the Wiener process simultaneously. If one is only interested in the simulation of the analytical solution, the random variable in (VI.24), for example, can be generated using one normally distributed random variable only instead of m, if $a_1 = 0$, or $2m$, if $a_1 \neq 0$, standard-normally distributed random variables.

In the following, we distinguish the cases $a_1 = 0$ and $a_1 \in \mathbb{R}\setminus\{0\}$. We start with the case of $a_1 = 0$. The random vectors

$$\left(\int_{t_n}^{t_{n+1}} \mathrm{d}W_s^j,\ \int_{t_n}^{t_{n+1}} (s - t_n)\, \mathrm{d}W_s^j,\ \int_{t_n}^{t_{n+1}} (s - t_n)^2\, \mathrm{d}W_s^j \right)^{\mathrm{T}},$$

$j \in \{1, \ldots, m\}$, are independent and normally distributed with expectation $0_{3\times 1}$ and covariance

$$\begin{pmatrix} \hbar_M & \frac{1}{2}\hbar_M^2 & \frac{1}{3}\hbar_M^3 \\ \frac{1}{2}\hbar_M^2 & \frac{1}{3}\hbar_M^3 & \frac{1}{4}\hbar_M^4 \\ \frac{1}{3}\hbar_M^3 & \frac{1}{4}\hbar_M^4 & \frac{1}{5}\hbar_M^5 \end{pmatrix} = \begin{pmatrix} \hbar_M^{\frac{1}{2}} & 0 & 0 \\ \frac{1}{2}\hbar_M^{\frac{3}{2}} & \frac{1}{\sqrt{12}}\hbar_M^{\frac{3}{2}} & 0 \\ \frac{1}{3}\hbar_M^{\frac{5}{2}} & \frac{1}{\sqrt{12}}\hbar_M^{\frac{5}{2}} & \frac{1}{\sqrt{180}}\hbar_M^{\frac{5}{2}} \end{pmatrix} \begin{pmatrix} \hbar_M^{\frac{1}{2}} & \frac{1}{2}\hbar_M^{\frac{3}{2}} & \frac{1}{3}\hbar_M^{\frac{5}{2}} \\ 0 & \frac{1}{\sqrt{12}}\hbar_M^{\frac{3}{2}} & \frac{1}{\sqrt{12}}\hbar_M^{\frac{5}{2}} \\ 0 & 0 & \frac{1}{\sqrt{180}}\hbar_M^{\frac{5}{2}} \end{pmatrix},$$

where the factorization follows from the Cholesky decomposition. According to [74, Corollary 6.11], there exist independently $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variables B_n, $G_{1,n}$, and $G_{2,n}$ such that

$$\int_{t_n}^{t_{n+1}} \mathrm{d}W_s^j = \hbar_M^{\frac{1}{2}} B_n^j,$$

$$\int_{t_n}^{t_{n+1}} (s - t_n)\, \mathrm{d}W_s^j = \frac{1}{2}\hbar_M^{\frac{3}{2}} B_n^j + \frac{1}{\sqrt{12}}\hbar_M^{\frac{3}{2}} G_{1,n}^j,$$

and

$$\int_{t_n}^{t_{n+1}} (s - t_n)^2\, \mathrm{d}W_s^j = \frac{1}{3}\hbar_M^{\frac{5}{2}} B_n^j + \frac{1}{\sqrt{12}}\hbar_M^{\frac{5}{2}} G_{1,n}^j + \frac{1}{\sqrt{180}}\hbar_M^{\frac{5}{2}} G_{2,n}^j$$

P-almost surely for $j \in \{1, \ldots, m\}$. Inserting these three equations into equations (VI.21), (VI.22), and (VI.23), we obtain formulas that can be simulated directly. We summarize this in the following example in case of $a_1 = 0$.

Example VI.1
Consider the linear SDDE with additive noise

$$X_t = \begin{cases} \xi_t & \text{if } t \in [t_0 - \tau, t_0] \text{ and} \\ \xi_{t_0} + \int_{t_0}^{t} a_2 X_{s-\tau} + a_3 \,\mathrm{d}s + \sum_{j=1}^{m} \int_{t_0}^{t} b_3^j \,\mathrm{d}W_s^j & \text{if } t \in \,]t_0, t_0 + 3\tau], \end{cases} \tag{VI.25}$$

where $\xi \in C([t_0 - \tau, t_0]; \mathbb{R})$ and $a_2, a_3, b_3^j \in \mathbb{R}$, $j \in \{1, \ldots, m\}$ are constants. Then, the analytical solution of SDDE (VI.25) *is exactly simulated on the points in time given in formula* (VI.20) *for $M \in \mathbb{N}$ as follows.*

i) *For $n = 0, 1, \ldots, M-1$, generate $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variable B_n, and calculate*

$$X_{t_{n+1}} = X_{t_n} + a_2 \int_{t_n}^{t_{n+1}} \xi_{s-\tau} \,\mathrm{d}s + a_3 \hbar_M + \hbar_M^{\frac{1}{2}} \sum_{j=1}^{m} b_3^j B_n^j.$$

ii) *For $n = M, M+1, \ldots, 2M-1$, generate independently $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variables B_n and $G_{1,n-M}$, and calculate*

$$\begin{aligned} X_{t_{n+1}} = X_{t_n} &+ a_2 \hbar_M X_{t_{n-M}} + a_2^2 \int_{t_{n-M}}^{t_{n-M+1}} \int_{t_{n-M}}^{s} \xi_{u-\tau} \,\mathrm{d}u \,\mathrm{d}s + a_3 \hbar_M + \frac{1}{2} a_2 a_3 \hbar_M^2 \\ &+ \frac{1}{2} a_2 \hbar_M^{\frac{3}{2}} \sum_{j=1}^{m} b_3^j \left(B_{n-M}^j - \frac{G_{1,n-M}^j}{\sqrt{3}} \right) + \hbar_M^{\frac{1}{2}} \sum_{j=1}^{m} b_3^j B_n^j. \end{aligned}$$

iii) *For $n = 2M, 2M+1, \ldots, 3M-1$, generate independently $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variables B_n, $G_{1,n-M}$, and $G_{2,n-2M}$, and calculate*

$$\begin{aligned} X_{t_{n+1}} = X_{t_n} &+ a_2 \hbar_M X_{t_{n-M}} + \frac{1}{2} a_2^2 \hbar_M^2 X_{t_{n-2M}} \\ &+ a_2^3 \int_{t_{n-2M}}^{t_{n-2M+1}} \int_{t_{n-2M}}^{s} \int_{t_{n-2M}}^{u} \xi_{r-\tau} \,\mathrm{d}r \,\mathrm{d}u \,\mathrm{d}s \\ &+ a_3 \hbar_M + \frac{1}{2} a_2 a_3 \hbar_M^2 + \frac{1}{6} a_2^2 a_3 \hbar_M^3 \\ &+ \frac{1}{2} a_2^2 \hbar_M^{\frac{5}{2}} \sum_{j=1}^{m} b_3^j \left(\frac{B_{n-2M}^j}{3} - \frac{G_{1,n-2M}^j}{\sqrt{12}} + \frac{G_{2,n-2M}^j}{\sqrt{180}} \right) \\ &+ \frac{1}{2} a_2 \hbar_M^{\frac{3}{2}} \sum_{j=1}^{m} b_3^j \left(B_{n-M}^j - \frac{G_{1,n-M}^j}{\sqrt{3}} \right) + \hbar_M^{\frac{1}{2}} \sum_{j=1}^{m} b_3^j B_n^j. \end{aligned}$$

Now, we proceed with the case $a_1 \in \mathbb{R} \setminus \{0\}$. The random vectors

$$\begin{aligned} \bigg(\int_{t_n}^{t_{n+1}} \mathrm{d}W_s^j, \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)} \,\mathrm{d}W_s^j, \int_{t_n}^{t_{n+1}} (s-t_n) \,\mathrm{e}^{-a_1(s-t_n)} \,\mathrm{d}W_s^j, \ldots \\ \int_{t_n}^{t_{n+1}} (s-t_n)^2 \,\mathrm{e}^{-a_1(s-t_n)} \,\mathrm{d}W_s^j \bigg)^{\mathrm{T}}, \end{aligned}$$

$j \in \{1, \ldots, m\}$, are independent and normally distributed with expectation $0_{4\times 1}$ and covariance $\Sigma \in \mathbb{R}^{4\times 4}$ where $\Sigma_{1,1} = \hbar_M$,

$$\Sigma_{2,1} = \Sigma_{1,2} = \int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}s = \frac{1}{a_1}\big(1 - \mathrm{e}^{-a_1\hbar_M}\big),$$

$$\Sigma_{2,2} = \int_{t_n}^{t_{n+1}} \mathrm{e}^{-2a_1(s-t_n)}\,\mathrm{d}s = \frac{1}{2a_1}\big(1 - \mathrm{e}^{-2a_1\hbar_M}\big),$$

$$\Sigma_{3,1} = \Sigma_{1,3} = \int_{t_n}^{t_{n+1}} (s-t_n)\,\mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}s = \frac{1}{a_1^2}\big(1 - \mathrm{e}^{-a_1\hbar_M}(1 + a_1\hbar_M)\big),$$

$$\Sigma_{3,2} = \Sigma_{2,3} = \int_{t_n}^{t_{n+1}} (s-t_n)\,\mathrm{e}^{-2a_1(s-t_n)}\,\mathrm{d}s = \frac{1}{4a_1^2}\big(1 - \mathrm{e}^{-2a_1\hbar_M}(1 + 2a_1\hbar_M)\big),$$

$$\begin{aligned}\Sigma_{3,3} = \Sigma_{4,2} = \Sigma_{2,4} &= \int_{t_n}^{t_{n+1}} (s-t_n)^2\,\mathrm{e}^{-2a_1(s-t_n)}\,\mathrm{d}s \\ &= \frac{1}{4a_1^3}\big(1 - \mathrm{e}^{-2a_1\hbar_M}(1 + 2a_1\hbar_M + 2a_1^2\hbar_M^2)\big),\end{aligned}$$

$$\Sigma_{4,1} = \Sigma_{1,4} = \int_{t_n}^{t_{n+1}} (s-t_n)^2\,\mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}s = \frac{1}{a_1^3}\big(2 - \mathrm{e}^{-a_1\hbar_M}(2 + 2a_1\hbar_M + a_1^2\hbar_M^2)\big),$$

$$\begin{aligned}\Sigma_{4,3} = \Sigma_{3,4} &= \int_{t_n}^{t_{n+1}} (s-t_n)^3\,\mathrm{e}^{-2a_1(s-t_n)}\,\mathrm{d}s \\ &= \frac{1}{8a_1^4}\big(3 - \mathrm{e}^{-2a_1\hbar_M}(3 + 6a_1\hbar_M + 6a_1^2\hbar_M^2 + 4a_1^3\hbar_M^3)\big),\end{aligned}$$

and

$$\begin{aligned}\Sigma_{4,4} &= \int_{t_n}^{t_{n+1}} (s-t_n)^4\,\mathrm{e}^{-2a_1(s-t_n)}\,\mathrm{d}s \\ &= \frac{1}{4a_1^5}\big(3 - \mathrm{e}^{-2a_1\hbar_M}(3 + 6a_1\hbar_M + 6a_1^2\hbar_M^2 + 4a_1^3\hbar_M^3 + 2a_1^4\hbar_M^4)\big).\end{aligned}$$

Consider the Cholesky decomposition $\Sigma = L_\Sigma L_\Sigma^{\mathrm{T}}$, where

$$L_\Sigma = \begin{pmatrix} \ell_{1,1} & 0 & 0 & 0 \\ \ell_{2,1} & \ell_{2,2} & 0 & 0 \\ \ell_{3,1} & \ell_{3,2} & \ell_{3,3} & 0 \\ \ell_{4,1} & \ell_{4,2} & \ell_{4,3} & \ell_{4,4} \end{pmatrix} \tag{VI.26}$$

is given in the example below. Then, according to [74, Corollary 6.11], there exist independent and $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variables B_n, $G_{0,n}$, $G_{1,n}$, and $G_{2,n}$ such that

$$\int_{t_n}^{t_{n+1}} \mathrm{d}W_s^j = \ell_{1,1} B_n^j = \hbar_M^{\frac{1}{2}} B_n^j, \tag{VI.27}$$

$$\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}W_s^j = \ell_{2,1}B_n^j + \ell_{2,2}G_{0,n}^j, \tag{VI.28}$$

$$\int_{t_n}^{t_{n+1}} (s-t_n)\,\mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}W_s^j = \ell_{3,1}B_n^j + \ell_{3,2}G_{0,n}^j + \ell_{3,3}G_{1,n}^j, \tag{VI.29}$$

and

$$\int_{t_n}^{t_{n+1}} (s-t_n)^2\,\mathrm{e}^{-a_1(s-t_n)}\,\mathrm{d}W_s^j = \ell_{4,1}B_n^j + \ell_{4,2}G_{0,n}^j + \ell_{4,3}G_{2,n}^j + \ell_{4,4}G_{3,n}^j \tag{VI.30}$$

P-almost surely, where $j \in \{1,\ldots,m\}$. Inserting these equations (VI.27), (VI.27), (VI.27), and (VI.30) into equations (VI.21), (VI.22), and (VI.23), we can simulate the analytical solution. We summarize this in the following example.

Example VI.2
Consider the linear SDDE with additive noise

$$X_t = \begin{cases} \xi_t & \text{if } t \in [t_0-\tau, t_0] \text{ and} \\ \xi_{t_0} + \int_{t_0}^{t} a_1X_s + a_2X_{s-\tau} + a_3\,\mathrm{d}s + \sum_{j=1}^{m}\int_{t_0}^{t} b_3^j\,\mathrm{d}W_s^j & \text{if } t \in\,]t_0, t_0+3\tau], \end{cases} \tag{VI.31}$$

where $\xi \in C([t_0-\tau, t_0];\mathbb{R})$ and $a_1, a_2, a_3, b_3^j \in \mathbb{R}$, $j \in \{1,\ldots,m\}$, are constants. In case of $a_1 = 0$, we refer to Example VI.1. So let $a_1 \neq 0$ in the following. Further, define $\hbar_M \coloneqq \frac{\tau}{M}$ for some $M \in \mathbb{N}$.

At first, compute the Cholesky decomposition of the covariance matrix $\Sigma = L_\Sigma L_\Sigma^{\mathrm{T}}$. That is, the entries of the matrix L_Σ, see formula (VI.26)*, are given by $\ell_{1,1} = \sqrt{\hbar_M}$,*

$$\ell_{2,1} = \frac{\Sigma_{2,1}}{\ell_{1,1}}, \quad \ell_{3,1} = \frac{\Sigma_{3,1}}{\ell_{1,1}}, \quad \ell_{4,1} = \frac{\Sigma_{4,1}}{\ell_{1,1}}, \quad \ell_{2,2} = \big(\Sigma_{2,2} - \ell_{2,1}^2\big)^{\frac{1}{2}},$$

$$\ell_{3,2} = \frac{\Sigma_{3,2} - \ell_{2,1}\ell_{3,1}}{\ell_{2,2}}, \quad \ell_{4,2} = \frac{\Sigma_{4,2} - \ell_{2,1}\ell_{4,1}}{\ell_{2,2}}, \quad \ell_{3,3} = \big(\Sigma_{3,3} - \ell_{3,1}^2 - \ell_{3,2}^2\big)^{\frac{1}{2}},$$

$$\ell_{4,3} = \frac{\Sigma_{4,3} - \ell_{3,1}\ell_{4,1} - \ell_{3,2}\ell_{4,2}}{\ell_{3,3}}, \quad \text{and} \quad \ell_{4,4} = \big(\Sigma_{4,4} - \ell_{4,1}^2 - \ell_{4,2}^2 - \ell_{4,3}^2\big)^{\frac{1}{2}}.$$

Then, the analytical solution of the SDDE (VI.31) *is exactly simulated on the points in time given in* (VI.20) *for $M \in \mathbb{N}$ as follows.*

i) For $n = 0, 1, \ldots, M-1$, generate independently $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variables B_n and $G_{0,n}$, and calculate

$$X_{t_{n+1}} = \mathrm{e}^{a_1\hbar_M}\bigg(X_{t_n} + a_2\int_{t_n}^{t_{n+1}} \mathrm{e}^{-a_1(s-t_n)}\xi_{s-\tau}\,\mathrm{d}s + \frac{a_3}{a_1}\big(1-\mathrm{e}^{-a_1\hbar_M}\big) + \sum_{j=1}^{m} b_3^j\big(\ell_{2,1}B_n^j + \ell_{2,2}G_{0,n}^j\big)\bigg).$$

ii) For $n = M, M+1, \dots, 2M-1$, generate independently $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variables B_n, $G_{0,n}$, and $G_{1,n-M}$, and calculate

$$X_{t_{n+1}} = \mathrm{e}^{a_1 \hbar_M}\Bigg(X_{t_n} + a_2 \hbar_M X_{t_{n-M}} + a_2^2 \int_{t_{n-M}}^{t_{n-M+1}} \int_{t_{n-M}}^{s} \mathrm{e}^{-a_1(u - t_{n-M})} \xi_{u-\tau}\, \mathrm{d}u\, \mathrm{d}s$$
$$+ \frac{a_3}{a_1}\big(1 - \mathrm{e}^{-a_1 \hbar_M}\big) + \frac{a_2 a_3}{a_1^2}\big(\mathrm{e}^{-a_1 \hbar_M} - 1 + a_1 \hbar_M\big)$$
$$+ a_2 \sum_{j=1}^{m} b_3^j \Big((\hbar_M \ell_{2,1} - \ell_{3,1}) B_{n-M}^j + (\hbar_M \ell_{2,2} - \ell_{3,2}) G_{0,n-M}^j - \ell_{3,3} G_{1,n-M}^j \Big)$$
$$+ \sum_{j=1}^{m} b_3^j \big(\ell_{2,1} B_n^j + \ell_{2,2} G_{0,n}^j \big) \Bigg).$$

