Journal of Computational and Applied Mathematics 383 (2021) 113112

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Randomized derivative-free Milstein algorithm for efficient approximation of solutions of SDEs under noisy information

Paweł M. Morkisz [a,b], Paweł Przybyłowicz [a,*]

[a] AGH University of Science and Technology, Faculty of Applied Mathematics, Al. A. Mickiewicza 30, 30-059 Kraków, Poland
[b] NVIDIA Poland

ARTICLE INFO

Article history:
Received 18 March 2020
Received in revised form 13 July 2020

MSC:
68Q25
65C30

Keywords:
SDEs
Standard noisy information
Pointwise approximation
Randomized Milstein algorithm
*n*th minimal error
Optimality

ABSTRACT

We deal with pointwise approximation of solutions of scalar stochastic differential equations in the presence of informational noise about underlying drift and diffusion coefficients. We define a randomized derivative-free version of Milstein algorithm $\bar{\mathcal{A}}_n^{df-RM}$ and investigate its error. We also study the lower bounds on the error of arbitrary algorithm. It turns out that in some case the scheme $\bar{\mathcal{A}}_n^{df-RM}$ is the optimal one. Finally, in order to test the algorithm $\bar{\mathcal{A}}_n^{df-RM}$ in practice, we report performed numerical experiments.

1. Introduction

In this paper we deal with pointwise approximation of solutions of the following scalar stochastic differential equations (SDEs)

$$\begin{cases} \mathrm{d}X(t) = a(t, X(t))\,\mathrm{d}t + b(t, X(t))\,\mathrm{d}W(t), & t \in [0, T], \\ X(0) = \eta, \end{cases} \tag{1}$$

where $T > 0$, η is an initial-value, and $W = \{W(t)\}_{t\geq 0}$ is a standard one-dimensional Wiener process on some probability space $(\Omega, \Sigma, \mathbb{P})$. We will assume that only noisy evaluations of a and b are allowed. The aim is to find an efficient approximation of $X(T)$ with an (asymptotic) error as small as possible.

The problem of approximation of solutions of SDEs under exact information about coefficients is well studied in literature, see, for example, the standard reference [1]. Much less is known when values of drift and diffusion coefficients are corrupted by some noise. Therefore, in this paper we assume that evaluations of the underlying coefficients are permissible only at certain precision levels. Such a disturbance may be caused by, for example, measurement errors, rounding errors, and lowering precision when performing computations on GPUs, see Remark 2 and [2,3] for further discussions and examples.

In literature there are many results on numerical problems under noisy information, such as integrating or approximation of regular functions [3–5], L_p approximation of piecewise regular functions [6], solutions of IVPs [7] or PDEs [8,9].

* Corresponding author.
E-mail addresses: morkiszp@agh.edu.pl (P.M. Morkisz), pprzybyl@agh.edu.pl (P. Przybyłowicz).

https://doi.org/10.1016/j.cam.2020.113112

For stochastic case we refer to [10] and [11] where authors studied approximation of, respectively, SDEs under noisy information by randomized Euler scheme and stochastic Itô integration in the case when also the values of the Wiener process W were inexact.

In this paper we extend the results obtained in [10]. Namely, we study approximation of solutions of SDEs by a randomized version of Milstein scheme under noisy information. For exact information such a version of the Milstein scheme was investigated in [12]. Here, however, we use its derivative free version in order to cover also the case of inexact information. Hence, our proof technique differs from that used in [12].

We use a suitable computation setting that allows us to model a situation when values of a's and b's are perturbed by some deterministic noise, see [10]. Namely, let $\delta_1, \delta_2 \in [0, 1]$ be the precision levels corresponding to drift and diffusion coefficients, respectively. (The case of $\delta_1 = \delta_2 = 0$ corresponds to the exact information.) Available standard information about each coefficient consists of noisy evaluations of the coefficients at a finite number of points $(t_i, y_i) \in [0, T] \times \mathbb{R}$. This means that, for example, for diffusion coefficient b and for a given point $(t_i, y_i) \in [0, T] \times \mathbb{R}$ evaluation returns $\tilde{b}(t_i, y_i)$ with the property that $|b(t_i, y_i) - \tilde{b}(t_i, y_i)| \leq \delta_2(1 + |y_i|)$. Moreover, as in [10] for $a = a(t, y)$ we allow randomized choices of sample points with respect to the time variable t. For the Wiener process W we assume that the information is exact, i.e. it is given by the values of W at a finite number of points $s_k \in [0, T]$. (See, however, Remark 4.) The error of the algorithm, using the information above, is measured in the qth mean ($q \geq 1$) maximized over the class of input data (a, b, η) and over all permissible information about (a, b, η) with the given precisions $\delta_1, \delta_2 \geq 0$.