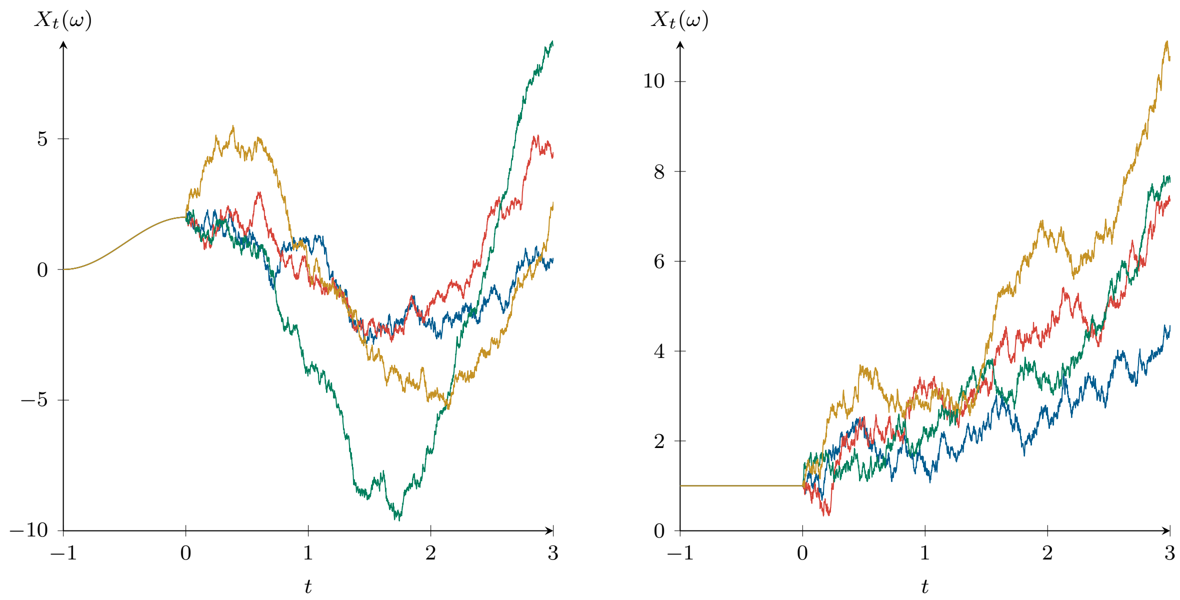
iii) For $n = 2M, 2M+1, \dots, 3M-1$, generate independently $\mathrm{N}(0_{m\times 1}, I_m)$-distributed random variables B_n, $G_{0,n}$, $G_{1,n-M}$, and $G_{2,n-2M}$, and calculate

$$X_{t_{n+1}} = \mathrm{e}^{a_1 \hbar_M}\Bigg(X_{t_n} + a_2 \hbar_M X_{t_{n-M}} + \frac{1}{2} a_2^2 \hbar_M^2 X_{t_{n-2M}}$$
$$+ a_2^3 \int_{t_{n-2M}}^{t_{n-2M+1}} \int_{t_{n-2M}}^{s} \int_{t_{n-2M}}^{u} \mathrm{e}^{-a_1(r - t_{n-2M})} \xi_{r-\tau}\, \mathrm{d}r\, \mathrm{d}u\, \mathrm{d}s$$
$$+ \frac{a_3}{a_1}\big(1 - \mathrm{e}^{-a_1 \hbar_M}\big) + \frac{a_2 a_3}{a_1^2}\big(\mathrm{e}^{-a_1 \hbar_M} - 1 + a_1 \hbar_M\big)$$
$$+ \frac{a_2^2 a_3}{2 a_1^3}\big(2 - 2\mathrm{e}^{-a_1 \hbar_M} - 2 a_1 \hbar_M + a_1^2 \hbar_M^2\big)$$
$$+ \frac{1}{2} a_2^2 \sum_{j=1}^{m} b_3^j \Big((\hbar_M^2 \ell_{2,1} - 2\hbar_M \ell_{3,1} + \ell_{4,1}) B_{n-2M}^j$$
$$+ (\hbar_M^2 \ell_{2,2} - 2\hbar_M \ell_{3,2} + \ell_{4,2}) G_{0,n-2M}^j$$
$$- (2\hbar_M \ell_{3,3} - \ell_{4,3}) G_{1,n-2M}^j + \ell_{4,4} G_{2,n-2M}^j \Big)$$
$$+ a_2 \sum_{j=1}^{m} b_3^j \Big((\hbar_M \ell_{2,1} - \ell_{3,1}) B_{n-M}^j + (\hbar_M \ell_{2,2} - \ell_{3,2}) G_{0,n-M}^j - \ell_{3,3} G_{1,n-M}^j \Big)$$
$$+ \sum_{j=1}^{m} b_3^j \big(\ell_{2,1} B_n^j + \ell_{2,2} G_{0,n}^j \big) \Bigg).$$

In Figure VI.3, we present simulations of Example VI.1 and Example VI.2 that are produced using software The MathWorks, Inc., MATLAB Release 2018b, [102].

In both Figure VI.3 i) and Figure VI.3 ii), we show four realizations of the analytical solution of SDDEs (VI.25) and (VI.31), respectively. For more details on the parameters of the SDDEs, we refer to the captions of the figures.

The larger M is in Example VI.2, the smaller are the entries of the matrix Σ. Due to this, we use the command `vpa` in MATLAB, which uses at least 32 significant digits to evaluate the calculations. Then, we compute matrix L_Σ using command `L = chol(vpa(sigma),'lower');`, where `sigma` $= \Sigma$ and `L` $= L_\Sigma$. In Figure VI.3 ii), we use $M = 2^{10}$ and $\tau = 1$. Hence, we have



i) Realizations of analytical solution X of SDDE (VI.25) with coefficients $a_1 = 0$, $a_2 = -2$, $a_3 = 1$, and $b_3^1 = 2$, and initial condition $\xi_t = 1 + \cos(\pi t)$ for $t \in [-1, 0]$.

ii) Realizations of analytical solution X of SDDE (VI.31) with coefficients $a_1 = -2$, $a_2 = 4$, $a_3 = 1$, and $b_3^1 = \sqrt{2}$, and initial condition $\xi_t = 1$ for $t \in [-1, 0]$.

Figure VI.3. Four realizations of the analytical solution of linear SDDEs with additive noise (VI.25) and (VI.31) are presented. We set $t_0 = 0$, $\tau = 1$, $T = 3$, and $m = 1$, and using Example VI.1 and Example VI.2 with $M = 2^{10}$, we simulate the analytical solutions error-free.

$\hbar_M = 2^{-10}$ and for example obtain $\ell_{4,4} \approx$ `L(4,4)` $= 3.67032\ldots \cdot 10^{-13}$ numerically. Without using the command `vpa`, we received an error from the Cholesky decomposition `chol` that matrix `sigma` must be positive definite. Thus, simulating Example VI.2, one should take these numerical issues into account if M is large.

In order to compare the Milstein scheme with the Euler-Maruyama scheme in numerical simulations, we need analytical solutions of SDDEs that do not just have additive noise. Recall that the Milstein scheme coincides with the Euler-Maruyama scheme if an SDDE has additive noise. So far, to our knowledge, there are no error-free simulations of analytical solutions of SDDEs with more general than additive noise published. Using Itô's formula, we deduce and correctly simulate analytical solutions of more general SDDEs in the following.

Our goal is to find SDDEs

$$Z_t = \begin{cases} \zeta_t & \text{if } t \in [t_0 - \tau, t_0] \text{ and} \\ \zeta_{t_0} + \int_{t_0}^t a(Z_s, Z_{s-\tau})\,\mathrm{d}s + \sum_{j=1}^m \int_{t_0}^t b^j(Z_s)\,\mathrm{d}W_s^j & \text{if } t \in]t_0, T] \end{cases}$$

such that $(f(Z_t))_{t \in [t_0, T]}$ is the unique solution of a linear SDDE with additive noise as in Example VI.1 or Example VI.2.

To begin with, we follow a similar approach as in [46, Section 2.5] for SODEs. Let $f \in C^2(\mathbb{R}; \mathbb{R})$ be a strictly monotone function. We denote by f' and f'' the first and second derivative of f, respectively. Using Itô's formula, see e. g. [64] or [75, p. 153], we obtain

$$f(Z_t) = \begin{cases} f(\zeta_t) & \text{if } t \in [t_0 - \tau, t_0], \\ f(\zeta_{t_0}) + \int_{t_0}^{t} f'(Z_s)a(Z_s, Z_{s-\tau}) + \frac{1}{2} f''(Z_s) \sum_{j=1}^{m} (b^j(Z_s))^2 \, \mathrm{d}s \\ \quad + \sum_{j=1}^{m} \int_{t_0}^{t} f'(Z_s) b^j(Z_s) \, \mathrm{d}W_s^j & \text{if } t \in \,]t_0, T] \end{cases}$$

for all $t \in [t_0 - \tau, T]$ P-almost surely. Since f is strictly monotone, it has an inverse function g such that $f(g(y)) = y$ for all $y \in \operatorname{im} f$ and $g(f(x)) = x$ for all $x \in \operatorname{dom} f$. Moreover, it holds $g \in C^2(\operatorname{im} f; \operatorname{dom} f)$, where

$$g'(x) = \frac{1}{f'(g(x))}$$

and

$$g''(x) = \frac{-f''(g(x))}{(f'(g(x)))^3}$$

for all $x \in \operatorname{im} f$, see e. g. [57, p. 300]. We set $X_t = f(Z_t)$ for all $t \in [t_0 - \tau, T]$, where in particular $\xi_t = f(\zeta_t)$ for all $t \in [t_0 - \tau, t_0]$. Using that g is the inverse function of f, it holds $Z_t = g(X_t)$, and the SDDE above can be rewritten to

$$X_t = \begin{cases} \xi_t \quad \text{if } t \in [t_0 - \tau, t_0], \\ \xi_{t_0} + \int_{t_0}^{t} f'(g(X_s))a(g(X_s), g(X_{s-\tau})) + \frac{1}{2} f''(g(X_s)) \sum_{j=1}^{m} \left(b^j(g(X_s))\right)^2 \mathrm{d}s \\ \quad + \sum_{j=1}^{m} \int_{t_0}^{t} f'(g(X_s)) b^j(g(X_s)) \, \mathrm{d}W_s^j \quad \text{if } t \in \,]t_0, T]. \end{cases}$$

This SDDE equals the SDDEs in Example VI.1 and Example VI.2 if and only if

$$a_1 x + a_2 y + a_3 = f'(g(x))a(g(x), g(y)) + \frac{1}{2} f''(g(x)) \sum_{j=1}^{m} \left(b^j(g(x))\right)^2 \tag{VI.32}$$

and

$$b_3^j = f'(g(x)) b^j(g(x)) \tag{VI.33}$$

for all $x, y \in \operatorname{im} f$ and $j \in \{1, \ldots, m\}$, where $a_1, a_2, a_3, b_3^j \in \mathbb{R}$. Since function f is strictly monotone, it holds $f'(x) \neq 0$ for all $x \in \operatorname{dom} f$, and thus, we obtain from conditions (VI.32) and (VI.33) that

$$b^j(x) = \frac{b_3^j}{f'(x)}$$

and

$$a(x, y) = \frac{a_1 f(x) + a_2 f(y) + a_3 - \frac{1}{2} f''(x) \sum_{j=1}^{m} \left(\frac{b_3^j}{f'(x)}\right)^2}{f'(x)}$$

for all $x, y \in \operatorname{dom} f$. We summarize these considerations in the following example.

Example VI.4
Let X be the solution of the linear SDDE with additive noise

$$X_t = \begin{cases} \xi_t & \text{if } t \in [t_0-\tau, t_0], \\ \xi_{t_0} + \int_{t_0}^{t} a_1 X_s + a_2 X_{s-\tau} + a_3 \,\mathrm{d}s + \sum_{j=1}^{m} \int_{t_0}^{t} b_3^j \,\mathrm{d}W_s^j & \text{if } t \in]t_0, T], \end{cases} \tag{VI.34}$$

where $\xi \in C([t_0-\tau, t_0]; \mathbb{R})$ and $a_1, a_2, a_3, b_3^j \in \mathbb{R}$, $j \in \{1, \dots, m\}$. In addition, let $f \in C^2(\mathbb{R};\mathbb{R})$ be a strictly monotone function and $g \in C^2(\mathbb{R};\mathbb{R})$ its inverse function. Then, $(Z_t)_{t\in[t_0-\tau,T]} = (g(X_t))_{t\in[t_0-\tau,T]}$ is a unique strong solution of the SDDE

$$Z_t = \begin{cases} g(\xi_t) & \text{if } t \in [t_0-\tau, t_0] \text{ and} \\ g(\xi_{t_0}) + \int_{t_0}^{t} \frac{a_1 f(Z_s) + a_2 f(Z_{s-\tau}) + a_3 - \frac{1}{2} f''(Z_s) \sum_{j=1}^{m} \left(\frac{b_3^j}{f'(Z_s)} \right)^2}{f'(Z_s)} \,\mathrm{d}s \\ \qquad + \sum_{j=1}^{m} \int_{t_0}^{t} \frac{b_3^j}{f'(Z_s)} \,\mathrm{d}W_s^j & \text{if } t \in]t_0, T]. \end{cases} \tag{VI.35}$$

Example VI.4 allows us to simulate the analytical solution of the SDDE with more general noise, see equation (VI.35), error-free as follows. First, we simulate the analytical solution of linear SDDE with additive noise (VI.34), see Example VI.1 and Example VI.2. Then, we set $Z = g(X)$ to obtain the solution of SDDE (VI.35).

Let us provide an example. Set $f(x) = \ln(x^2)$ for $x \in \mathbb{R}$. Then, $Z = g(X)$ with $g(x) = \sqrt{\mathrm{e}^x}$, $x \in \mathbb{R}$, is the unique strong solution of SDDE

$$Z_t = \begin{cases} \sqrt{\mathrm{e}^{\xi_t}} & \text{if } t \in [t_0-\tau, t_0] \text{ and} \\ \sqrt{\mathrm{e}^{\xi_{t_0}}} + \int_{t_0}^{t} \frac{1}{2}\big(a_1 \ln(Z_s^2) + a_2 \ln(Z_{s-\tau}^2) + a_3\big) Z_s + \frac{1}{8} \sum_{j=1}^{m} (b_3^j)^2 Z_s \,\mathrm{d}s \\ \qquad + \sum_{j=1}^{m} \int_{t_0}^{t} \frac{1}{2} b_3^j Z_s \,\mathrm{d}W_s^j & \text{if } t \in]t_0, T], \end{cases} \tag{VI.36}$$

which has multiplicative noise.

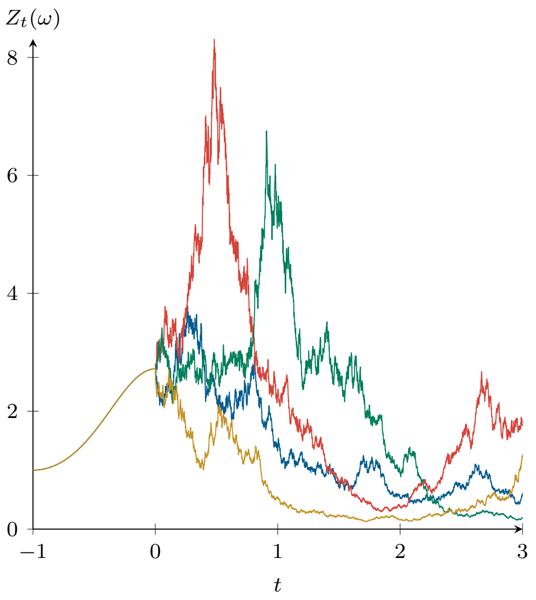
In Figure VI.5, we present some simulations of the analytical solution Z.

In order to find more strictly monotone functions $f \in C^2(\mathbb{R};\mathbb{R})$, we can use the Lamperti transformation, see [86, Theorem 2] and [46, p. 34].

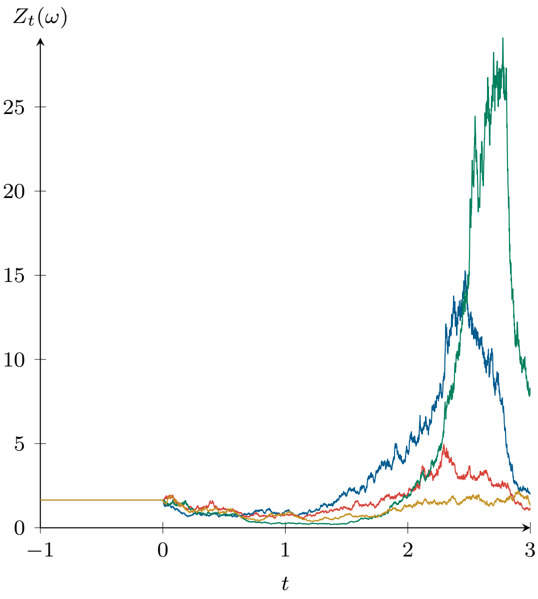
Example VI.6
Let $\rho \in C^1(\mathbb{R};\mathbb{R})$ such that $|\rho(x)| > 0$ for all $x \in \mathbb{R}$. We set

$$f(x) = \int_0^x \frac{1}{\rho(y)} \,\mathrm{d}y$$

for all $x \in \mathbb{R}$. Then, $f \in C^2(\mathbb{R},\mathbb{R})$ is a strictly monotone function, and we have $f'(x) = \frac{1}{\rho(x)}$ and $f''(x) = -\frac{\rho'(x)}{\rho^2(x)}$ for all $x \in \mathbb{R}$.



i) Realizations of analytical solution Z of SDDE (VI.36) with coefficients $a_1 = 0$, $a_2 = -1$, $a_3 = \frac{1}{100}$, and $b_3^1 = \sqrt{2}$, and initial condition $\xi_t = 1 + \cos(\pi t)$ for $t \in [-1, 0]$.



ii) Realizations of analytical solution Z of SDDE (VI.36) with coefficients $a_1 = -\frac{1}{2}$, $a_2 = -3$, $a_3 = 1$, and $b_3^1 = \sqrt{2}$, and initial condition $\xi_t = 1$ for $t \in [-1, 0]$.