Theorem 2, which is the main result of the paper, states that the nth minimal error (under suitably regular informational noise) is asymptotically equal to $\Theta(n^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}} + \delta_1 + \delta_2)$ where the factors in Θ do not depend on δ_1, δ_2. (Here, $\gamma_1, \gamma_2 \in (0, 1]$ are the Hölder exponents, with respect to time variable, of drift and diffusion coefficients, respectively.) A randomized derivative-free version $\bar{\mathcal{A}}_n^{df-RM}$ of the classical Milstein algorithm is defined, which uses noisy evaluations of drift and diffusion coefficients, and attains the desired rate of convergence. When the disturbances for a and b are more rough, then error term for the scheme $\bar{\mathcal{A}}_n^{df-RM}$ also depends on $\delta_2 n^{1/2}$, see Theorem 1 (ii). This implies that in order to obtain any convergence rate it is necessary to tend with both precision levels to zero suitably fast with respect to n.

The paper is organized as follows. Section 2 consists of the problem formulation, basic notions and definitions. Randomized derivative-free Milstein algorithm $\bar{\mathcal{A}}_n^{df-RM}$ under perturbed information together with the upper bounds on its error are presented in Section 3. In Section 4 we show the lower bound on the worst case error for an arbitrary algorithm (Lemma 3). This leads to the conclusion that randomized Milstein algorithm $\bar{\mathcal{A}}_n^{df-RM}$ is optimal (Theorem 2). Section 5 reports the numerical experiments performed for the algorithm $\bar{\mathcal{A}}_n^{df-RM}$. Finally, the Appendix contains auxiliary facts used in the paper.

2. Preliminaries

Let $T > 0$. We denote by $\mathbb{N} = \{1, 2, \ldots\}$. Let $W = \{W(t)\}_{t\geq 0}$ be a standard one-dimensional Wiener process on a complete probability space $(\Omega, \Sigma, \mathbb{P})$. We denote by $\{\Sigma_t\}_{t\geq 0}$ a filtration, satisfying the usual conditions, such that W is a Wiener process on $(\Omega, \Sigma, \mathbb{P})$ with respect to $\{\Sigma_t\}_{t\geq 0}$. Let $\Sigma_\infty = \sigma\left(\bigcup_{t\geq 0} \Sigma_t\right)$. For a random variable $X : \Omega \to \mathbb{R}$ we write $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$, where $q \in [2, +\infty)$. A continuous function $f : [0, T] \times \mathbb{R} \to \mathbb{R}$ belongs to $C^{0,j}([0, T] \times \mathbb{R})$, with $j \in \mathbb{N} \cup \{0\}$, provided that for all $k = 0, 1, \ldots, j$ the partial derivatives $\partial^k f/\partial y^k$ exist and are continuous on $[0, T] \times \mathbb{R}$.

For any $f \in C^{0,1}([0, T] \times \mathbb{R})$ by L_1 we mean the following differential operator

$$L_1 f(t, y) = f(t, y) \cdot \frac{\partial f}{\partial y}(t, y).$$

We will also use its derivative-free version. Namely, for $f \in C([0, T] \times \mathbb{R})$ and $h > 0$ the difference operator $\mathcal{L}_{1,h}$ is given as follows

$$\mathcal{L}_{1,h} f(t, y) = f(t, y) \cdot \Delta_h f(t, y),$$

where

$$\Delta_h f(t, y) = \frac{f(t, y + h) - f(t, y)}{h}.$$

(Basic properties of L_1 and $\mathcal{L}_{1,h}$, used in the paper, are gathered in the Appendix.)

Let $K > 0$ and $\gamma \in (0, 1]$. We say that $f : [0, T] \times \mathbb{R} \to \mathbb{R}$ belongs to the function class F_K^γ iff for all $t, s \in [0, T]$ and all $y, z \in \mathbb{R}$ it satisfies the following assumptions:

(i) $f \in C^{0,2}([0, T] \times \mathbb{R})$,
(ii) $|f(0, 0)| \leq K$,
(iii) $|f(t, y) - f(t, z)| \leq K|y - z|$,
(iv) $|f(t, y) - f(s, y)| \leq K(1 + |y|)|t - s|^\gamma$,
(v) $\left|\frac{\partial f}{\partial y}(t, y) - \frac{\partial f}{\partial y}(t, z)\right| \leq K|y - z|$.

In this paper we will be considering drift coefficients a from the following class

$$\mathcal{A}_K^{\gamma_1} = \left\{ f \in F_K^{\gamma_1} \;\middle|\; \left| \frac{\partial f}{\partial y}(t,y) - \frac{\partial f}{\partial y}(s,y) \right| \leq K(1+|y|)|t-s|^{\gamma_1} \text{ for all } t,s \in [0,T], y \in \mathbb{R} \right\},$$

while we will be assuming that diffusion coefficients are from

$$\mathcal{B}_K^{\gamma_2} = \left\{ f \in F_K^{\gamma_2} \;\middle|\; |L_1 f(t,y) - L_1 f(t,z)| \leq K|y-z| \text{ for all } t \in [0,T], y,z \in \mathbb{R} \right\}.$$

Moreover, let

$$\mathcal{J}_K^q = \{\eta : \Omega \to \mathbb{R} \mid \eta \text{ is } \Sigma_0 - \text{measurable}, \mathbb{E}|\eta|^{2q} \leq K\}.$$