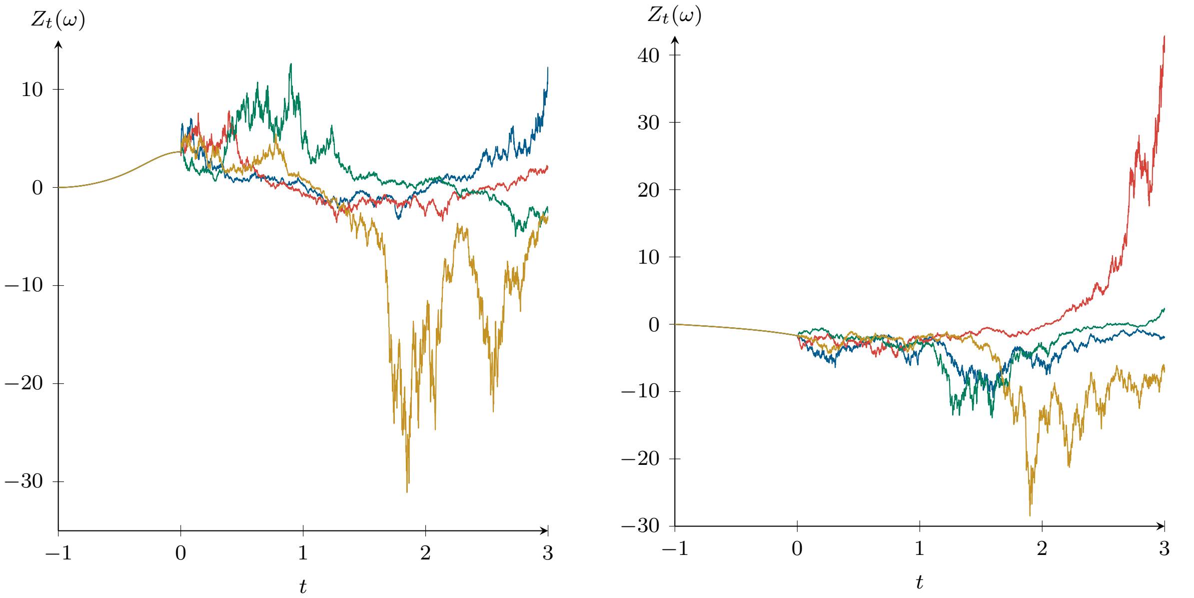
Figure VI.5. Four realizations of the analytical solution Z of SDDE with multiplicative noise (VI.36) are presented. We set $t_0 = 0$, $\tau = 1$, $T = 3$, and $m = 1$, and using Example VI.1 and Example VI.2 with $M = 2^{10}$, we simulate analytical solution X of linear SDDE with additive noise (VI.34) error-free. Then, using Example VI.4, we obtain with $Z = g(X)$, where $g(x) = \sqrt{e^x}$ for $x \in \mathbb{R}$, the analytical solution Z.

Setting e. g. $\rho(x) = \sqrt{x^2 + 1}$, $x \in \mathbb{R}$, in Example VI.6, we obtain $f = \operatorname{arcsinh}$ and $g = \sinh$. Then, SDDE (VI.35) reads as

$$Z_t = \begin{cases} \sinh(\xi_t) & \text{if } t \in [t_0 - \tau, t_0] \text{ and} \\ \sinh(\xi_{t_0}) + \displaystyle\int_{t_0}^{t} \big(a_1 \operatorname{arcsinh}(Z_s) + a_2 \operatorname{arcsinh}(Z_{s-\tau}) + a_3\big)\sqrt{Z_s^2 + 1} + \frac{1}{2}\sum_{j=1}^{m} (b_3^j)^2 Z_s \,\mathrm{d}s & \\ \qquad + \displaystyle\sum_{j=1}^{m} \int_{t_0}^{t} b_3^j \sqrt{Z_s^2 + 1}\,\mathrm{d}W_s^j & \text{if } t \in \,]t_0, T]. \end{cases} \tag{VI.37}$$

Simulations of analytical solution Z are presented in Figure VI.7.

We remark that the drift coefficients of SDDEs (VI.36) and (VI.37) are not Lipschitz continuous. In order to conserve Lipschitz continuous coefficients, we can choose function ρ in Example VI.6 to be bounded and assume that its first derivative ρ' is Lipschitz continuous. Of course constant functions fulfill these conditions for example. But then, SDDE (VI.35) is still linear and has additive noise. Function ρ with $\rho(x) = 2 + \arctan(x)$ for $x \in \mathbb{R}$ provides another example of a bounded and continuous function with Lipschitz continuous derivative. Then, we have $f(x) = 2x + x\arctan(x) - \frac{1}{2}\ln(1 + x^2)$ for $x \in \mathbb{R}$. However, the inverse function g of f is not known to us so that the analytical solution $Z = g(X)$ cannot be calculated. Nevertheless, Example VI.6 provides analytical solutions of SDDEs with more general than additive noise like SDDEs (VI.36) and (VI.37), which can be used for numerical tests of the Milstein scheme.



i) Realizations of analytical solution Z of SDDE (VI.37) where $m = 3$ with coefficients $a_1 = 0$, $a_2 = -\frac{3}{2}$, $a_3 = -\frac{1}{8}$, $b_3^1 = \frac{1}{\sqrt{3}}$, $b_3^2 = -\sqrt{2}$, and $b_3^3 = -\frac{1}{4}$, and initial condition $\xi_t = 1 + \cos(\pi t)$ for $t \in [-1, 0]$.

ii) Realizations of analytical solution Z of SDDE (VI.37) where $m = 2$ with coefficients $a_1 = \frac{1}{2}$, $a_2 = -\frac{1}{\sqrt{2}}$, $a_3 = 0$, $b_3^1 = \frac{2}{3}$, and $b_3^2 = 1$, and initial condition $\xi_t = 2 - 2\mathrm{e}^{\frac{1}{2}(t+1)}$ for $t \in [-1, 0]$.

Figure VI.7. Four realizations of the analytical solution Z of SDDE with commutative noise (VI.37) are presented. We set $t_0 = 0$, $\tau = 1$, and $T = 3$, and using Example VI.1 and Example VI.2 with $M = 2^{10}$, we simulate analytical solution X of linear SDDE with additive noise (VI.34) error-free. Then, using Example VI.4, we obtain with $Z = g(X)$, where $g(x) = \sinh(x)$ for $x \in \mathbb{R}$, the analytical solution Z.

VI.2. Numerical Examples

In order to illustrate and confirm our theoretical results on the strong and pathwise convergence of Euler-Maruyama and the Milstein scheme from Chapter IV, we provide some numerical examples in this section. Using the analytical solutions of SDDEs derived in the previous section, we are able to calculate and compare the errors made by the Euler-Maruyama and Milstein approximation. For the sake of clarity and better comparability, the illustrations of the simulation studies are postponed to the end of this section.

To begin with, we consider the linear SDDE with additive noise from equation (VI.34). As the derivatives of the diffusion coefficients vanish, the Milstein scheme coincides with Euler-Maruyama scheme.

In Figure VI.8 and Figure VI.9, we present simulation studies of the Euler-Maruyama approximations of the linear SDDEs with multidimensional additive noise from Example VI.1 and Example VI.2 where $a_1 = 0$ and $a_1 = 1$, respectively. For more details on the parameters of the SDDEs and the simulation studies, we refer to the captions of Figure VI.8 and Figure VI.9.

Figure VI.8 i) and Figure VI.9 i) show the empirical error of the Euler-Maruyama scheme Y^h versus step size h in the strong sense for $p \in \{2, 7, 50\}$. As both axes are scaled logarithmically in the figures, the slopes of the graphs equal the empirical strong orders of convergence. We see that the slopes of the graphs are approximately $\alpha \approx 1$ as expected. This confirms our theoretical result from Corollary IV.13. Further, we obtain for larger p a larger empirical error in the strong sense. This is consistent with our theoretical error estimates, cf. the proof of Theorem IV.9.

Considering the empirical pathwise error in Figure VI.8 iii) and Figure VI.9 iii), we see that the empirical order of convergence is approximately $\alpha \approx 1$, too. This confirms our theoretical result as well. In Corollary IV.14, we proved that the Euler-Maruyama scheme is pathwise convergent with order $\alpha = 1 - \varepsilon$ for arbitrary $\varepsilon > 0$ in case of additive noise. Whereas Figure VI.8 iii) and Figure VI.9 iii) display only four realizations, the histograms in Figure VI.8 iv) and Figure VI.9 iv) present the relative frequency of the pathwise error of the Euler-Maruyama approximation with step size $h = 2^{-16}$ and $h = 2^{-12}$, respectively, over 10^3 realizations.

Further, Figure VI.8 ii) and Figure VI.9 ii) indicate the convergence of the Euler-Maruyama scheme for step sizes $h = 2^{-i}$ where $i \in \{0, 1, \dots, 5\}$. According to the continuous formulation of the scheme and due to the evaluation of the analytical solution and Euler-Maruyama approximations on the fine grid $\mathcal{I} = \{-1 + n \cdot 2^{-16} : n \in \{0, 1, \dots, 4 \cdot 2^{16}\}\}$ and $\mathcal{I} = \{-1 + n \cdot 2^{-12} : n \in \{0, 1, \dots, 4 \cdot 2^{12}\}\}$, respectively, we see that the approximations approach the movement of the trajectory of the solution between the grid points obtained by step size h of the Euler-Maruyama method.

In the following, we consider SDDEs with commutative noise where the Euler-Maruyama scheme and the Milstein scheme do not coincide. To the best of our knowledge, the numerical examples below are the first that compare the numerical approximations to the exactly simulated analytical solutions of SDDEs.

If the diffusion coefficients of an SDDE do not depend on the past history of the solution and satisfy the commutativity condition (V.1), the Milstein scheme (IV.33) simplifies by equations (V.2) and (V.3) to

$$Y_t = \begin{cases} \xi_t \quad \text{for } t \in [t_0 - \tau, t_0] \text{ and} \\ Y_{t_n} + \Big(a(t_n, t_n - \tau_1, \dots, t_n - \tau_D, Y_{t_n}, Y_{t_n - \tau_1}, \dots, Y_{t_n - \tau_D}) \\ \qquad - \frac{1}{2} \sum_{j=1}^{m} \sum_{i=1}^{d} \partial_{x_0^i} b^j(t_n, t_n - \tau_1, \dots, t_n - \tau_D, Y_{t_n}) \\ \qquad\qquad \times b^{i,j}(t_n, t_n - \tau_1, \dots, t_n - \tau_D, Y_{t_n}) \Big)(t - t_n) \\ \quad + \sum_{j=1}^{m} b^j(t_n, t_n - \tau_1, \dots, t_n - \tau_D, Y_{t_n}) \big(W_t^j - W_{t_n}^j\big) \\ \quad + \frac{1}{2} \sum_{j_1, j_2 = 1}^{m} \sum_{i=1}^{d} \partial_{x_0^i} b^{j_1}(t_n, t_n - \tau_1, \dots, t_n - \tau_D, Y_{t_n}) \\ \qquad \times b^{i,j_2}(t_n, t_n - \tau_1, \dots, t_n - \tau_D, Y_{t_n}) \big(W_t^{j_1} - W_{t_n}^j\big)\big(W_t^{j_2} - W_{t_n}^j\big) \\ \quad \text{for } t \in \,]t_n, t_{n+1}] \text{ where } n = 0, 1, \dots, N. \end{cases} \tag{VI.38}$$

Thus, we only have to generate the increments of the Wiener process like with the Euler-Maruyama scheme in order to simulate the Milstein approximations.

In figures VI.10, VI.11, VI.12, and VI.13 below, we present simulation studies of the Euler-Maruyama scheme and the Milstein scheme of SDDEs (VI.36) and (VI.37) in case of $m = 1$. In the simulation study in Figure VI.14, we consider the SDDE

$$Z_t = \begin{cases} \arctan(\xi_t) & \text{if } t \in [t_0 - \tau, t_0] \text{ and} \\ \arctan(\xi_{t_0}) + \displaystyle\int_{t_0}^{t} \Big(\big(a_1 \tan(Z_s) + a_2 \tan(Z_{s-\tau}) + a_3\big) \cos^2(Z_s) \\ \qquad - \displaystyle\sum_{j=1}^{m} (b_3^j)^2 \sin(Z_s) \cos^3(Z_s) \Big) \, \mathrm{d}s \\ \qquad + \displaystyle\sum_{j=1}^{m} \int_{t_0}^{t} b_3^j \cos^2(Z_s) \, \mathrm{d}W_s^j & \text{if } t \in \,]t_0, T], \end{cases} \tag{VI.39}$$

where the dimension of the Wiener process is $m = 10$.

The analytical solutions of underlying SDDEs are simulated using Example VI.4 as well as Example VI.1 in figures VI.10, VI.12, and VI.14 where $a_1 = 0$, and Example VI.2 in Figure VI.11 and Figure VI.13 where $a_1 = -2$. We generated 10^3 realizations where $h = 2^{-i}$ with $i \in \{0, 1, \ldots, 16\}$ in case of $a_1 = 0$ and 10^4 realizations where $h = 2^{-i}$ with $i \in \{0, 1, \ldots, 12\}$ in case of $a_1 = -2$. For more details on the parameters of the SDDEs and the simulation studies, we refer to the captions of the figures.

We remark that SDDEs (VI.36), (VI.37), and (VI.39) do not fulfill the assumptions of Theorem IV.6 and Theorem IV.9, nor of their corollaries on the convergence of the numerical schemes. However, we are able to simulate the analytical solution exactly. This makes these SDDEs valuable. As we will see below, they nevertheless confirm and illustrate our theoretical results, although, we supposed stronger conditions on the underlying SDDEs in the theorems and corollaries.

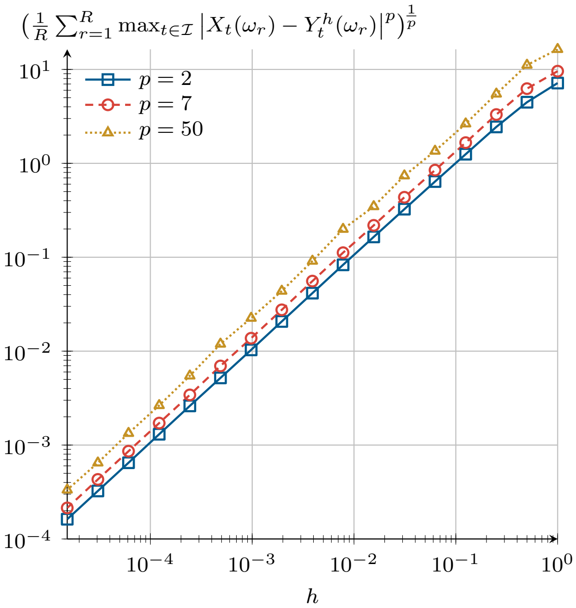
In figures VI.10 i), VI.11 i), VI.12 i), VI.13 i), and VI.14 i) with logarithmically scaled axes, the empirical error in the strong sense for $p \in \{2, 7\}$ of the Euler-Maruyama and the Milstein scheme versus step size h is presented. Here, we see that the empirical strong order of convergence of the Euler-Maruyama scheme is approximately $\alpha \approx \frac{1}{2}$ whereas the Milstein scheme converges approximately with order $\alpha \approx 1$ in the strong sense. Moreover, for larger p, a larger empirical error in the strong sense is obtained. This confirms our theoretical results in Theorem IV.6 and Theorem IV.9, and this is consistent with our theoretical error estimates, cf. inequality (IV.58) and the proof of Theorem IV.9.

The four realizations of the empirical pathwise error of the Euler-Maruyama and the Milstein scheme in each of the figures VI.10 iii), VI.11 iii), VI.12 iii), VI.13 iii), and VI.14 iii) confirm our theoretical results in Corollary IV.7 and Corollary IV.12 as well. The empirical pathwise order of convergence is approximately $\alpha \approx \frac{1}{2}$ for the Euler-Maruyama scheme and approximately $\alpha \approx 1$ for the Milstein scheme.

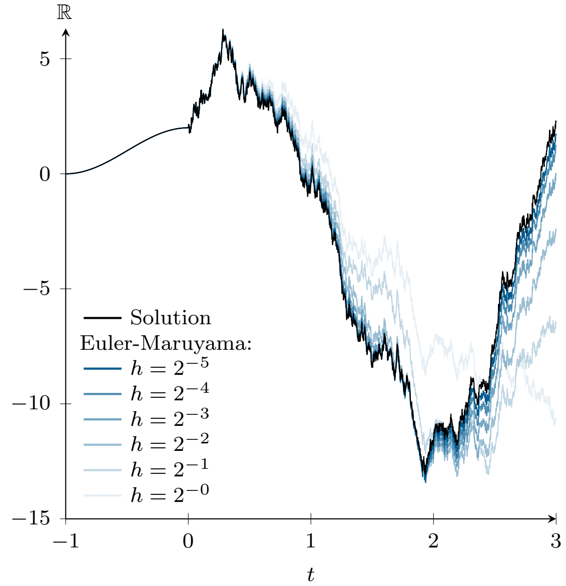
The histograms show the relative frequency of the pathwise error of the Euler-Maruyama scheme (blue) and the Milstein scheme (red) with step size $h = 2^{-16}$ over 10^3 realizations, see figures VI.10 iv), VI.12 iv), and VI.14 iv), and with step size $h = 2^{-12}$ over 10^4 realizations, see Figure VI.11 iv) and Figure VI.13 iv). The abscissa is logarithmically scaled and thus corresponds to the ordinate of the empirical pathwise error plots in figures VI.10 iii), VI.11 iii), VI.12 iii), VI.13 iii), and VI.14 iii). Since larger values have a higher contribution to the $L^p(\Omega;\mathbb{R})$-norm, we see from the histograms in figures VI.10 iv), VI.11 iv), VI.12 iv), VI.13 iv), and VI.14 iv) that there are many realizations, which have a pathwise error that is much smaller than the empirical error in the $S^p([-1,3]\times\Omega;\mathbb{R})$-norm, cf. the sub-figures i) and iii) in figures VI.10, VI.11, VI.12, VI.13, and VI.14. In this context, also note the skewness of the distribution of the relative frequency of the pathwise error, in particular for the Milstein scheme. In Figure VI.11 iii) and Figure VI.13 iii), we especially see that the pathwise error of some realizations of the Milstein approximation can be approximately as large as or even larger than the pathwise error of realizations of the Euler-Maruyama approximation. However, if we only consider single realizations and compare the pathwise error of the Euler-Maruyama with corresponding pathwise error of the Milstein scheme on the same realization, the empirical pathwise error of the Milstein scheme is asymptotically smaller than the one of the Euler-Maruyama scheme. See the yellow lines in Figure VI.11 iii) as well as the blue and green lines in Figure VI.13 iii).

In figures VI.10 ii), VI.11 ii), VI.12 ii), VI.13 ii), and VI.14 ii), the Euler-Maruyama approximation (blue) and the Milstein approximation (red) with step size $h = 2^{-4}$ of a single trajectory are presented together with the exactly simulated analytical solution (black). Here, we see that the Milstein approximation is most of the time closer to the exact solution than the Euler-Maruyama approximation. Especially, the Milstein scheme performs better when the solution quickly changes over time.