For $\gamma_1, \gamma_2 \in (0,1]$, $q \in [2,+\infty)$, $K \in (0,+\infty)$ we consider the following class of admissible input data

$$\mathcal{F}(\gamma_1,\gamma_2,q,K) = \mathcal{A}_K^{\gamma_1} \times \mathcal{B}_K^{\gamma_2} \times \mathcal{J}_K^q.$$

For all $(a,b,\eta) \in \mathcal{F}(\gamma_1,\gamma_2,q,K)$ Eq. (1) has a unique strong solution $\{X(t)\}_{t\in[0,T]}$, that is adapted to $\{\Sigma_t\}_{t\in[0,T]}$, see, for example, [13]. The numbers $T, K, q, \gamma_1, \gamma_2$ will be called parameters of the class $\mathcal{F}(\gamma_1,\gamma_2,q,K)$. Except for T the parameters are, in general, not known and the algorithms presented later on will not use them as input parameters.

Under some minor modifications, we recall from [10] a model of computation under inexact information about a's and b's. To do that we need to introduce the following auxiliary classes:

$$\mathcal{K}^1 = \{p : [0,T] \times \mathbb{R} \to \mathbb{R} \mid p - \text{ Borel measurable }, |p(t,y)| \leq 1 + |y|, t \in [0,T], y \in \mathbb{R}\},$$

$$\mathcal{K}^1_{\mathrm{Lip}} = \{p \in \mathcal{K}^1 \mid |p(t,y) - p(t,z)| \leq |y-z|, t \in [0,T], y,z \in \mathbb{R}\},$$

and

$$\mathcal{K}^2 = \{p \in \mathcal{K}^1 \mid |p(t,y)| \leq 1, t \in [0,T], y \in \mathbb{R}\},$$

see also [11]. The classes $\mathcal{K}^1_{\mathrm{Lip}}$, $\mathcal{K}^2$ are nonempty and contain constant functions. (This is an important fact from the point of view of lower error bounds, see [10].) Let δ_1, $\delta_2 \in [0,1]$. We refer to δ_1, δ_2 as to precision parameters. For $a \in \mathcal{A}_K^{\gamma_1}$ we define the following class of corrupted drift coefficients

$$V_a(\delta_1) = \{\tilde{a} \mid \exists_{p_a \in \mathcal{K}^1} \, : \, \tilde{a} = a + \delta_1 \cdot p_a\},$$

while for $b \in \mathcal{B}_K^{\gamma_2}$ we consider the following two classes of corrupted diffusion coefficients

$$V_b^1(\delta_2) = \{\tilde{b} \mid \exists_{p_b \in \mathcal{K}^1_{\mathrm{Lip}}} \, : \, \tilde{b} = b + \delta_2 \cdot p_b\},$$

and

$$V_b^2(\delta_2) = \{\tilde{b} \mid \exists_{p_b \in \mathcal{K}^2} \, : \, \tilde{b} = b + \delta_2 \cdot p_b\}.$$

Note that we impose more smoothness for corrupting functions p_b's than for p_a's. This is due to some technicalities, see Remark 3. We have that $\{a\} = V_a(0) \subset V_a(\delta_1) \subset V_a(\delta_1')$ for $0 \leq \delta_1 \leq \delta_1' \leq 1$, and $\{b\} = V_b^i(0) \subset V_b^i(\delta_2) \subset V_b^i(\delta_2')$ for $0 \leq \delta_2 \leq \delta_2' \leq 1$, for $i = 1, 2$.

For $(a,b,\eta) \in \mathcal{F}(\gamma_1,\gamma_2,q,K)$ let $(\tilde{a},\tilde{b}) \in V_a(\delta_1) \times (V_b^1(\delta_2) \cup V_b^2(\delta_2))$. We assume that the approximation method is based on discrete noisy information about (a,b) and exact information about W, and η. Hence, a vector of noisy information has the following form

$$\begin{aligned} \mathcal{N}(\tilde{a},\tilde{b},\eta,W) = \Big[\, & \tilde{a}(\xi_0,y_0), \tilde{a}(\xi_1,y_1), \ldots, \tilde{a}(\xi_{i_1-1},y_{i_1-1}), \\ & \tilde{b}(t_0,z_0), \tilde{b}(t_1,z_1), \ldots, \tilde{b}(t_{i_1-1},z_{i_1-1}), \\ & \tilde{b}(t_0,u_0), \tilde{b}(t_1,u_1), \ldots, \tilde{b}(t_{i_1-1},u_{i_1-1}), \\ & W(s_0), W(s_1), \ldots, W(s_{i_2-1}), \eta \, \Big], \end{aligned}$$