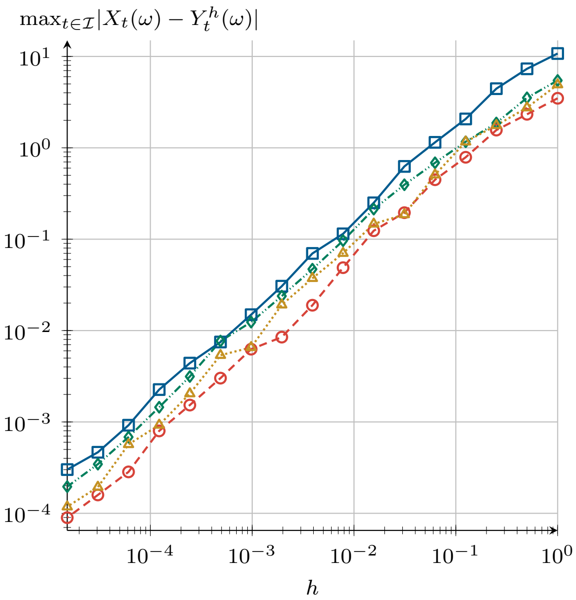
In these simulation studies, our theoretical results are confirmed, and we see that the Milstein scheme outperforms the Euler-Maruyama scheme. Especially, the strong and pathwise orders of convergence can excellently be seen. We remark that the derivative of the drift coefficient with respect to the delay argument does not vanish and is not constant in the examples above. Hence, remainder term R_5^l in the proof of Theorem IV.9, see inequality (IV.145), does not vanish and converges to zero in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ with order $\alpha = 1$ as $h \to 0$. Thus, the derivative of the drift coefficient does not have to be incorporated in the numerical scheme in order to obtain a strong convergence of order $\alpha = 1$ as proven, and the Taylor expansions presented in [104, 124] are not optimal.



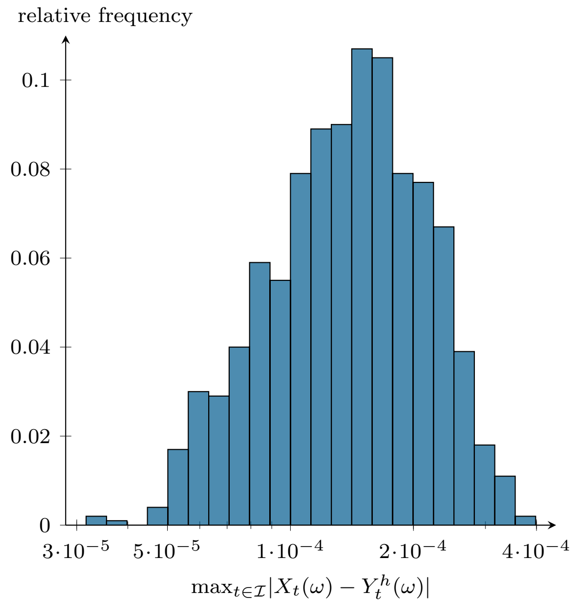
i) Empirical error of the Euler-Maruyama approximation Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \dots, 16\}$, in the strong sense ($S^p([-1, 3] \times \Omega; \mathbb{R})$-norm) for $p \in \{2, 7, 50\}$ over $R = 10^3$ realizations. The scales of both axes are logarithmic.



ii) Trajectories of one realization of the analytical solution and its Euler-Maruyama approximations with respect to step sizes $h = 2^{-i}$, $i \in \{0, 1, \dots, 5\}$.

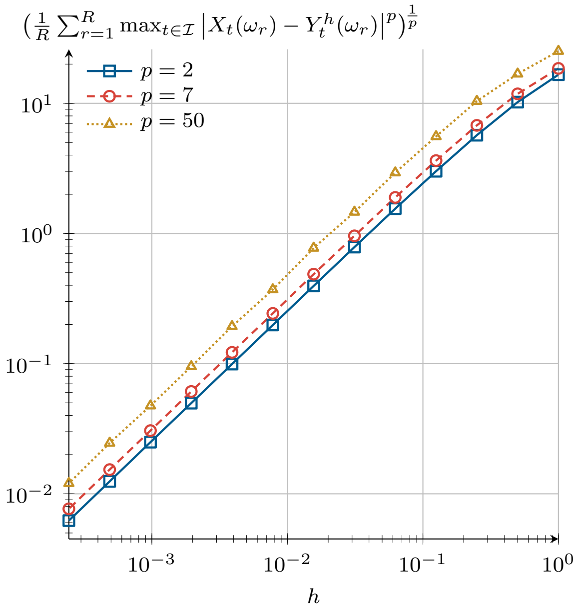


iii) Empirical error of the Euler-Maruyama approximation Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \dots, 16\}$, in the pathwise sense for four realizations. The scales of both axes are logarithmic.

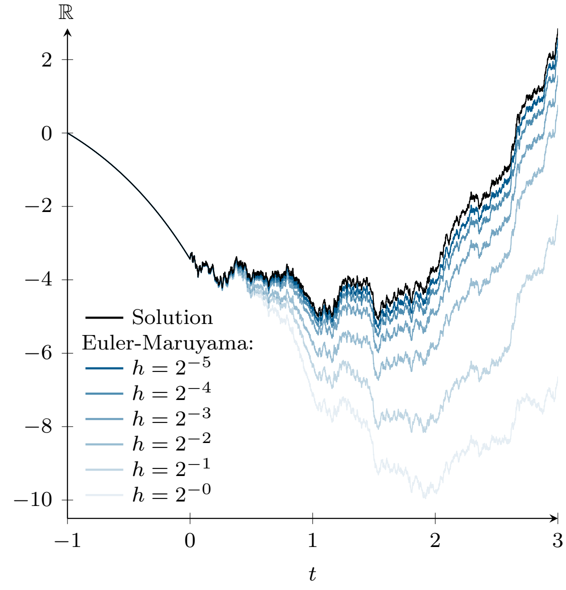


iv) Histogram of the empirical pathwise error over 10^3 realizations of the Euler-Maruyama approximation Y^h with step size $h = 2^{-16}$. The abscissa is logarithmically scaled.

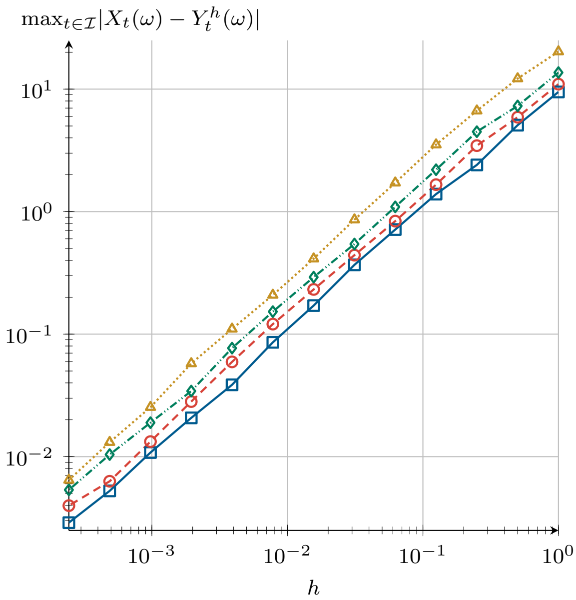
Figure VI.8. Simulation study of the Euler-Maruyama scheme and linear SDDE (VI.25) with additive noise. We set $t_0 = 0$, $\tau = 1$, $T = 3$, $a_1 = 0$, $a_2 = -2$, $a_3 = 1$, $m = 5$, $b_3^1 = -\frac{1}{2}$, $b_3^2 = 1$, $b_3^3 = 2$, $b_3^4 = -2$, $b_3^5 = 1$, and $\xi_t = 1 + \cos(\pi t)$ for $t \in [-1, 0]$. We simulated analytical solution X of SDDE (VI.25) error-free using Example VI.1 where $M = 2^{16}$ and thus $\hbar_M = 2^{-16}$. The analytical solution and the numerical approximations are evaluated at the points in time $\mathcal{I} = \{-1 + n\,\hbar_M : n \in \{0, 1, \dots, 4M\}\}$.



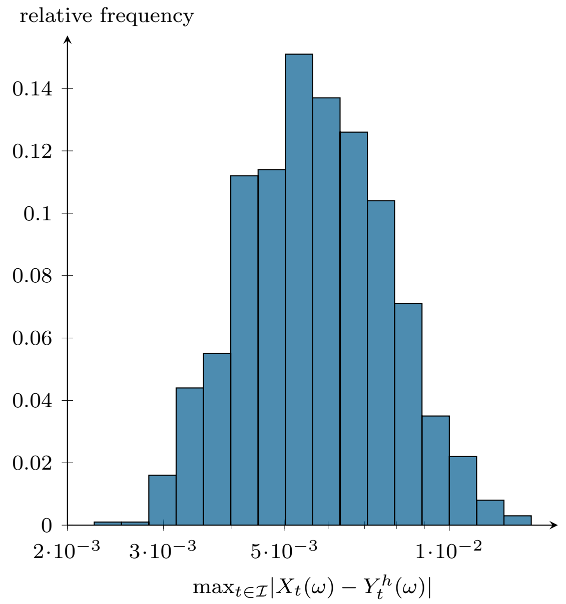
i) Empirical error of the Euler-Maruyama approximation Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 12\}$, in the strong sense ($S^p([-1,3] \times \Omega; \mathbb{R})$-norm) for $p \in \{2, 7, 50\}$ over $R = 10^3$ realizations. The scales of both axes are logarithmic.



ii) Trajectories of one realization of the analytical solution and its Euler-Maruyama approximations with respect to step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 5\}$.

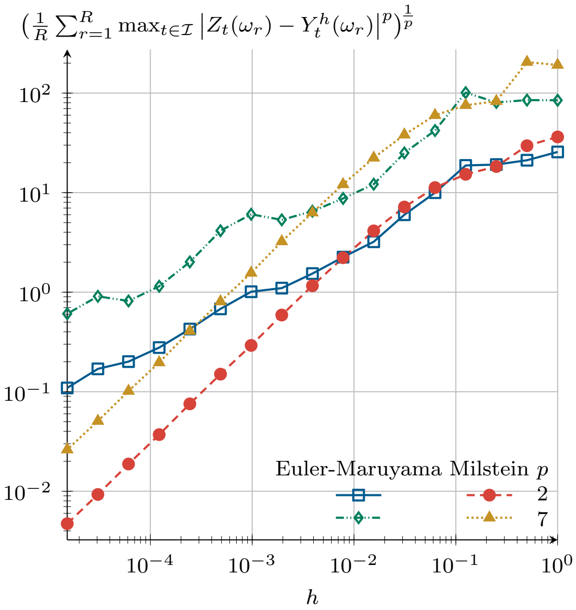


iii) Empirical error of the Euler-Maruyama approximation Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 12\}$, in the pathwise sense for four realizations. The scales of both axes are logarithmic.

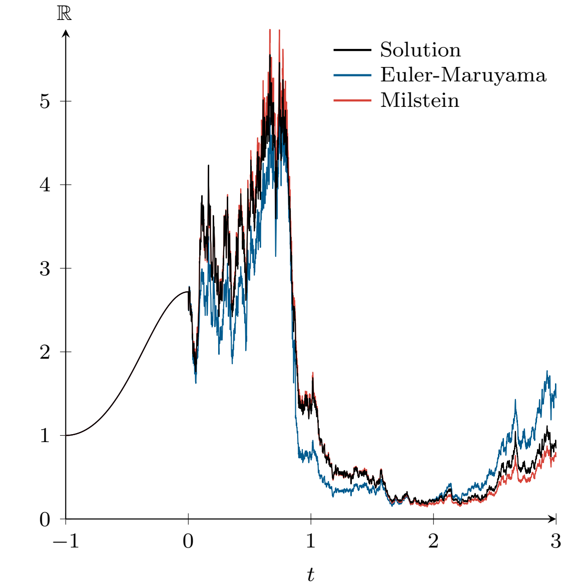


iv) Histogram of the empirical pathwise error over 10^3 realizations of the Euler-Maruyama approximation Y^h with step size $h = 2^{-12}$. The abscissa is logarithmically scaled.

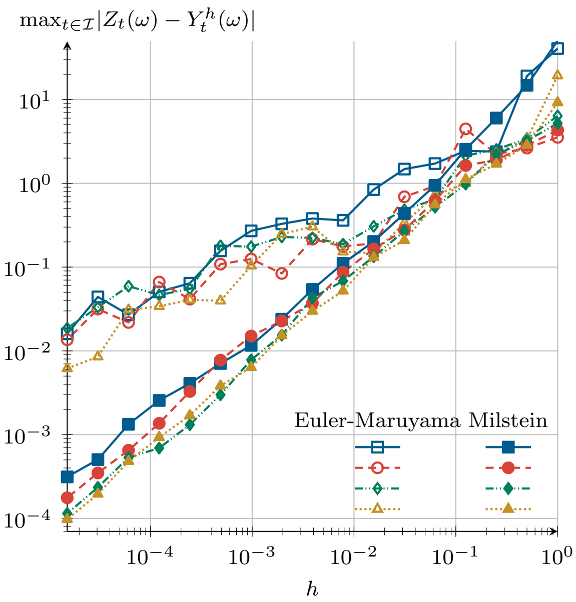
Figure VI.9. Simulation study of the Euler-Maruyama scheme and linear SDDE (VI.31) with additive noise. We set $t_0 = 0$, $\tau = 1$, $T = 3$, $a_1 = 1$, $a_2 = -2$, $a_3 = -\frac{1}{2}$, $m = 2$, $b_3^1 = \frac{2}{3}$, $b_3^2 = 1$, and $\xi_t = 2 - 2e^{t+1}$ for $t \in [-1, 0]$. We simulated analytical solution X of SDDE (VI.31) error-free using Example VI.2 where $M = 2^{12}$ and thus $\hbar_M = 2^{-12}$. The analytical solution and the numerical approximations are evaluated at the points in time $\mathcal{I} = \{ -1 + n\,\hbar_M : n \in \{0, 1, \ldots, 4M\}\}$.



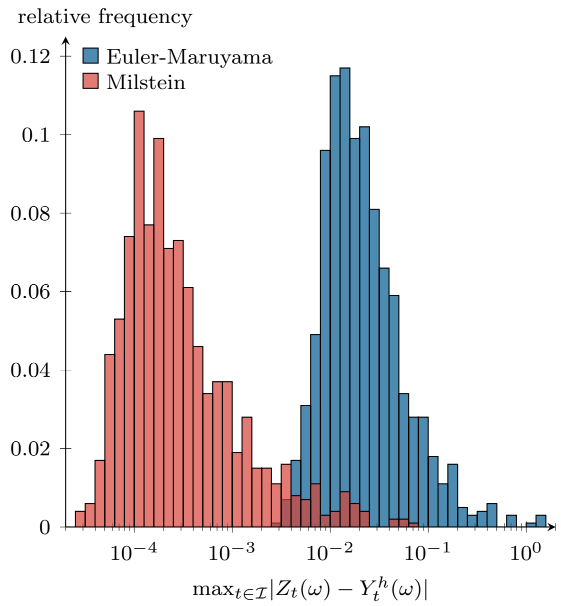
i) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 16\}$, in the strong sense ($S^p([-1,3] \times \Omega; \mathbb{R})$-norm) for $p \in \{2, 7\}$ over $R = 10^3$ realizations. The scales of both axes are logarithmic.



ii) Trajectories of one realization of the analytical solution, and its Euler-Maruyama and Milstein approximations with respect to step size $h = 2^{-4}$.

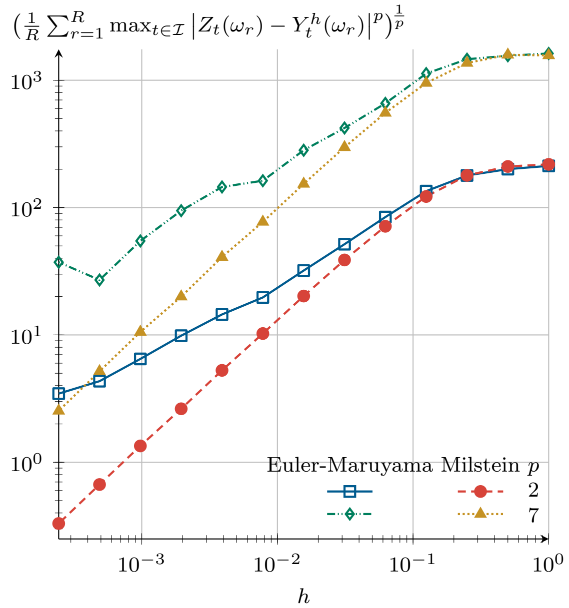


iii) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 16\}$, in the pathwise sense for four realizations. The scales of both axes are logarithmic.

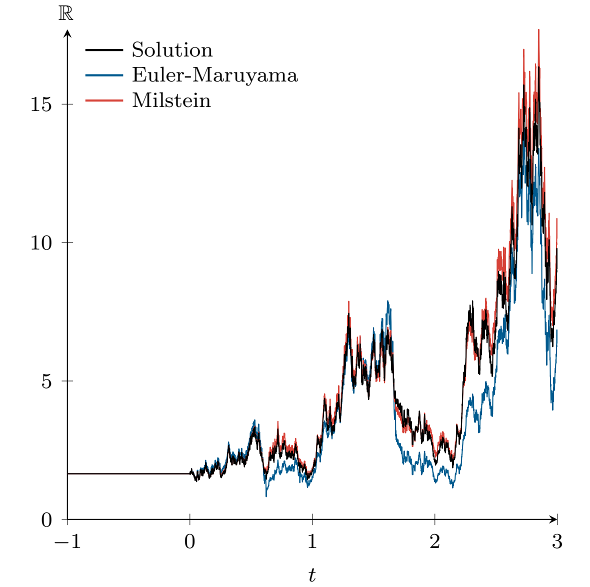


iv) Histogram of the empirical pathwise error over 10^3 realizations of the Euler-Maruyama and the Milstein approximations Y^h with step size $h = 2^{-16}$. The abscissa is logarithmically scaled.

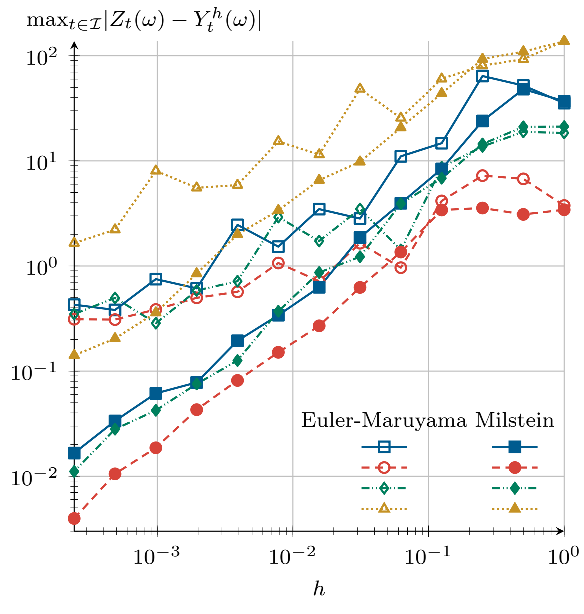
Figure VI.10. Simulation study of the Euler-Maruyama and Milstein scheme regarding SDDE (VI.36) with commutative noise. We set $t_0 = 0$, $\tau = 1$, $T = 3$, $a_1 = 0$, $a_2 = -2$, $a_3 = 1$, $m = 1$, $b_3^1 = 2$, and $\xi_t = 1 + \cos(\pi t)$ for $t \in [-1, 0]$. Let X be the analytical solution of linear SDDE (VI.25) with additive noise and $g(x) = \sqrt{\mathrm{e}^x}$ for $x \in \mathbb{R}$. We simulated the analytical solution $Z = g(X)$ of SDDE (VI.36) error-free using Example VI.1 and Example VI.4 where $M = 2^{16}$ and thus $\hbar_M = 2^{-16}$. The analytical solution and the numerical approximations are evaluated at the points in time $\mathcal{I} = \{-1 + n\,\hbar_M : n \in \{0, 1, \ldots, 4M\}\}$.



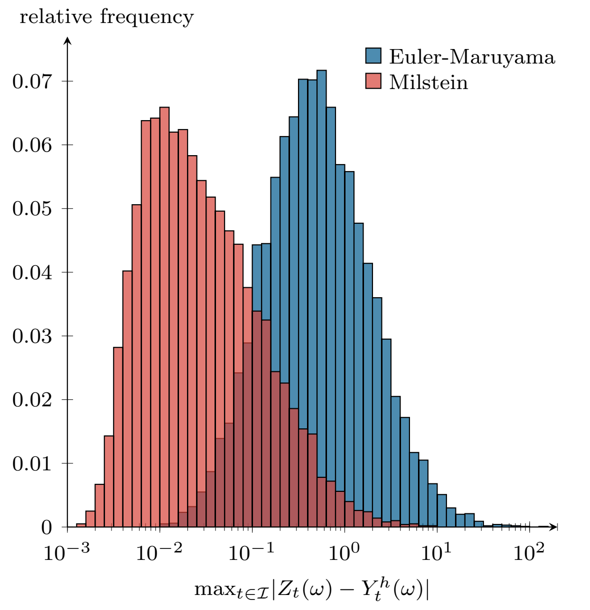
i) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 12\}$, in the strong sense ($S^p([-1, 3] \times \Omega; \mathbb{R})$-norm) for $p \in \{2, 7\}$ over $R = 10^4$ realizations. The scales of both axes are logarithmic.



ii) Trajectories of one realization of the analytical solution, and its Euler-Maruyama and Milstein approximations with respect to step size $h = 2^{-4}$.

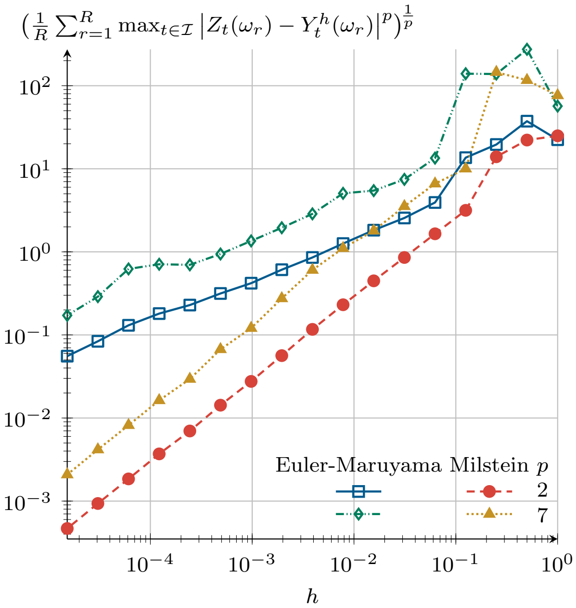


iii) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 12\}$, in the pathwise sense for four realizations. The scales of both axes are logarithmic.

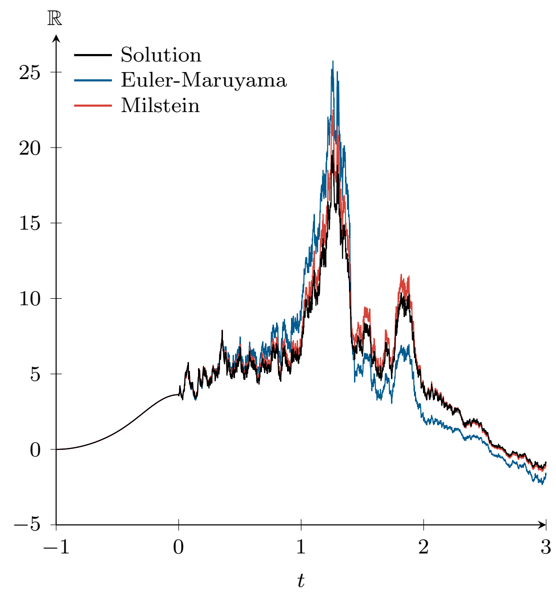


iv) Histogram of the empirical pathwise error over 10^4 realizations of the Euler-Maruyama and the Milstein approximations Y^h with step size $h = 2^{-12}$. The abscissa is logarithmically scaled.

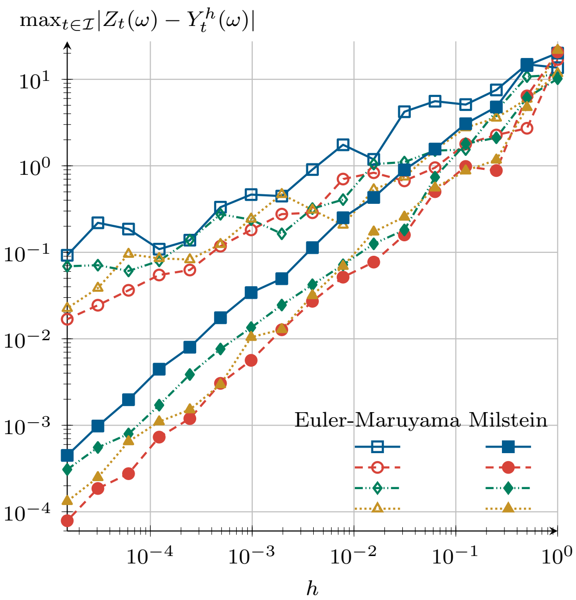
Figure VI.11. Simulation study of the Euler-Maruyama and Milstein scheme regarding SDDE (VI.36) with commutative noise. We set $t_0 = 0$, $\tau = 1$, $T = 3$, $a_1 = -2$, $a_2 = 4$, $a_3 = 1$, $m = 1$, $b_3^1 = 2$, and $\xi_t = 1$ for $t \in [-1, 0]$. Let X be the analytical solution of linear SDDE (VI.31) with additive noise and $g(x) = \sqrt{e^x}$ for $x \in \mathbb{R}$. We simulated the analytical solution $Z = g(X)$ of SDDE (VI.36) error-free using Example VI.2 and Example VI.4 where $M = 2^{12}$ and thus $\hbar_M = 2^{-12}$. The analytical solution and the numerical approximations are evaluated at the points in time $\mathcal{I} = \{-1 + n\,\hbar_M : n \in \{0, 1, \ldots, 4M\}\}$.



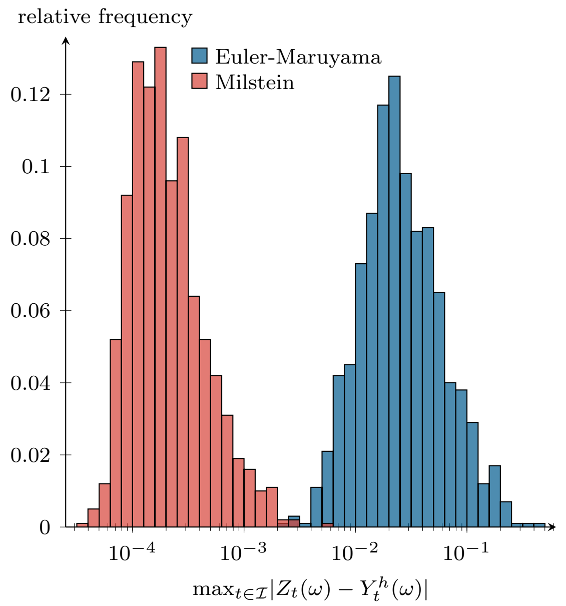
i) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \dots, 16\}$, in the strong sense ($S^p([-1,3] \times \Omega; \mathbb{R})$-norm) for $p \in \{2, 7\}$ over $R = 10^3$ realizations. The scales of both axes are logarithmic.



ii) Trajectories of one realization of the analytical solution, and its Euler-Maruyama and Milstein approximations with respect to step size $h = 2^{-4}$.

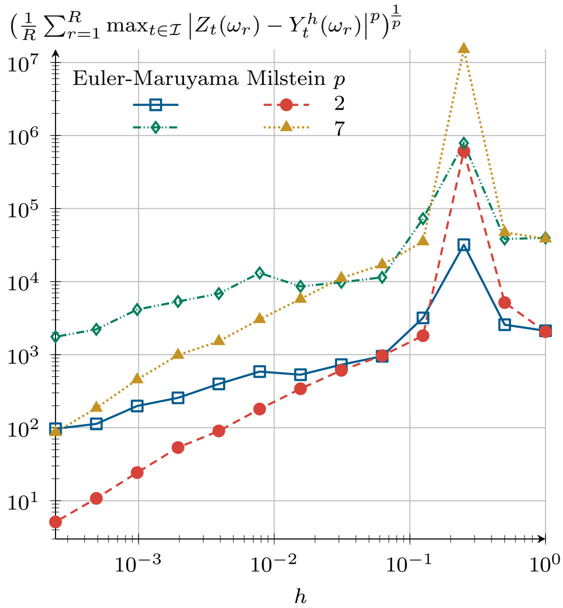


iii) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \dots, 16\}$, in the pathwise sense for four realizations. The scales of both axes are logarithmic.

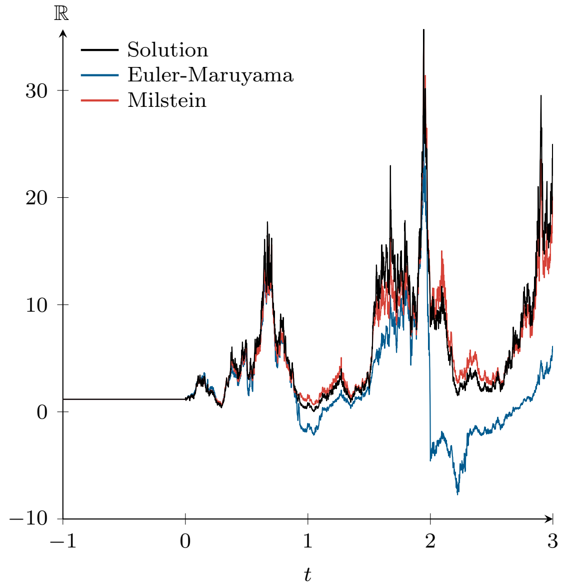


iv) Histogram of the empirical pathwise error over 10^3 realizations of the Euler-Maruyama and the Milstein approximations Y^h with step size $h = 2^{-16}$. The abscissa is logarithmically scaled.

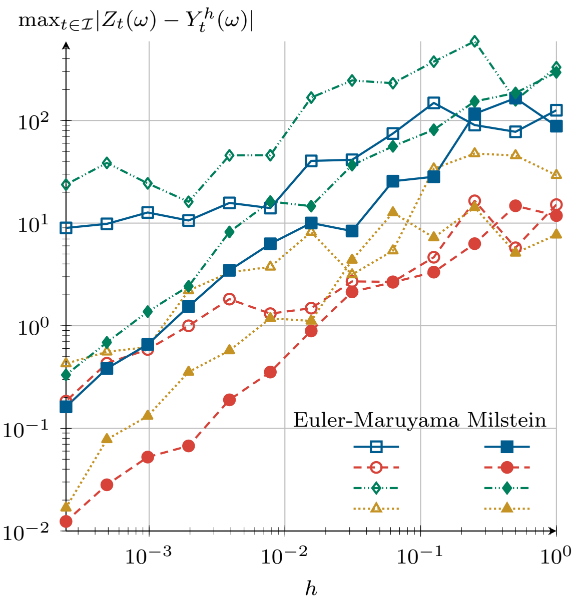
Figure VI.12. Simulation study of the Euler-Maruyama and Milstein scheme regarding SDDE (VI.37) with commutative noise. We set $t_0 = 0$, $\tau = 1$, $T = 3$, $a_1 = 0$, $a_2 = -1$, $a_3 = 1$, $m = 1$, $b_3^1 = -1$, and $\xi_t = 1 + \cos(\pi t)$ for $t \in [-1, 0]$. Let X be the analytical solution of linear SDDE (VI.25) with additive noise and $g(x) = \sinh(x)$ for $x \in \mathbb{R}$. We simulated the analytical solution $Z = g(X)$ of SDDE (VI.37) error-free using Example VI.1 and Example VI.4 where $M = 2^{16}$ and thus $\hbar_M = 2^{-16}$. The analytical solution and the numerical approximations are evaluated at the points in time $\mathcal{I} = \{-1 + n\,\hbar_M : n \in \{0, 1, \dots, 4M\}\}$.



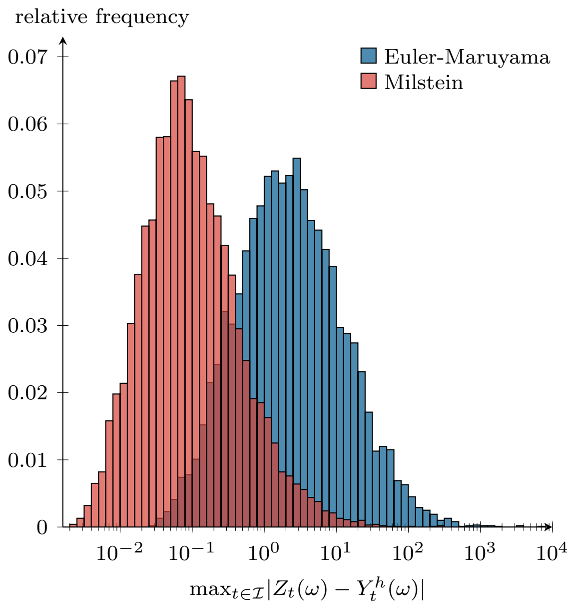
i) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 12\}$, in the strong sense ($S^p([-1, 3] \times \Omega; \mathbb{R})$-norm) for $p \in \{2, 7\}$ over $R = 10^4$ realizations. The scales of both axes are logarithmic.



ii) Trajectories of one realization of the analytical solution, and its Euler-Maruyama and Milstein approximations with respect to step size $h = 2^{-4}$.

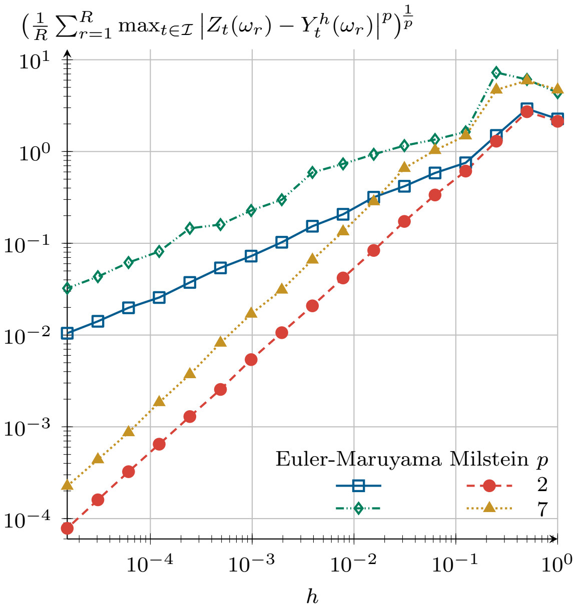


iii) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 12\}$, in the pathwise sense for four realizations. The scales of both axes are logarithmic.

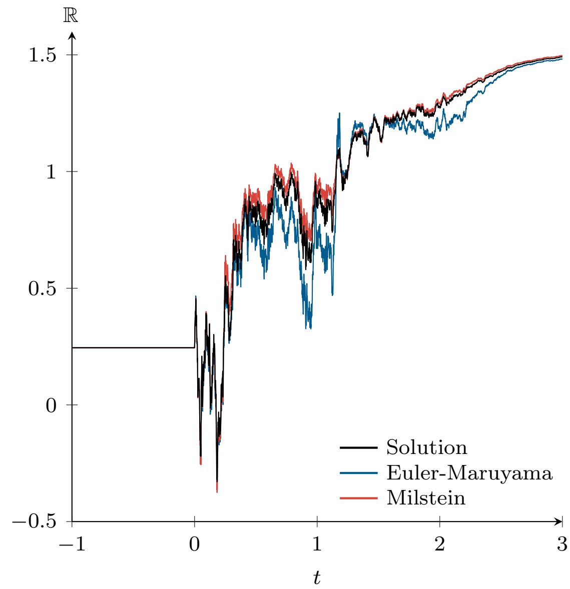


iv) Histogram of the empirical pathwise error over 10^4 realizations of the Euler-Maruyama and the Milstein approximations Y^h with step size $h = 2^{-12}$. The abscissa is logarithmically scaled.

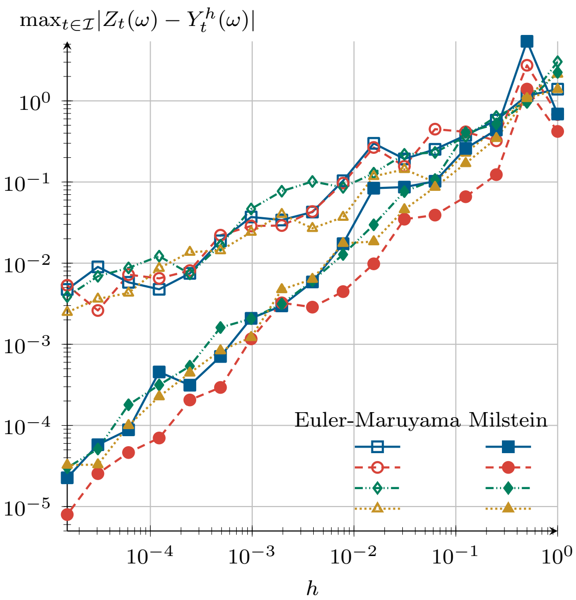
Figure VI.13. Simulation study of the Euler-Maruyama and Milstein scheme regarding SDDE (VI.37) with commutative noise. We set $t_0 = 0$, $\tau = 1$, $T = 3$, $a_1 = -2$, $a_2 = 3$, $a_3 = 1$, $m = 1$, $b_3^1 = 2$, and $\xi_t = 1$ for $t \in [-1, 0]$. Let X be the analytical solution of linear SDDE (VI.31) with additive noise and $g(x) = \sinh(x)$ for $x \in \mathbb{R}$. We simulated the analytical solution $Z = g(X)$ of SDDE (VI.37) error-free using Example VI.2 and Example VI.4 where $M = 2^{12}$ and thus $\hbar_M = 2^{-12}$. The analytical solution and the numerical approximations are evaluated at the points in time $\mathcal{I} = \big\{ -1 + n\,\hbar_M : n \in \{0, 1, \ldots, 4M\}\big\}$.