where $i_1, i_2 \in \mathbb{N}$ and $(\xi_0,\xi_1,\ldots,\xi_{i_1-1})$ is a random vector on $(\Omega,\Sigma,\mathbb{P})$ which takes values in $[0,T]^{i_1}$. We assume that the σ-fields $\sigma(\xi_0,\xi_1,\ldots,\xi_{i_1-1})$ and Σ_∞ are independent. Moreover, $t_0,t_1,\ldots,t_{i_1-1} \in [0,T]$ and $s_0,s_1,\ldots,s_{i_2-1} \in [0,T]$ are given time points. We assume that $t_i \neq t_j$, $s_i \neq s_j$ for all $i \neq j$. The evaluation points y_j, z_j, u_j for the spatial variables y,z of $a(\cdot,y)$, $b(\cdot,z)$, and $b(\cdot,u)$ can be given in adaptive way with respect to (a,b,η) and W. This means that there exist Borel measurable mappings $\psi_0 : \mathbb{R}^{i_2} \times \mathbb{R} \to \mathbb{R}^3$, $\psi_j : \mathbb{R}^j \times \mathbb{R}^j \times \mathbb{R}^j \times \mathbb{R}^{i_2} \times \mathbb{R} \to \mathbb{R}^3$, $j = 1,2,\ldots,i_1-1$, such that the successive points y_j, z_j are computed in the following way:

$$(y_0,z_0,u_0) = \psi_0\Big(W(s_0), W(s_1), \ldots, W(s_{i_2-1}), \eta\Big),$$

where

$$
\begin{aligned}
(y_j, z_j, u_j) = \psi_j \Big(& \tilde{a}(\xi_0, y_0), \tilde{a}(\xi_1, y_1), \dots, \tilde{a}(\xi_{j-1}, y_{j-1}), \\
& \tilde{b}(t_0, z_0), \tilde{b}(t_1, z_1), \dots, \tilde{b}(t_{j-1}, z_{j-1}), \\
& \tilde{b}(t_0, u_0), \tilde{b}(t_1, u_1), \dots, \tilde{b}(t_{j-1}, u_{j-1}), \\
& W(s_0), W(s_1), \dots, W(s_{i_2-1}), \eta \Big),
\end{aligned}
$$

for $j = 1, 2, \dots, i_1 - 1$. The total number of (noisy) evaluations of a, b and W is equal to $l = 3i_1 + i_2$.

Any algorithm $\mathcal{A}$ using $\mathcal{N}(\tilde{a}, \tilde{b}, \eta, W)$, that computes approximation to $X(T)$ is of the form

$$\mathcal{A}(\tilde{a}, \tilde{b}, \eta, W, \delta_1, \delta_2) = \varphi(\mathcal{N}(\tilde{a}, \tilde{b}, \eta, W)), \tag{2}$$

for some Borel measurable mapping $\varphi : \mathbb{R}^{3i_1+i_2+1} \to \mathbb{R}$. For a given $n \in \mathbb{N}$ we denote by Φ_n a class of all algorithms of the form (2) for which the total number of evaluations l is at most n.

Let $\mathcal{F} \subset \mathcal{F}(\gamma_1, \gamma_2, q, K)$. (In this paper we use certain subclasses of $\mathcal{F}(\gamma_1, \gamma_2, q, K)$ when establishing the lower bounds, see (61) and (62).) For a fixed $(a, b, \eta) \in \mathcal{F}$ the error of $\mathcal{A} \in \Phi_n$ is defined in the following way

$$e^{(q)}(\mathcal{A}, a, b, \eta, W, V^i, \delta_1, \delta_2) = \sup_{(\tilde{a}, \tilde{b}) \in V_a(\delta_1) \times V_b^i(\delta_2)} \|X(T) - \mathcal{A}(\tilde{a}, \tilde{b}, \eta, W, \delta_1, \delta_2)\|_q.$$

for $i = 1, 2$, where $X(T) = X(a, b, \eta, W)(T)$. Hence, we are considering the worst error with respect to any $(\tilde{a}, \tilde{b})$ that can be given to us for a fixed (a, b, η). The worst-case error of the algorithm $\mathcal{A}$ in the class $\mathcal{F}$ is defined as

$$e^{(q)}(\mathcal{A}, \mathcal{F}, W, V^i, \delta_1, \delta_2) = \sup_{(a,b,\eta) \in \mathcal{F}} e^{(q)}(\mathcal{A}, a, b, \eta, W, V^i, \delta_1, \delta_2), \tag{3}$$

see [3] and [14]. Finally, we define the nth minimal error as follows

$$e_n^{(q)}(\mathcal{F}, W, V^i, \delta_1, \delta_2) = \inf_{\mathcal{A} \in \Phi_n} e^{(q)}(\mathcal{A}, \mathcal{F}, W, V^i, \delta_1, \delta_2), \;\; i = 1, 2.$$

Our aim is to find possibly sharp bounds on the nth minimal error $e_n^{(q)}(\mathcal{F}, W, V^i, \delta_1, \delta_2)$, i.e., the lower and upper bounds which match up to constants. We are also interested in defining an algorithm for which the infimum in $e_n^{(q)}(\mathcal{F}, W, V^i, \delta_1, \delta_2)$ is asymptotically attained.

Unless otherwise stated, all constants appearing in this paper (including those in the "O", "Ω", and "Θ" notation) will only depend on the parameters of the respective classes. Furthermore, the same symbol may be used for different constants.