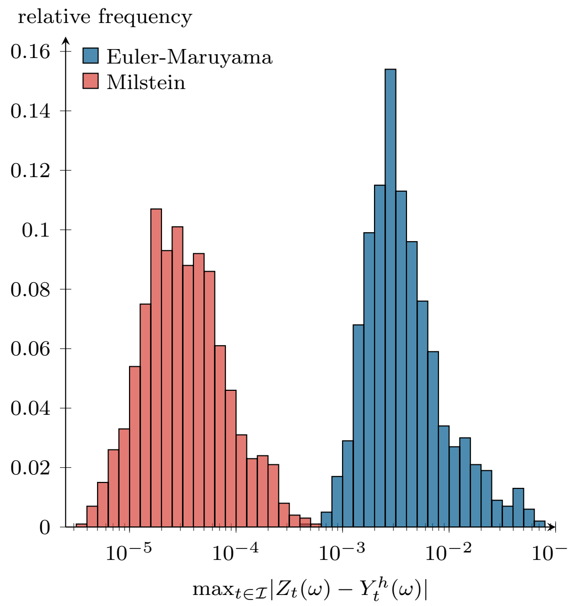
i) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 16\}$, in the strong sense ($S^p([-1, 3] \times \Omega; \mathbb{R})$-norm) for $p \in \{2, 7\}$ over $R = 10^3$ realizations. The scales of both axes are logarithmic.



ii) Trajectories of one realization of the analytical solution, and its Euler-Maruyama and Milstein approximations with respect to step size $h = 2^{-4}$.



iii) Empirical error of the Euler-Maruyama and the Milstein approximations Y^h with step sizes $h = 2^{-i}$, $i \in \{0, 1, \ldots, 16\}$, in the pathwise sense for four realizations. The scales of both axes are logarithmic.



iv) Histogram of the empirical pathwise error over 10^3 realizations of the Euler-Maruyama and the Milstein approximations Y^h with step size $h = 2^{-16}$. The abscissa is logarithmically scaled.

Figure VI.14. Simulation study of the Euler-Maruyama and Milstein scheme regarding SDDE (VI.39) with commutative noise. We set $t_0 = 0$, $\tau = 1$, $T = 3$, $a_1 = 0$, $a_2 = 4$, $a_3 = -\frac{1}{2}$, $m = 10$, $b_3^1 = -\frac{1}{2}$, $b_3^2 = 1$, $b_3^3 = \frac{1}{4}$, $b_3^4 = \frac{1}{10}$, $b_3^5 = -\frac{1}{8}$, $b_3^6 = -\frac{1}{4}$, $b_3^7 = -\frac{1}{10}$, $b_3^8 = \frac{1}{2}$, $b_3^9 = -\frac{1}{8}$, $b_3^{10} = -\frac{1}{8}$, and $\xi_t = \frac{1}{4}$ for $t \in [-1, 0]$. Let X be the analytical solution of linear SDDE (VI.25) with additive noise and $g(x) = \arctan(x)$ for $x \in \mathbb{R}$. We simulated the analytical solution $Z = g(X)$ of SDDE (VI.39) error-free using Example VI.1 and Example VI.4 where $M = 2^{16}$ and thus $\hbar_M = 2^{-16}$. The analytical solution and the numerical approximations are evaluated at the points in time $\mathcal{I} = \{-1 + n\,\hbar_M : n \in \{0, 1, \ldots, 4M\}\}$.

VII
Conclusion and Some Open Problems

Several new results on the convergence of the Milstein scheme are presented in this thesis. In the following, we highlight the most important results and depict some open problems.

In Chapter IV, we proved that the Milstein scheme for SDDEs converges in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ for arbitrary $p\in[1,\infty[$ with order $\alpha=1$, see Theorem IV.9 and Corollary IV.11. The considered SDDEs are allowed to have random initial conditions that we thoroughly took into account in our analysis of the convergence. Most terms of the expansion of the global error are estimated by standard Itô calculus, whereas one term lacks the martingale property. Here, it is more difficult to handle the supremum over time inside the expectation, and more sophisticated techniques are needed in order to obtain the desired order of convergence $\alpha=1$. The supremum over time was estimated by means of Lemma IV.22. Further, we used techniques from the Malliavin calculus. In this regard, we emphasize Lemma IV.19, which makes the Malliavin calculus applicable. The result of Lemma IV.19 and the techniques used in its proof might be useful in other contexts as well. They separate the SDDE's initial condition, which is independent of the Wiener process, from those random variables that are generated by the Wiener process. The latter were analyzed with the Malliavin calculus. Here, we looked at arbitrary complete probability spaces and did not limit ourselves to product probability spaces. As we assume a polynomial growth condition on the second partial derivatives of the drift coefficient, we needed a more general chain rule, which we stated in Theorem III.9. In the proof of Theorem IV.9, we further used that the solution of SDDE (II.1) with a deterministic initial condition is differentiable in the sense of Malliavin, see Theorem III.26.

Having proved that the Milstein scheme for SDDEs is convergent in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ for arbitrary $p\in[1,\infty[$, we obtained various corollaries. Using Lemma IV.3, the Milstein scheme converges pathwise with order $\alpha=1-\varepsilon$ for arbitrary $\varepsilon>0$, see Corollary IV.12. Moreover, if the SDDE under consideration has additive noise, the Milstein scheme coincides with the Euler-Maruyama scheme, and thus, the Euler-Maruyama scheme converges in this case in $S^p([t_0-\tau,T]\times\Omega;\mathbb{R}^d)$ for arbitrary $p\in[1,\infty[$ with order $\alpha=1$ and pathwise with order $\alpha=1-\varepsilon$ for arbitrary $\varepsilon>0$ as well, see Corollary IV.13 and Corollary IV.14.

If the SDDE under consideration does not satisfy commutativity condition (V.1), the Milstein scheme involves iterated stochastic integrals that need to be approximated. Various approximation were analyzed in Chapter V. We proved that the simple Fourier method converges in $L^p(\Omega;\mathbb{R})$ for arbitrary $p\in[2,\infty[$, see Theorem V.2, and provided an algorithm for the approximation of the delayed- and nondelayed-iterated stochastic integrals occurring in the Milstein scheme, see Algorithm V.4. If the diffusion coefficients do not depend on the past history of the SDDE's solution, only non-delayed iterated stochastic integrals occur in the Milstein scheme as in the case of SODEs. In this case, we further improved the approximation algorithm. We proved the convergence of this new method in $L^p(\Omega;\mathbb{R})$ for arbitrary $p \in [2,\infty[$ as well, see Theorem V.8, Theorem V.9, and Algorithm V.10. In Theorem V.11 and Theorem V.12, we further presented a variant of Algorithm V.10, which might be valuable in case of high-dimensional Wiener processes. The computational costs of these approximations were compared in Section V.3. Especially here, it turned out that the methods derived in Section V.2 are much more efficient than the approximation introduced by Wiktorsson in [136].

As we proved that our approximations of the iterated stochastic integrals are convergent in $L^p(\Omega;\mathbb{R})$ for arbitrary $p \in [2,\infty[$, we obtained, using Lemma V.16 and Corollary V.17, the convergence of the Milstein scheme based on these approximations in $L^p(\Omega;\mathbb{R}^d)$ for arbitrary $p \in [2,\infty[$ with order $\alpha = 1$ and pathwise convergence with order $\alpha = 1-\varepsilon$ for arbitrary $\varepsilon > 0$ as well, see Theorem V.18 and Theorem V.19.

Most stochastic integrals that occur in the proofs of the results mentioned above can be estimated using the Burkholder inequality. However, in Section II.2, we derived more sophisticated inequalities for time-discrete and time-continuous martingales whose constants are smaller than those of the Burkholder inequalities, see Theorem II.5 and Theorem II.6. Due to the smaller constants, these inequalities are highly valuable in stochastic analysis for accurate estimates, as for example in the numerical analysis of approximations of SDEs. As in the case of the time-discrete and time-continuous Burkholder inequalities, the constants in Theorem II.5 for the discrete Burkholder-type inequalities are best possible. However, it is an open problem whether the constants are best possible in case of the time-continuous martingale inequalities in Theorem II.6 as well.

In Chapter VI, we finally provided some simulations that illustrate and confirm our theoretical results on the convergence of the Milstein scheme. At first, we focused on analytical solutions of SDDEs that can be simulated error-free, see Section VI.1. We derived solutions for various SDDEs driven by multidimensional Wiener processes. Here, not only SDDEs with additive noise are considered but also more general SDDEs that satisfy commutativity condition (V.1), see Example VI.1, Example VI.2, and Example VI.4. Using these analytical solutions, we provided some numerical simulation studies in Section VI.2. These are the first examples that compare the Milstein approximation with the exactly simulated analytical solution.

In the following, we address some open problems that arose in the focus of this thesis and provide motivations for further and future research.

Concerning the Malliavin calculus and the continuity of the Skorohod integral operator, the precise constant in inequality (III.18) from Proposition III.25 seems to be unknown so far. As the inequality is used in the numerical analysis, it is natural to ask for the best possible constant of this inequality. We refer to [6, 88, 115] for some result on this constant.

For the convergence of the Milstein scheme, we supposed, among others, classical global Lipschitz conditions on the SDDE's coefficients in Assumption IV.8. In further research, the convergence under local Lipschitz conditions can be analyzed. The results in [2, 70] give inspirations on how to prove the pathwise convergence. In addition to [2], the convergence of the Euler-Maruyama scheme for SDDEs under local Lipschitz conditions is also analyzed in [83, 100].

In order to reduce the number of function evaluations of the Milstein scheme and to make the scheme easier applicable in numerical toolboxes, there is a high demand for efficient Runge-Kutta scheme for SDDEs, cf. [125] in case of SODEs and [89] in case of SPDEs. Further, we refer to [110] for a first approach in case of SDDEs, where $d = D = m = 1$.

Another type of Milstein scheme is the drift-randomized Milstein scheme that is proposed by Kruse and Wu, see [81]. They considered SODEs with nondifferentiable drift coefficients and prove the convergence by randomizing the drift coefficients. This introduced randomization causes a martingale property concerning the drift coefficients. A similar approach might be promising for SDDEs as well. Introducing this additional randomness in the Milstein scheme, the order of convergence $\alpha = 1$ could be proven without the Malliavin calculus. However, we emphasize that this approach results in a different type of scheme than the Milstein scheme considered in this thesis.

A further open problem is the efficient approximation of delayed-iterated stochastic integrals. The difficulty compared to Algorithm V.10 in Section V.2 is that the dependencies of the remainders of expansions (V.18) and (V.19) as well as random variables (V.25) on all intervals between the discretization points must be taken into account in order to approximate the remainders properly. Thus, for a fixed $K \in \mathbb{N}$, the remainders of the iterated stochastic integral approximations $I^K_{(i,j),n,\tau_l}$ introduced in Section V.1 have to be analyzed at once for all $i, j \in \{1, \ldots, m\}$, $l \in \{0, 1, \ldots, D\}$, and $n \in \{0, 1, \ldots, N-1\}$. Although the distributional properties of the remainders must be analyzed all at once, it seems that an algorithm generating these iterated stochastic integral approximations can be formulated sequentially like Algorithm V.4 if a Cholesky-type decomposition of the resulting conditional covariance matrix is used.

Furthermore, the savings of computational effort by Algorithm V.10 and its variant analyzed in Theorem V.11 and Theorem V.12 are especially of interest if the underlying Wiener process is high-dimensional, see Section V.3. Thus, it is promising to extend the approximations introduced in Section V.2 to iterated stochastic integrals driven by Q-Wiener processes, cf. [90], where an extension of Wiktorsson's method is presented. The dimension of the Wiener process approximating the Q-Wiener process driving SPDEs has to increase in order to obtain a higher accuracy of the Milstein approximation of the SPDE's solution. Hence, an extension of our methods in Section V.2 to the case of iterated stochastic integrals driven by a Q-Wiener process could supplant the algorithm proposed by Leonhard and Rößler in [90].

In this thesis, we focused on the strong and pathwise convergence of the Milstein scheme for SDDEs. We did not consider weak approximations of solution X of SDDE (II.1). The weak convergence of the Euler-Maruyama scheme for SDDEs is analyzed in [18, 140] for example. However, efficient weak approximations like multilevel Monte-Carlo methods for SDDEs have not been developed yet in contrast to SODEs, see e. g. [49, 56]. Our results on the convergence of the Milstein scheme for SDDEs could be valuable developing efficient weak approximations. In case of SODEs, we refer to [8, 30, 48, 50] for efficient algorithms involving the Milstein scheme.

Notations

$0_{i\times j}$	matrix of zeros of size $i \times j$
$\lambda \otimes \mathrm{P}$	product measure
$\mathscr{B}([t_0, T]) \otimes \mathscr{F}$	product-σ-algebra
$x \otimes z$	linear operator in Chapter III, see p. 28
$x \otimes y$	Kronecker product in Chapter V, see p. 137
$\mathbb{1}_A$	indicator function of set A
$\lvert\cdot\rvert$	absolute value
a	drift coefficient of SDDE (II.1)
a^i	ith component of a, see p. 9
$a^j_{k,n} \coloneqq a^j_{k,n,\tau_0}$	see p. 133
$a_{k,n}$	$(a^1_{k,n}, \ldots, a^m_{k,n})^{\mathrm{T}}$, cf. p. 137
a^j_{k,n,τ_l}	see p. 133
a_{k,n,τ_l}	$(a^1_{k,n,\tau_l}, \ldots, a^m_{k,n,\tau_l})^{\mathrm{T}}$, see p. 137
$A_{(i,j),n} \coloneqq A_{(i,j),n,\tau_0}$	see equation (V.15)
$A_{(i,j),n,\tau_l}$	see equation (V.17)
$A^K_{(i,j),n} \coloneqq A^K_{(i,j),n,\tau_0}$	see equation (V.18)
$A^K_{(i,j),n,\tau_l}$	see equation (V.19)
$A^K_n \coloneqq A^K_{n,\tau_0}$	see p. 137
A^K_{n,τ_l}	see equation (V.19)
α	order of convergence, see Definition (IV.1) and Definition (IV.2)
additive noise	see p. 78
$\mathscr{B}(E)$	Borel-σ-algebra on separable Banach space E
b^j	diffusion coefficient of SDDE (II.1), where $j \in \{1, \ldots, m\}$
$b^{i,j}$	ith component of b^j, see p. 9
$b^j_{k,n} \coloneqq b^j_{k,n,\tau_0}$	see p. 133
$b_{k,n}$	$(b^1_{k,n}, \ldots, b^m_{k,n})^{\mathrm{T}}$, cf. p. 137
b^j_{k,n,τ_l}	see p. 133

b_{k,n,τ_l}	$(b^1_{k,n,\tau_l}, \dots, b^m_{k,n,\tau_l})^{\mathrm{T}}$, see p. 137
β	see Assumption IV.8 *iii)*
commutative noise	see condition (V.1) on p. 129
$C(A;\mathbb{R}^d)$	see p. 9
$C^1(\mathbb{R}^{d\times(D+1)};\mathbb{R}^d)$	see p. 9
$C^2(\mathbb{R}^{d\times(D+1)};\mathbb{R}^d)$	see p. 9
$C_p^\infty(\mathbb{R}^K;\mathbb{R})$	see p. 27
$c_{\delta,p}$	constant in inequality (III.18)
$C_{\mathrm{D},p}$	constant in inequality (III.22)
$C_{\bar{I},p}$	constant in inequality (V.54)
cf.	abbreviation of the Latin word *confer* – compare (to/with)
$\mathrm{cost}[\cdot]$	see equations (V.45), (V.46), (V.47), and (V.48)
$\mathrm{cost}[\cdot \mid \mathrm{MSE} = Ch_n^3]$	see equations (V.49), (V.50), (V.51), and (V.52)
$\mathrm{Cov}[\cdot,\cdot]$	covariance
d	dimension of SDDE (II.1)
D	number of different positive delays in SDDE (II.1)
D	Malliavin derivative, see Definition III.3, p. 31, Definition III.13, and p. 34
δ	divergence operator, see Definition III.10 and equation (III.16)
$\mathscr{D}^p(\Omega;\mathbb{R})$	see Definition III.6
$\mathscr{D}^p(\Omega;E)$	see Definition III.15
$\mathrm{D}_t^j F$	see p. 31
$\mathrm{D}_t^j F_s$	see p. 35
$\mathrm{D}_t^j F_s^l$	see p. 35
$\mathrm{D}_t^j X_t$	see Theorem III.26
$\mathrm{dom}\, f$	domain of function or operator f
$\mathrm{E}[\cdot]$	expectation on $(\Omega,\mathscr{F},\mathrm{P})$
$\mathrm{E}[\cdot \mid \mathscr{G}]$	conditional expectation
e_i	ith unit vector if not otherwise stated
e. g.	abbreviation of Latin *exempli gratia* – for example
$\mathcal{E}_{\bar{I},p}(h_n,K_n)$	bound in inequality (V.55), also see p. 149
$\mathscr{F}$	σ-algebra of $(\Omega,\mathscr{F},\mathrm{P})$
$(\mathscr{F}_t)_{t\in[t_0-\tau,T]}$	filtration that satisfies the usual conditions, see p. 5
$\mathscr{G} := \mathscr{G}_T$	σ-algebra, see p. 7

$(\mathscr{G}_t)_{t\in[t_0,T]}$	filtration, see p. 7
$\Gamma(\cdot)$	Gamma function, see p. 135 and cf. Lemma V.20
γ_a	see inequality (IV.16) and Assumption IV.8 *vi)*
γ_b	see inequality (IV.17) and Assumption IV.8 *vi)*
h	maximum step size of discretization $\{t_0, t_1, \ldots, t_N\}$, see equation (IV.1)
$\hbar_M$	equidistant step size in Chapter VI, see formula (VI.20)
$\mathrm{H} := \mathrm{H}_{\mathbb{R}}$	abbreviation for $L^2([t_0,T]; L_{HS}(\mathbb{R}^m;\mathbb{R}))$
H_E	abbreviation for $L^2([t_0,T]; L_{HS}(\mathbb{R}^m;E))$, see equation (III.2)
H_m	selection matrix, see equation (V.27)
$H^p(A\times\Omega;E)$	see p. 8
$\langle\cdot,\cdot\rangle_E$	inner product of real separable Hilbert space E
$\langle\cdot,\cdot\rangle_{\mathrm{H}}$	see equation (III.4)
$\langle\cdot,\cdot\rangle_{L_{HS}(E_1;E_2)}$	see p. 28
I_m	identity matrix in $\mathbb{R}^{m\times m}$
$\operatorname{im} f$	image of function or operator f
$I_{(i,j),n} := I_{(i,j),n,\tau_0}$	see p. 132
I_n	see p. 139
$I_{(i,j),n,\tau_l}$	see p. 132
I_{n,τ_l}	cf. pp. 139 and 140
$I^K_{(i,j),n} := I^K_{(i,j),n,\tau_0}$	see equation (V.20)
$I^{(K)'}_{(i,j),n}$	cf. [136, Theorem 4.1] and p. 145
I^K_n	cf. p. 139 and Algorithm V.4
I^{K+}_n	cf. equation (V.41) and Algorithm V.10
$\tilde{I}^{K+}_n$	cf. equation (V.44)
$I^{(K)'}_n$	cf. [136, Theorem 4.1] and p. 145
$\bar{I}^{K_n}_{(i,j),n,\tau_l}$	see p. 155
$I^K_{(i,j),n,\tau_l}$	see equation (V.21)
I^K_{n,τ_l}	cf. p. 139 and Algorithm V.4
$\int_{t_0}^T h(s)\,\mathrm{d}W_s$	short notation of $\sum_{j=1}^m \int_{t_0}^T h^j(s)\,\mathrm{d}W^j_s$, see equation (III.5)
$\sum_{j=1}^m \int_{t_0}^T F^j_t\,\delta W^j_t$	notation for $\int_{t_0}^T F_t\,\delta W_t := \delta(F)$, see p. 38
K_a	constant of linear growth condition (II.10)
K_b	constant of linear growth condition (II.11)
$K_{\partial^2 a}$	see Assumption IV.8 *v)*

$K_{\partial^2 b}$	see Assumption IV.8 *v)*
λ	Lebesgue-measure on $\mathbb{R}$
$\ell_{i,j}$	see equation (VI.26) and Example VI.2
L_a	global Lipschitz constant of drift coefficient a, see inequality (II.8)
L_b	global Lipschitz constant of drift coefficient b^j, see inequality (II.9)
$L_{\partial b}$	see Assumption IV.8 *iii)*
L_Σ	see equation (VI.26) and Example VI.2
$L_{\mathrm{t},a}$	see inequality (IV.16) and Assumption IV.8 *vi)*
$L_{\mathrm{t},b}$	see inequality (IV.17) and Assumption IV.8 *vi)*
L_ξ	see inequality (IV.18) and Assumption IV.8 *vii)*
$L^p(\Omega; E)$	see p. 7
$L^p_{\mathscr{G}}(\Omega; E)$	$\mathscr{G}$-measurable random variables in $L^p(\Omega; E)$, see p. 29
$L_{HS}(E_1; E_2)$	space of Hilbert-Schmidt operators from E_1 to E_2, see p. 28
$L^2([t_n, t_{n+1}]; \mathbb{R})$	see p. 133
m	dimension of Wiener process W
$x \vee y$	maximum of x and y
$x \wedge y$	minimum of x and y
$\mathbb{N}$	set of natural numbers
$\mathbb{N}_0$	set of natural numbers with zero, that is, $\mathbb{N} \cup \{0\}$
$\lVert\cdot\rVert$	Euclidean norm
$\lVert\cdot\rVert_{\mathscr{D}^p(\Omega;\mathbb{R})}$	see Definition III.6
$\lVert\cdot\rVert_{\mathscr{D}^p(\Omega;E)}$	see Definition III.15
$\lVert\cdot\rVert_E$	norm on vector space E
$\lVert\cdot\rVert_{\mathrm{F}}$	Frobenius norm
$\lVert\cdot\rVert_{H^p(A\times\Omega;E)}$	see p. 8
$\lVert\cdot\rVert_{L^p(\Omega;E)}$	see p. 7
$\lVert\cdot\rVert_{L^p_{\mathscr{G}}(\Omega;E)}$	see p. 29
$\lVert\cdot\rVert_{L_{HS}(E_1;E_2)}$	see p. 28
$\lVert\cdot\rVert_{S^p(A\times\Omega;E)}$	see p. 8
$\mathrm{N}(\mu, \varSigma)$	normal distribution with expectation μ and covariance $\varSigma$
Ω	sample space of $(\Omega, \mathscr{F}, \mathrm{P})$
$\mathcal{O}$	Landau symbol, cf. [87, pp. 31, 59]
$\mathscr{o}$	Landau symbol, cf. [87, p. 61]
$\partial_{x_l^i} f$	partial derivative, see p. 9

P	probability measure of $(\Omega, \mathscr{F}, \mathrm{P})$
$\mathrm{P}\vert_{\mathscr{G}}$	probability measure P restricted to σ-algebra $\mathscr{G}$
P_m	permutation matrix, see equation (V.29)
$(\Omega, \mathscr{F}, \mathrm{P})$	complete probability space
$\Phi_{s,t}$	fundamental solution, see p. 198
$\mathbb{R} = \mathbb{R}^1$	real line
$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathbb{R}^{i\times j}$	space of real matrices of size $i \times j$
$\mathbb{R}_+$	set of all nonnegative real numbers
$\mathcal{R}_r$	see pp. 87–88
$\mathcal{R}_5^l$	see equation (IV.61)
$\mathcal{R}_5^{l,z_k^r}$	see inequality (IV.109)
$\mathcal{R}_5'$	see equation (IV.116)
$\mathcal{R}_5''$	see equation (IV.117)
ϱ_a	see Assumption IV.8 *v)*
ϱ_b	see Assumption IV.8 *v)*
SDE	*S*tochastic *D*ifferential *E*quation
SDDE	*S*tochastic *D*elay *D*ifferential *E*quation
SFDE	*S*tochastic *F*unctional *D*ifferential *E*quation
SODE	*S*tochastic *O*rdinary *D*ifferential *E*quation
SPDE	*S*tochastic *P*artial *D*ifferential *E*quation
$\sim$	with distribution
$\lfloor s \rfloor$	see equation (IV.14)
$\lceil s \rceil$	see equation (IV.15)
$\sigma(\mathcal{E})$	σ-algebra generated by set $\mathcal{E}$
$\sigma_2^{K_n}$	see p. 147
$\sigma_4^{K_n}$	see p. 147
$\Sigma_{i,j}$	covariance entries, see p. 209
$\mathit{\Sigma}_{1,n}^K$	covariance matrix, see equation (V.34)
$\sqrt{\mathit{\Sigma}_{1,n}^K}$	square root matrix of $\mathit{\Sigma}_{1,n}^K$ so that $\mathit{\Sigma}_{1,n}^K = \sqrt{\mathit{\Sigma}_{1,n}^K}\sqrt{\mathit{\Sigma}_{1,n}^K}$, see p. 143
$\mathit{\Sigma}_{2,n}^K$	covariance matrix, see equation (V.35)
$\mathit{\Sigma}_{3,n}^K$	covariance matrix, see equation (V.36)
S_n^K	Schur complement, see equation (V.37)
$\sqrt{S_n^K}$	square root matrix of S_n^K so that $S_n^K = \sqrt{S_n^K}\sqrt{S_n^K}$, see p. 143

$\sqrt{\mathrm{E}[S_n^K]}$	see Lemma V.6
$\sqrt{\Sigma_\infty}$	see [136, Equations (4.5) and (4.7)] and cf. p. 144
$S^p(A \times \Omega; E)$	see p. 8
$\mathscr{S}(\Omega; \mathbb{R})$	set of $\mathbb{R}$-valued smooth random variables, see Definition III.1
$\mathscr{S}(\Omega; E)$	set of E-valued smooth random variables, see Definition III.11
T	finite time horizon of SDDE (II.1)
t_0	starting point of SDDE (II.1)
t_n	point of discretization $\{t_0, t_1, \dots, t_N\}$, where $t_N = T$
τ	see p. 6
$\tau_0 := 0$	see p. 6
τ_l	positive time lag in SDDE (II.1), see p. 6
A^{T}	transpose of matrix or vector A
$\mathcal{T}(t, X_t)$	short notation of $(t, t-\tau_1, \dots, t-\tau_D, X_t, X_{t-\tau_1}, \dots, X_{t-\tau_D})$, see p. 7
$\mathcal{T}(t, Y_t)$	short notation of $(t, t-\tau_1, \dots, t-\tau_D, Y_t, Y_{t-\tau_1}, \dots, Y_{t-\tau_D})$, see p. 64
$\mathcal{T}(\lfloor s \rfloor, X_{\lfloor s \rfloor} + \theta(X_s - X_{\lfloor s \rfloor}))$	short notation, see p. 73
ϑ	cf. Lemma IV.22
$\mathrm{Var}[\cdot]$	variance
$\mathrm{vec}[\cdot]$	cf. equation (V.23)
W	m-dimensional Wiener process, see Definition II.1
W^j	jth component of Wiener process W, where $j \in \{1, \dots, m\}$
$\mathrm{d}W_s^j$	stochastic integration with respect to W^j in the sense of Itô
δW_s^j	stochastic integration with respect to W^j in the sense of Skorohod, see p. 38
$\Delta W_n^j := \Delta W_{n,\tau_0}^j$	see p. 132
ΔW_n	$(\Delta W_n^1, \dots, \Delta W_n^m)^{\mathrm{T}}$, see p. 137
$\Delta W_{n,\tau_l}^j$	see p. 132
$\Delta W_{n,\tau_l}$	$(\Delta W_{n,\tau_l}^1, \dots, \Delta W_{n,\tau_l}^m)^{\mathrm{T}}$, see p. 137
X	solution of SDDE (II.1)
X^ξ	for emphasis of initial condition ξ of solution X of SDDE (II.1)
ξ	initial condition of SDDE (II.1)
$\tilde{\xi}$	initial condition, see equation (IV.82)
Y	approximation of solution X of SDDE (II.1), for example Euler-Maruyama scheme (IV.13), Milstein scheme (IV.33), Milstein scheme (VI.38) for SDDEs with commutative noise

Y^h	for emphasis of maximum step size h of Y
$\overline{Y}$	Milstein scheme based on approximated iterated stochastic integrals, see equations (V.56) and (V.58)
$\overline{Y}^h$	for emphasis of maximum step size h of $\overline{Y}$
$\mathbb{Z}$	set of all integers
z_k^r	deterministic initial condition, see p. 95
ζ^r	initial condition, see p. 95

Bibliography

[1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover Publications, Inc., 1972.

[2] B. Akhtari, E. Babolian, and A. Neuenkirch. An Euler scheme for stochastic delay differential equations on unbounded domains: Pathwise convergence. *Discrete Contin. Dyn. Syst. Ser. B*, 20(1):23–38, 2015. DOI: 10.3934/dcdsb.2015.20.23.

[3] E. Alòs and D. Nualart. A maximal inequality for the Skorohod integral. In I. Csiszár and G. Michaletzky, editors, *Stochastic Differential and Difference Equations*, volume 23 of *Progress in Systems and Control Theory*, pages 241–251. Birkhäuser, Boston, MA, 1997. DOI: 10.1007/978-1-4612-1980-4_18.

[4] E. Alòs and D. Nualart. Stochastic integration with respect to the fractional Brownian motion. *Stochastics Stochastics Rep.*, 75(3):129–152, 2003. DOI: 10.1080/1045112031000078917.

[5] G. E. Andrews, R. Askey, and R. Roy. *Special functions*, volume 71 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1999. DOI: 10.1017/CBO9781107325937.

[6] N. Arcozzi. Riesz transforms on compact Lie groups, spheres and Gauss space. *Ark. Mat.*, 36(2):201–231, 1998. DOI: 10.1007/BF02384766.

[7] L. Arnold. *Stochastic Differential Equations: Theory and Applications.* John Wiley & Sons, New York, NY, 1974.

[8] R. Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance Stoch.*, 13(3):381–401, 2009. DOI: 10.1007/s00780-009-0099-7.

[9] C. T. H. Baker and E. Buckwar. Numerical analysis of explicit one-step methods for stochastic delay differential equations. *LMS J. Comput. Math.*, 3:315–335, 2000. DOI: 10.1112/s1461157000000322.

[10] A. V. Balakrishnan. *Applied functional analysis*, volume 3 of *Applications of Mathematics.* Springer, New York, NY, second edition, 1981. DOI: 10.1007/978-1-4612-5865-0.

[11] A. Barth and A. Lang. L^p and almost sure convergence of a Milstein scheme for stochastic partial differential equations. *Stochastic Process. Appl.*, 123(5):1563–1587, 2013. DOI: 10.1016/j.spa.2013.01.003.

[12] H. Bauer. *Wahrscheinlichkeitstheorie.* Walter de Gruyter, Berlin, 1991.

[13] S. Becker, A. Jentzen, and P. E. Kloeden. An exponential Wagner-Platen type scheme for SPDEs. *SIAM J. Numer. Anal.*, 54(4):2389–2426, 2016. DOI: 10.1137/15M1008762.

[14] Y. M. Berezansky, Z. G. Sheftel, and G. F. Us. *Functional Analysis. Vol. II*, volume 86 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1996. DOI: 10.1007/978-3-0348-9024-3.

[15] K. Bichteler and S. J. Lin. On the stochastic Fubini theorem. *Stochastics Stochastics Rep.*, 54(3-4):271–279, 1995. DOI: 10.1080/17442509508834009.

[16] L. Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. DOI: 10.1137/1.9781611971286.

[17] E. Buckwar. Introduction to the numerical analysis of stochastic delay differential equations. *J. Comput. Appl. Math.*, 125(1-2):297–307, 2000. DOI: 10.1016/S0377-0427(00)00475-1.

[18] E. Buckwar, R. Kuske, S.-E. A. Mohammed, and T. Shardlow. Weak convergence of the Euler scheme for stochastic differential delay equations. *LMS J. Comput. Math.*, 11:60–99, 2008. DOI: 10.1112/s146115700000053x.

[19] D. L. Burkholder. Sharp inequalities for martingales and stochastic integrals. *Astérisque*, (157-158):75–94, 1988.

[20] D. L. Burkholder. Explorations in martingale theory and its applications. In *École d'Été de Probabilités de Saint-Flour, XIX - 1989*, volume 1464 of *Lecture Notes in Mathematics*, pages 1–66. Springer, Berlin, 1991. DOI: 10.1007/BFb0085167.

[21] K. L. Chung and R. J. Williams. *Introduction to Stochastic Integration*. Modern Birkhäuser Classics. Birkhäuser, Basel, second edition, 2014. DOI: 10.1007/978-1-4614-9587-1.

[22] D. S. Clark. Short proof of a discrete Gronwall inequality. *Discrete Appl. Math.*, 16(3):279–281, 1987. DOI: 10.1016/0166-218X(87)90064-3.

[23] J. M. C. Clark and R. J. Cameron. The maximum rate of convergence of discrete approximations for stochastic differential equations. In *Stochastic Differential Systems Filtering and Control: Proceedings of the IFIP-WG 7/1 Working Conference Vilnius, Lithuania, USSR, Aug. 28–Sept. 2, 1978*, volume 25 of *Lecture Notes in Control and Information Sciences*, pages 162–171. Springer, Berlin, 1980. DOI: 10.1007/BFb0004007.

[24] D. L. Cohn. *Measure theory*. Birkhäuser Advanced Texts Basler Lehrbücher. Birkhäuser, New York, NY, second edition, 2013. DOI: 10.1007/978-1-4614-6956-8.

[25] S. G. Cox and J. van Neerven. Pathwise Hölder convergence of the implicit-linear Euler scheme for semi-linear SPDEs with multiplicative noise. *Numer. Math.*, 125(2):259–345, 2013. DOI: 10.1007/s00211-013-0538-4.

[26] G. Da Prato, S. Kwapień, and J. Zabczyk. Regularity of solutions of linear stochastic equations in Hilbert spaces. *Stochastics*, 23(1):1–23, 1987. DOI: 10.1080/17442508708833480.

[27] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 45 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992. DOI: `10.1017/cbo9780511666223`.

[28] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 2014. DOI: `10.1017/cbo9781107295513`.

[29] A. L. Dawidowicz and K. Twardowska. On the relation between the Stratonovich and Itô integrals with integrands of delayed argument. *Demonstratio Math.*, 28(2):465–478, 1995. DOI: `10.1515/dema-1995-0225`.

[30] K. Debrabant and A. Rössler. On the acceleration of the multi-level Monte Carlo method. *J. Appl. Probab.*, 52(2):307–322, 2015. DOI: `10.1239/jap/1437658600`.

[31] A. Defant and K. Floret. *Tensor norms and operator ideals*, volume 176 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1993.

[32] C. Dellacherie and P.-A. Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.

[33] C. Dellacherie and P.-A. Meyer. *Probabilities and potential B. Theory of martingales*, volume 72 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1982.

[34] C. Doléans. Variation quadratique des martingales continues à droite. *Ann. Math. Statist*, 40:284–289, 1969. DOI: `10.1214/aoms/1177697823`.

[35] J. L. Doob. *Stochastic Processes*. John Wiley & Sons, New York, NY, 1953.

[36] L. E. Dubins and D. Gilat. On the distribution of maxima of martingales. *Proc. Amer. Math. Soc.*, 68(3):337–338, 1978. DOI: `10.2307/2043117`.

[37] N. Dunford and J. T. Schwartz. *Linear Operators. Part I: General Theory*. Wiley Classics Library. John Wiley & Sons, Inc., New York, NY, 1988.

[38] J. Elstrodt. *Maß- und Integrationstheorie*. Springer-Lehrbuch. Springer, Berlin, 2011. DOI: `10.1007/978-3-642-17905-1`.

[39] L. Euler. *Institutionum Calculi Integralis*. Impensis Academiae Imperialis Scientiarum, Petropoli, 1768.

[40] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010. DOI: `10.1090/gsm/019`.

[41] O. Faure. *Simulation du mouvement brownien et des diffusions*. Thèse de doctorat, Ecole Nationale des Ponts et Chaussées, Français, 1992.

[42] J. G. Gaines and T. J. Lyons. Random generation of stochastic area integrals. *SIAM J. Appl. Math.*, 54(4):1132–1146, 1994. DOI: `10.1137/S0036139992235706`.

[43] P. Gänssler and W. Stute. *Wahrscheinlichkeitstheorie*. Springer, Berlin, 1977. DOI: `10.1007/978-3-642-66749-7`.

[44] A. M. Garsia. *Martingale inequalities: Seminar notes on recent progress.* Mathematics Lecture Notes Series. W. A. Benjamin, Inc., Reading, MA, 1973.

[45] R. K. Getoor and M. J. Sharpe. Conformal martingales. *Invent. Math.*, 16(4):271–308, 1972. DOI: 10.1007/BF01425714.

[46] I. I. Gihman and A. V. Skorohod. *Stochastic Differential Equations*, volume 72 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 2. Folge.* Springer, Berlin, 1972.

[47] I. I. Gikhman and A. V. Skorokhod. *Introduction to the theory of random processes.* W. B. Saunders Co., Philadelphia, PA, 1969.

[48] M. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer, Berlin, 2008. DOI: 10.1007/978-3-540-74496-2_20.