Remark 1. Let $\alpha : [0, T] \to \mathbb{R}$ be a γ_1-Hölder continuous function, while let $\beta : [0, T] \to \mathbb{R}$ a γ_2-Hölder continuous function. Moreover, let $G, H : \mathbb{R} \to \mathbb{R}$ be any $C^2(\mathbb{R})$ functions that satisfy the following conditions:

- G, G', H, H' are globally Lipschitz continuous,
- H is bounded on $\mathbb{R}$.

Then there exists $K \in (0, +\infty)$ such that $a(t, y) = G(y \cdot \alpha(t))$ belongs to $\mathcal{A}_K^{\gamma_1}$, while $b(t, y) = H(y \cdot \beta(t))$ belongs to $\mathcal{B}_K^{\gamma_2}$.

Remark 2. It is worth mentioning that the proposed computation and error setting includes the phenomenon of lowering precision of computations. Namely, we can model relative roundoff errors by considering disturbing functions p_f, $f \in \{a, b\}$, of the form

$$p_f(t, y) = \alpha(t, y) \cdot f(t, y), \;\; (t, y) \in [0, T] \times \mathbb{R}, \tag{4}$$

for some function α that is Borel measurable and bounded on $[0, T] \times \mathbb{R}$. That is a frequent case for efficient computations using both CPUs and GPUs. An example could be the current state-of-the-art GPU — NVIDIA Tesla V100, which performance behaves as follows — 7 TeraFLOPS for *double precision*, 14 TeraFLOPS for *single precision*, and up to 112 TeraFLOPS for *half precision* of very specific type (repeatable operations of matrix multiplications and additions). We refer to [11] where Monte Carlo simulations were performed on GPUs.

3. Randomized derivative-free Milstein algorithm for noisy information

Below we define randomized derivative-free Milstein algorithm in presence of informational noise for a and b. Let $n \in \mathbb{N}$ and let

$$t_i = ih, \quad h = T/n, \quad i = 0, 1, \dots, n, \tag{5}$$

be the equidistant discretization on $[0,T]$. Moreover, we take

$$\Delta W_i = W(t_{i+1}) - W(t_i),$$
$$\mathcal{I}_{t_i,t}(W,W) = \int_{t_i}^{t}\int_{t_i}^{s} \mathrm{d}W(u)\mathrm{d}W(s) = \frac{1}{2}\Big((W(t)-W(t_i))^2-(t-t_i)\Big),$$

for $t\in[t_i,t_{i+1}]$, $i=0,1,\ldots,n-1$. Let $\{\xi_i\}_{i=0}^{n-1}$ be independent random variables on the probability space $(\Omega,\Sigma,\mathbb{P})$, such that the σ-fields $\sigma(\xi_0,\xi_1,\ldots,\xi_{n-1})$ and Σ_∞ are independent, with ξ_i being uniformly distributed on $[t_i,t_{i+1}]$. Then for any $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$, $(\tilde a,\tilde b)\in V_a(\delta_1)\times(V_b^1(\delta_2)\cup V_b^2(\delta_2))$ we set

$$\begin{cases} \bar X_n^{df-RM}(0) &= \eta,\\ \bar X_n^{df-RM}(t_{i+1}) &= \bar X_n^{df-RM}(t_i) + \tilde a(\xi_i,\bar X_n^{df-RM}(t_i))\cdot\frac{T}{n} + \tilde b(t_i,\bar X_n^{df-RM}(t_i))\cdot\Delta W_i\\ &\quad + \mathcal{L}_{1,h}\tilde b(t_i,\bar X_n^{df-RM}(t_i))\cdot\mathcal{I}_{t_i,t_{i+1}}(W,W), \end{cases} \tag{6}$$

for $i=0,1,\ldots,n-1$, where

$$\mathcal{L}_{1,h}\tilde b(t_i,\bar X_n^{df-RM}(t_i)) = \tilde b(t_i,\bar X_n^{df-RM}(t_i))\cdot\frac{\tilde b(t_i,\bar X_n^{df-RM}(t_i)+h)-\tilde b(t_i,\bar X_n^{df-RM}(t_i))}{h}.$$

The algorithm $\bar{\mathcal{A}}_n^{df-RM}$ is defined as

$$\bar{\mathcal{A}}_n^{df-RM}(\tilde a,\tilde b,\eta,W,\delta_1,\delta_2) := \bar X_n^{df-RM}(T). \tag{7}$$

In case of exact information (i.e., $\delta_1=\delta_2=0$) we write X_n^{df-RM} and $\mathcal{A}_n^{df-RM}$ instead of $\bar X_n^{df-RM}$, and $\bar{\mathcal{A}}_n^{df-RM}$, respectively. The total number of evaluations of a, b, and W used for computing $\bar{\mathcal{A}}_n^{df-RM}$ is $4n$. Therefore, $\bar{\mathcal{A}}_n^{df-RM}\in\Phi_{4n}$. Moreover, the combinatorial cost consists of $O(n)$ arithmetic operations.

In the following theorem we state the upper bounds on the error of randomized derivative-free Milstein scheme under noisy information about a and b.

Theorem 1.