[49] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008. DOI: 10.1287/opre.1070.0496.

[50] M. B. Giles, K. Debrabant, and A. Rössler. Analysis of multilevel Monte Carlo path simulation using the Milstein discretisation. *Discrete Contin. Dyn. Syst. Ser. B*, 24(8):3881–3903, 2019. DOI: 10.3934/dcdsb.2018335.

[51] I. Gyöngy. A note on Euler's approximations. *Potential Anal.*, 8(3):205–216, 1998. DOI: 10.1023/A:1008605221617.

[52] I. Gyöngy and S. Sabanis. A note on Euler approximations for stochastic differential equations with delay. *Appl. Math. Optim.*, 68(3):391–412, 2013. DOI: 10.1007/s00245-013-9211-7.

[53] W. Hackenbroch and A. Thalmaier. *Stochastische Analysis. Eine Einführung in die Theorie der stetigen Semimartingale.* Mathematische Leitfäden. B. G. Teubner, Stuttgart, 1994. DOI: 10.1007/978-3-663-11527-4.

[54] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*, volume 8 of *Springer Series in Computational Mathematics.* Springer, Berlin, second revised edition, 1993. DOI: 10.1007/978-3-540-78862-1.

[55] P. R. Halmos and V. S. Sunder. *Bounded integral operators on L^2 spaces*, volume 96 of *Ergebnisse der Mathematik und ihrer Grenzgebiete.* Springer, Berlin, 1978. DOI: 10.1007/978-3-642-67016-9.

[56] S. Heinrich. Multilevel monte carlo methods. In S. Margenov, J. Waśniewski, and P. Yalamov, editors, *Large-Scale Scientific Computing*, volume 2179 of *Lecture Notes in Computer Science*, pages 58–67. Springer, Berlin, 2001. DOI: 10.1007/3-540-45346-6_5.

[57] H. Heuser. *Lehrbuch der Analysis. Teil 2.* Mathematische Leitfäden. B. G. Teubner, Stuttgart, 9. edition, 1991.

[58] F. Hirsch. Propriété d'absolue continuité pour les équations différentielles stochastiques dépendant du passé. *J. Funct. Anal.*, 76(1):193–216, 1988. DOI: 10.1016/0022-1236(88)90056-0.

[59] R. A. Horn and C. R. Johnson. *Matrix analysis.* Cambridge University Press, Cambridge, second edition, 2013. DOI: 10.1017/9781139020411.

[60] Y. Hu, S.-E. A. Mohammed, and F. Yan. Discrete-time approximations of stochastic delay equations: The Milstein scheme. *Ann. Probab.*, 32(1A):265–314, 2004. DOI: 10.1214/aop/1078415836.

[61] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces. Volume I: Martingales and Littlewood-Paley Theory*, volume 63 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics.* Springer, Cham, 2016. DOI: 10.1007/978-3-319-48520-1.

[62] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces. Volume II: Probabilistic Methods and Operator Theory*, volume 67 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics.* Springer, Cham, 2017. DOI: 10.1007/978-3-319-69808-3.

[63] K. Itô. Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20(8):519–524, 1944. DOI: 10.3792/pia/1195572786.

[64] K. Itô. On a formula concerning stochastic differentials. *Nagoya Math. J.*, 3:55–65, 1951.

[65] K. Itô and M. Nisio. On stationary solutions of a stochastic differential equation. *J. Math. Kyoto Univ.*, 4:1–75, 1964. DOI: 10.1215/kjm/1250524705.

[66] K. Itô and M. Nisio. On the convergence of sums of independent Banach space valued random variables. *Osaka J. Math.*, 5:35–48, 1968.

[67] J. Jacod and P. Protter. *Probability Essentials.* Universitext. Springer, Berlin, second edition, 2004. DOI: 10.1007/978-3-642-55682-1.

[68] S. Janson. *Gaussian Hilbert Spaces*, volume 129 of *Cambridge Tracts in Mathematics.* Cambridge University Press, 1997. DOI: 10.1017/CBO9780511526169.

[69] A. Jentzen and P. E. Kloeden. *Taylor Approximations for Stochastic Partial Differential Equations*, volume 83 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial & Applied Mathematics (SIAM), Philadelphia, PA, 2011. DOI: 10.1137/1.9781611972016.

[70] A. Jentzen, P. E. Kloeden, and A. Neuenkirch. Pathwise approximation of stochastic differential equations on domains: Higher order convergence rates without global Lipschitz coefficients. *Numer. Math.*, 112(1):41–64, 2009. DOI: 10.1007/s00211-008-0200-8.

[71] A. Jentzen and M. Röckner. A Milstein scheme for SPDEs. *Found. Comput. Math.*, 15(2):313–362, 2015. DOI: 10.1007/s10208-015-9247-y.

[72] S. Kaden and J. Potthoff. Progressive stochastic processes and an application to the Itô integral. *Stochastic Anal. Appl.*, 22(4):843–865, 2004. DOI: 10.1081/SAP-120037621.

[73] R. V. Kadison and J. R. Ringrose. *Fundamentals of the theory of operator algebras. Volume I: Elementary theory*, volume 100 of *Pure and Applied Mathematics.* Academic Press, Inc., New York, NY, 1983.

[74] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications. Springer, New York, NY, second edition, 2002. DOI: 10.1007/978-1-4757-4015-8.

[75] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer, New York, NY, 1998. DOI: 10.1007/978-1-4612-0949-2.

[76] A. Klenke. *Probability Theory. A Comprehensive Course.* Universitext. Springer, London, 2014. DOI: 10.1007/978-1-4471-5361-0.

[77] P. E. Kloeden and A. Neuenkirch. The pathwise convergence of approximation schemes for stochastic differential equations. *LMS J. Comput. Math.*, 10:235–253, 2007. DOI: 10.1112/S1461157000001388.

[78] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics.* Springer, Berlin, 1992. DOI: 10.1007/978-3-662-12616-5.

[79] P. E. Kloeden, E. Platen, and I. W. Wright. The approximation of multiple stochastic integrals. *Stochastic Anal. Appl.*, 10(4):431–441, 1992. DOI: 10.1080/07362999208809281.

[80] P. E. Kloeden and T. Shardlow. The Milstein scheme for stochastic delay differential equations without using anticipative calculus. *Stochastic Anal. Appl.*, 30(2):181–202, 2012. DOI: 10.1080/07362994.2012.628907.

[81] R. Kruse and Y. Wu. A randomized Milstein method for stochastic differential equations with non-differentiable drift coefficients. *Discrete Contin. Dyn. Syst. Ser. B*, 24(8):3475–3502, 2019. DOI: 10.3934/dcdsb.2018253.

[82] U. Küchler and E. Platen. Strong discrete time approximation of stochastic differential equations with time delay. *Math. Comput. Simul.*, 54(1-3):189–205, 2000. DOI: 10.1016/S0378-4754(00)00224-X.

[83] C. Kumar and S. Sabanis. Strong convergence of Euler approximations of stochastic differential equations with delay under local Lipschitz condition. *Stochastic Anal. Appl.*, 32(2):207–228, 2014. DOI: 10.1080/07362994.2014.858552.

[84] H.-H. Kuo. *Introduction to Stochastic Integration.* Universitext. Springer, New York, NY, 2006. DOI: 10.1007/0-387-31057-6.

[85] S. Kusuoka and D. Stroock. The partial Malliavin calculus and its application to nonlinear filtering. *Stochastics*, 12(2):83–142, 1984. DOI: 10.1080/17442508408833296.

[86] J. Lamperti. A simple construction of certain diffusion processes. *J. Math. Kyoto Univ.*, 4:161–170, 1964. DOI: 10.1215/kjm/1250524711.

[87] E. Landau. *Handbuch der Lehre von der Verteilung der Primzahlen. Erster Band.* B. G. Teubner, Leipzig, 1909.

[88] L. Larsson-Cohn. On the constants in the Meyer inequality. *Monatsh. Math.*, 137(1):51–56, 2002. DOI: 10.1007/s00605-002-0475-2.

[89] C. Leonhard and A. Rössler. Enhancing the order of the Milstein scheme for stochastic partial differential equations with commutative noise. *SIAM J. Numer. Anal.*, 56(4):2585–2622, 2018. DOI: 10.1137/16M1094087.

[90] C. Leonhard and A. Rössler. Iterated stochastic integrals in infinite dimensions: Approximation and error estimates. *Stoch PDE: Anal Comp*, 7(2):209–239, 2019. DOI: 10.1007/s40072-018-0126-9.

[91] P. Lévy. Wiener's random function, and other Laplacian random functions. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 171–187. University of California Press, Berkeley and Los Angeles, 1951.

[92] M. Loève. *Probability Theory II*, volume 46 of *Graduate Texts in Mathematics*. Springer, New York, fourth edition, 1978. DOI: 10.1007/978-1-4612-6257-2.

[93] J. Maas. Malliavin calculus and decoupling inequalities in Banach spaces. *J. Math. Anal. Appl.*, 363(2):383–398, 2010. DOI: 10.1016/j.jmaa.2009.08.041.

[94] J. Maas and J. M. A. M. van Neerven. A Clark-Ocone formula in UMD Banach spaces. *Electron. Commun. Probab.*, 13:151–164, 2008. DOI: 10.1214/ECP.v13-1361.

[95] J. R. Magnus and H. Neudecker. The commutation matrix: some properties and applications. *Ann. Statist.*, 7(2):381–394, 1979. DOI: 10.1214/aos/1176344621.

[96] S. J. A. Malham and A. Wiese. Efficient almost-exact Lévy area sampling. *Statist. Probab. Lett.*, 88:50–55, 2014. DOI: 10.1016/j.spl.2014.01.022.

[97] P. Malliavin. *Stochastic Analysis*, volume 313 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 1997. DOI: 10.1007/978-3-642-15074-6.

[98] X. Mao. *Stochastic Differential Equations and Applications*. Horwood Publishing, Chichester, second edition, 2007. DOI: 10.1533/9780857099402.

[99] X. Mao. Numerical solutions of stochastic differential delay equations under the generalized Khasminskii-type conditions. *Appl. Math. Comput.*, 217(12):5512–5524, 2011. DOI: 10.1016/j.amc.2010.12.023.

[100] X. Mao and S. Sabanis. Numerical solutions of stochastic differential delay equations under local Lipschitz condition. *J. Comput. Appl. Math.*, 151(1):215–227, 2003. DOI: 10.1016/S0377-0427(02)00750-1.

[101] G. Maruyama. Continuous Markov processes and stochastic equations. *Rend. Circ. Mat. Palermo (2)*, 4:48–90, 1955. DOI: 10.1007/BF02846028.

[102] The MathWorks, Inc., Natick, Massachusetts. *MATLAB Release 2018b*, 2018.

[103] P.-A. Meyer. Inégalités de normes pour les intégrales stochastiques. In C. Dellacherie, P.-A. Meyer, and M. Weil, editors, *Séminaire de Probabilités XII*, volume 649 of *Lecture Notes in Mathematics*, pages 757–762. Springer, Berlin, 1978. DOI: 10.1007/BFb0064636.

[104] M. Milošević and M. Jovanović. An application of Taylor series in the approximation of solutions to stochastic differential equations with time-dependent delay. *J. Comput. Appl. Math.*, 235(15):4439–4451, 2011. DOI: 10.1016/j.cam.2011.04.009.

[105] G. N. Milstein. *Numerical Integration of Stochastic Differential Equations*, volume 313 of *Mathematics and Its Applications*. Springer, Dordrecht, 1995. DOI: 10.1007/978-94-015-8455-5.

[106] G. N. Mil'shtein. Approximate integration of stochastic differential equations. *Theory Probab. Appl.*, 19(3):557–562, 1975. DOI: 10.1137/1119062.

[107] S.-E. A. Mohammed. *Stochastic functional differential equations*, volume 99 of *Research Notes in Mathematics*. Pitman Advanced Publishing Program, Boston, MA, 1984.

[108] S.-E. A. Mohammed. Nonlinear flows of stochastic linear delay equations. *Stochastics*, 17(3):207–213, 1986. DOI: 10.1080/17442508608833390.

[109] S.-E. A. Mohammed. Stochastic Differential Systems With Memory: Theory, Examples and Applications. In L. Decreusefond, J. Gjerde, B. Øksendal, and A. S. Üstünel, editors, *Stochastic Analysis and Related Topics VI: Proceedings of the Sixth Oslo—Silivri Workshop Geilo 1996*, volume 42 of *Progress in Probability*, pages 1–77. Birkhäuser, Boston, MA, 1998. DOI: 10.1007/978-1-4612-2022-0_1.

[110] Y. Niu, K. Burrage, and C. Zhang. A derivative-free explicit method with order 1.0 for solving stochastic delay differential equations. *J. Comput. Appl. Math.*, 253:51–65, 2013. DOI: 10.1016/j.cam.2013.03.049.

[111] I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus: From Stein's method to universality*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012. DOI: 10.1017/CBO9781139084659.

[112] D. Nualart. *The Malliavin Calculus and Related Topics*. Probability and Its Applications. Springer, New York, NY, 1995. DOI: 10.1007/978-1-4757-2437-0.

[113] D. Nualart. *The Malliavin Calculus and Related Topics*. Probability and Its Applications. Springer, Berlin, second edition, 2006. DOI: 10.1007/3-540-28329-3.

[114] D. Nualart and E. Pardoux. Stochastic calculus with anticipating integrands. *Probab. Th. Rel. Fields*, 78(4):535–581, 1988. DOI: 10.1007/BF00353876.

[115] G. Pisier. Riesz transforms: A simpler analytic proof of P.A. Meyer's inequality. In J. Azéma, P.-A. Meyer, and M. Yor, editors, *Séminaire de Probabilités XXII*, volume 1321 of *Lecture Notes in Mathematics*, pages 485–501. Springer, Berlin, 1988. DOI: 10.1007/BFb0084154.

[116] C. Prévôt and M. Röckner. *A Concise Course on Stochastic Partial Differential Equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. DOI: 10.1007/978-3-540-70781-3.

[117] M. Pronk and M. Veraar. Tools for Malliavin calculus in UMD Banach spaces. *Potential Anal.*, 40(4):307–344, 2014. DOI: 10.1007/s11118-013-9350-0.

[118] P. Protter. $\mathscr{H}^p$ stability of solutions of stochastic differential equations. *Z. Wahrsch. Verw. Gebiete*, 44(4):337–352, 1978. DOI: 10.1007/BF01013196.

[119] P. E. Protter. *Stochastic Integration and Differential Equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer, Berlin, 2003. DOI: 10.1007/978-3-662-10061-5.

[120] O. Purtukhia. Fubini type theorems for ordinary and stochastic integrals. *Proc. A. Razmadze Math. Inst.*, 130:101–114, 2002.

[121] M. M. Rao. *Foundations of Stochastic Analysis.* Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic Press, Inc., New York, NY, 1981.

[122] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften.* Springer, Berlin, third edition, 1999. DOI: 10.1007/978-3-662-06400-9.

[123] E. Rio. Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.*, 22(1):146–163, 2009. DOI: 10.1007/s10959-008-0155-9.

[124] N. Rosli, A. Bahar, S. H. Yeak, and X. Mao. A systematic derivation of stochastic Taylor methods for stochastic delay differential equations. *Bull. Malays. Math. Sci. Soc. (2)*, 36(3):555–576, 2013.

[125] A. Rössler. Runge-Kutta methods for the strong approximation of solutions of stochastic differential equations. *SIAM J. Numer. Anal.*, 48(3):922–952, 2010. DOI: 10.1137/09076636X.

[126] W. Rudin. *Principles of mathematical analysis.* International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York, NY, third edition, 1976.

[127] T. Rydén and M. Wiktorsson. On the simulation of iterated Itô integrals. *Stochastic Process. Appl.*, 91(1):151–168, 2001. DOI: 10.1016/S0304-4149(00)00053-3.

[128] A. V. Skorokhod. On a generalization of a stochastic integral. *Theory Probab. Appl.*, 20(2):219–233, 1975. DOI: 10.1137/1120030.

[129] R. L. Stratonovich. A new representation for stochastic integrals and equations. *SIAM J. Control*, 4:362–371, 1966. DOI: 10.1137/0304028.

[130] D. Talay. Résolution trajectorielle et analyse numérique des équations différentielles stochastiques. *Stochastics*, 9(4):275–306, 1983. DOI: 10.1080/17442508308833257.

[131] J. van Neerven. γ-radonifying operators – a survey. In *The AMSI-ANU Workshop on Spectral Theory and Harmonic Analysis*, volume 44 of *Proc. Centre Math. Appl. Austral. Nat. Univ.*, pages 1–61. Austral. Nat. Univ., Canberra, 2010.

[132] J. M. A. M. van Neerven, M. C. Veraar, and L. Weis. Stochastic integration in UMD Banach spaces. *Ann. Probab.*, 35(4):1438–1478, 2007. DOI: 10.1214/009117906000001006.

[133] J. M. A. M. van Neerven, M. C. Veraar, and L. Weis. Stochastic integration in Banach spaces – a survey. In R. C. Dalang, M. Dozzi, F. Flandoli, and F. Russo, editors, *Stochastic Analysis: A Series of Lectures: Centre Interfacultaire Bernoulli, January–June 2012, Ecole Polytechnique Fédérale de Lausanne, Switzerland*, volume 68 of *Progress in Probability*, pages 297–332. Birkhäuser, Basel, 2015. DOI: 10.1007/978-3-0348-0909-2_11.

[134] M. Veraar. The stochastic Fubini theorem revisited. *Stochastics*, 84(4):543–551, 2012. DOI: 10.1080/17442508.2011.618883.

[135] J. B. Walsh. A note on uniform convergence of stochastic processes. *Proc. Amer. Math. Soc.*, 18:129–132, 1967. DOI: 10.2307/2035239.

[136] M. Wiktorsson. Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *Ann. Appl. Probab.*, 11(2):470–487, 2001. DOI: 10.1214/aoap/1015345301.

[137] F. Yan. *Topics on stochastic delay equations.* Dissertation, Southern Illinois University, Carbondale, Illinois, 1999.

[138] K. Yosida. *Functional Analysis*, volume 123 of *Classics in Mathematics.* Springer, Berlin, sixth edition, 1995. DOI: 10.1007/978-3-642-61859-8.

[139] M. Zakai. Some moment inequalities for stochastic integrals and for solutions of stochastic differential equations. *Israel J. Math.*, 5(3):170–176, 1967. DOI: 10.1007/BF02771103.

[140] H. Zhang. Weak approximation of stochastic differential delay equations for bounded measurable function. *LMS J. Comput. Math.*, 16:319–343, 2013. DOI: 10.1112/S1461157013000120.

[141] A. Zygmund. *Trigonometric Series. Volumes I & II combined.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2002. DOI: 10.1017/CBO9781316036587.