(i) *There exists a positive constant C, depending only on the parameters of the class $\mathcal{F}(\gamma_1,\gamma_2,q,K)$, such that for all $n\in\mathbb{N}$, $\delta_1,\delta_2\in[0,1]$, $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$, $(\tilde a,\tilde b)\in V_a(\delta_1)\times V_b^1(\delta_2)$, we have*

$$\|X(a,b,\eta,W)(T)-\bar{\mathcal{A}}_n^{df-RM}(\tilde a,\tilde b,\eta,W,\delta_1,\delta_2)\|_q \le C\cdot\Big(n^{-\min\{\frac12+\gamma_1,\gamma_2\}}+\delta_1+\delta_2\Big).$$

(ii) *There exist positive constants C_1,C_2,C_3, depending only on the parameters of the class $\mathcal{F}(\gamma_1,\gamma_2,q,K)$ and q, such that for all $n\in\mathbb{N}$, $\delta_1,\delta_2\in[0,1]$, $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$, $(\tilde a,\tilde b)\in V_a(\delta_1)\times V_b^2(\delta_2)$, we have*

$$\|X(a,b,\eta,W)(T)-\bar{\mathcal{A}}_n^{df-RM}(\tilde a,\tilde b,\eta,W,\delta_1,\delta_2)\|_q \le C_1\cdot n^{-\min\{\frac12+\gamma_1,\gamma_2\}}$$
$$+C_2\cdot e^{C_3(\delta_2 n^{1/2})^q}\cdot(1+\delta_2 n^{1/2})\cdot\Big(\delta_1+\delta_2 n^{1/2}+n^{-3/2}\Big).$$

The aim of this section is to justify Theorem 1. Before we do that we need to prove several auxiliary results concerning, in particular, the upper bounds on the error of the following time-continuous version of randomized derivative-free Milstein algorithm $\bar{\mathcal{A}}_n^{df-RM}$ in presence of noise. Namely, let us take

$$\begin{cases} \tilde{\bar X}_n^{df-RM}(0) &= \eta,\\ \tilde{\bar X}_n^{df-RM}(t) &= \tilde{\bar X}_n^{df-RM}(t_i) + \tilde a(\xi_i,\tilde{\bar X}_n^{df-RM}(t_i))\cdot(t-t_i)\\ &\quad +\tilde b(t_i,\tilde{\bar X}_n^{df-RM}(t_i))\cdot(W(t)-W(t_i))\\ &\quad +\mathcal{L}_{1,h}\tilde b(t_i,\tilde{\bar X}_n^{df-RM}(t_i))\cdot\mathcal{I}_{t_i,t}(W,W), \end{cases} \tag{8}$$

for $t\in[t_i,t_{i+1}]$ and $i=0,1,\ldots,n-1$. In the case of exact information we write $\tilde X_n^{df-RM}=\{\tilde X_n^{df-RM}(t)\}_{t\in[0,T]}$ instead of $\tilde{\bar X}_n^{df-RM}=\{\tilde{\bar X}_n^{df-RM}(t)\}_{t\in[0,T]}$. It holds $\tilde{\bar X}_n^{df-RM}(t_i)=\bar X_n^{df-RM}(t_i)$ for $i=0,1,\ldots,n$. Hence, it is sufficient to analyze the error of $\tilde{\bar X}_n^{df-RM}$. We also extend the filtration $\{\Sigma_t\}_{t\ge0}$ in the same way as in [10]. Namely, let $\mathcal{G}^n=\sigma(\xi_0,\xi_1,\ldots,\xi_{n-1})$ and $\tilde\Sigma_t^n=\sigma\Big(\Sigma_t\cup\mathcal{G}^n\Big)$, $t\ge0$. Since the σ-fields Σ_∞ and $\mathcal{G}^n$ are independent, the process W is still a one-dimensional Wiener process on $(\Omega,\Sigma,\mathbb{P})$ with respect to $\{\tilde\Sigma_t^n\}_{t\ge0}$. In the sequel we will consider stochastic Itô integrals with respect to W of processes that are adapted to the filtration $\{\tilde\Sigma_t^n\}_{t\ge0}$. In particular, the following technical lemma assures suitable measurability of the process $\tilde{\bar X}_n^{df-RM}=\{\tilde{\bar X}_n^{df-RM}(t)\}_{t\in[0,T]}$ with respect to $\{\tilde\Sigma_t^n\}_{t\ge0}$.

Lemma 1. *Let $n\in\mathbb{N}$, $\delta_1,\delta_2\in[0,1]$, $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$ and $(\tilde a,\tilde b)\in V_a(\delta_1)\times(V_b^1(\delta_2)\cup V_b^2(\delta_2))$. Then the process $\tilde{\bar X}_n^{df-RM}=\{\tilde{\bar X}_n^{df-RM}(t)\}_{t\in[0,T]}$ is progressively measurable with respect to the filtration $\{\tilde\Sigma_t^n\}_{t\ge0}$.*

The lemma above follows from induction and Proposition 1.13 in [13]. Hence, we skip its proof.

In order to justify Theorem 1 we proceed as follows. First, in Section 3.1 we investigate the error of the randomized version of the classical Milstein algorithm when information about a and b is exact. Then, in Section 3.2 we show the upper bounds for the derivative-free version of randomized Milstein scheme also for exact information about a and b. Finally, combining the results obtained for these two methods we show the upper bounds on the error of $\bar{X}^{df-RM}$ in the presence of informational noise.

Remark 3. It is natural to ask about a version of Theorem 1 when corrupting functions p_b are from $\mathcal{K}^1$, as it is for p_a's. However, in this case we were unable to show any nontrivial upper bound for the algorithm $\bar{\mathcal{A}}_n^{df-RM}$. It turns out that for $p_b \in \mathcal{K}^1$ the function $(t,y) \to \mathcal{L}_{1,h}\tilde{b}(t,y)$ might be of super-linear growth with respect to y. Hence, we conjecture that some modification of the scheme $\bar{\mathcal{A}}_n^{df-RM}$ is needed in order to obtain analogous bounds as in Theorem 1. We postpone this problem to our future work.

Remark 4. In [11] the authors consider approximate stochastic Itô integration in the case when the values of the Wiener process are corrupted by informational noise. Preliminary estimates suggest that direct application of the techniques used in [11] to approximation of SDEs, under inexact information about W, is not possible. Therefore, further investigation in that direction is needed.

3.1. Performance of randomized Milstein algorithm for exact information

By randomizing evaluations of drift coefficient a in the classical Milstein scheme, we arrive at the following randomized Milstein algorithm. Take

$$\begin{cases} X_n^{RM}(0) &= \eta, \\ X_n^{RM}(t_{i+1}) &= X_n^{RM}(t_i) + a(\xi_i, X_n^{RM}(t_i)) \cdot \frac{T}{n} + b(t_i, X_n^{RM}(t_i)) \cdot \Delta W_i \\ & \quad + L_1 b(t_i, X_n^{RM}(t_i)) \cdot \mathcal{I}_{t_i,t_{i+1}}(W,W), \end{cases} \tag{9}$$

for $i = 0, 1, \ldots, n-1$. The algorithm $\mathcal{A}_n^{RM}$ is defined as

$$\mathcal{A}_n^{RM}(a,b,\eta,W,0,0) := X_n^{RM}(T). \tag{10}$$

Note that $\mathcal{A}_n^{RM} \notin \Phi_n$, since it uses values of the partial derivative of b. We refer to $\mathcal{A}_n^{RM}$ as to an auxiliary method that helps us to estimate the error of $\bar{\mathcal{A}}_n^{df-RM}$.

In order to investigate the error of the method $\mathcal{A}_n^{RM}$ we define the following time-continuous version of the scheme X_n^{RM} as follows:

$$\begin{cases} \tilde{X}_n^{RM}(0) = & \eta, \\ \tilde{X}_n^{RM}(t) = & \tilde{X}_n^{RM}(t_i) + a(\xi_i, \tilde{X}_n^{RM}(t_i)) \cdot (t - t_i) \\ & + b(t_i, \tilde{X}_n^{RM}(t_i)) \cdot (W(t) - W(t_i)) \\ & + L_1 b(t_i, \tilde{X}_n^{RM}(t_i)) \cdot \mathcal{I}_{t_i,t}(W,W), \end{cases} \tag{11}$$

for $t \in [t_i, t_{i+1}]$ and for $i = 0, 1, \ldots, n-1$. We have that $\tilde{X}_n^{RM}(t_i) = X_n^{RM}(t_i)$ for all $0 \le i \le n$. Hence, it is sufficient to analyze the error of time-continuous version of the algorithm. Moreover, for the process $\{\tilde{X}_n^{RM}(t)\}_{t\in[0,T]}$ the following version of Lemma 1 holds.

Lemma 2. *Let $n \in \mathbb{N}$, $(a,b,\eta) \in \mathcal{F}(\gamma_1,\gamma_2,q,K)$. Then the process $\tilde{X}_n^{RM} = \{\tilde{X}_n^{RM}(t)\}_{t\in[0,T]}$ is progressively measurable with respect to the filtration $\{\tilde{\Sigma}_t^n\}_{t\ge 0}$.*

We have the following result for the algorithm $\mathcal{A}_n^{RM}$.

Proposition 1. *There exists a positive constant C, depending only on the parameters of the class $\mathcal{F}(\gamma_1,\gamma_2,q,K)$, such that for all $n \in \mathbb{N}$ and all $(a,b,\eta) \in \mathcal{F}(\gamma_1,\gamma_2,q,K)$ we have*

$$\sup_{t\in[0,T]} \|X(t) - \tilde{X}_n^{RM}(t)\|_q \le Cn^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}, \tag{12}$$

and, in particular,

$$\|X(T) - \mathcal{A}_n^{RM}(a,b,\eta,W,0,0)\|_q \le Cn^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}.$$

Proof. We show the upper bound for $\sup_{t\in[0,T]} \|X(t) - \tilde{X}_n^{RM}(t)\|_q$, from which the desired result follows.

The solution X can be expressed in the following way

$$X(t) = \eta + A(t) + B(t),$$
$$A(t) = \int_0^t \sum_{i=0}^{n-1} a(s, X(s))\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s,$$
$$B(t) = \int_0^t \sum_{i=0}^{n-1} b(s, X(s))\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}W(s).$$

Let us denote by

$$U_i = (t_i, \tilde{X}_n^{RM}(t_i)), \quad V_i = (\xi_i, \tilde{X}_n^{RM}(t_i)).$$

Then, for the process $\{\tilde{X}_n^{RM}(t)\}_{t\in[0,T]}$ we can write

$$\tilde{X}_n^{RM}(t) = \eta + \tilde{A}_n^{RM}(t) + \tilde{B}_n^{RM}(t),$$
$$\tilde{A}_n^{RM}(t) = \int_0^t \sum_{i=0}^{n-1} a(V_i)\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s,$$
$$\tilde{B}_n^{RM}(t) = \int_0^t \sum_{i=0}^{n-1} \Big(b(U_i) + \int_{t_i}^s L_1 b(U_i)\mathrm{d}W(u)\Big)\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}W(s).$$

Note that the process

$$\Big\{\sum_{i=0}^{n-1} \Big(b(U_i) + \int_{t_i}^s L_1 b(U_i)\mathrm{d}W(u)\Big)\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\Big\}_{s\in[0,T]}$$

is adapted to $\{\tilde{\Sigma}_t^n\}_{t\in[0,T]}$ and has cádlág paths. Hence, the Itô integral above is well-defined.

We have that

$$\mathbb{E}|A(t) - \tilde{A}_n^{RM}(t)|^q \leq C\sum_{k=1}^{3} \mathbb{E}|\tilde{A}_{n,k}^{RM}(t)|^q,$$

where

$$\mathbb{E}|\tilde{A}_{n,1}^{RM}(t)|^q = \mathbb{E}\Big|\int_0^t \sum_{i=0}^{n-1}(a(s,X(s)) - a(s,X(t_i)))\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s\Big|^q,$$
$$\mathbb{E}|\tilde{A}_{n,2}^{RM}(t)|^q = \mathbb{E}\Big|\int_0^t \sum_{i=0}^{n-1}(a(s,X(t_i)) - a(\xi_i,X(t_i)))\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s\Big|^q,$$
$$\mathbb{E}|\tilde{A}_{n,3}^{RM}(t)|^q = \mathbb{E}\Big|\int_0^t \sum_{i=0}^{n-1}(a(\xi_i,X(t_i)) - a(V_i))\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s\Big|^q.$$

Since for all $s \in [0,T]$

$$\Big(\sum_{i=0}^{n-1} |X(t_i) - \tilde{X}_n^{RM}(t_i)|\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\Big)^q = \sum_{i=0}^{n-1} |X(t_i) - \tilde{X}_n^{RM}(t_i)|^q\cdot \mathbb{1}_{[t_i,t_{i+1})}(s),$$

we get, by using Lipschitz continuity of a and Hölder inequality, that

$$\begin{aligned}\mathbb{E}|\tilde{A}_{n,3}^{RM}(t)|^q &\leq \mathbb{E}\Big(\int_0^t \sum_{i=0}^{n-1} |a(\xi_i,X(t_i)) - a(\xi_i,\tilde{X}_n^{RM}(t_i))|\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s\Big)^q \\ &\leq K^q T^{q-1}\mathbb{E}\int_0^t \Big(\sum_{i=0}^{n-1} |X(t_i) - \tilde{X}_n^{RM}(t_i)|\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\Big)^q \mathrm{d}s \\ &= C\int_0^t \sum_{i=0}^{n-1} \mathbb{E}|X(t_i) - \tilde{X}_n^{RM}(t_i)|^q\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s. \end{aligned} \tag{13}$$

For any $(f,t) \in \{a,b\} \times [t_i,t_{i+1}]$ the function

$$\mathbb{R} \ni y \to f(t,y) \in \mathbb{R} \tag{14}$$

is in $C^2(\mathbb{R})$. Consider the process $\Big(f(t,X(s))\Big)_{s\in[t_i,t_{i+1}]}$ and apply the Itô formula (see, for example, Theorem 6.2, page 32 in [15]) to the function (14), and to the solution process $(X(s))_{s\in[t_i,t_{i+1}]}$. This gives the following (parametric) version of the Itô formula

$$f(t,X(s))-f(t,X(t_i))=\int_{t_i}^{s}\alpha(f,t,u)\mathrm{d}u+\int_{t_i}^{s}\beta(f,t,u)\mathrm{d}W(u), \tag{15}$$

where

$$\alpha(f,t,u)=\frac{\partial f}{\partial y}(t,X(u))\cdot a(u,X(u))+\frac{1}{2}\frac{\partial^2 f}{\partial y^2}(t,X(u))\cdot b^2(u,X(u)), \tag{16}$$

$$\beta(f,t,u)=\frac{\partial f}{\partial y}(t,X(u))\cdot b(u,X(u)), \tag{17}$$

for $f\in\{a,b\}$ and all $t,s,u\in[t_i,t_{i+1}]$. In particular, for all $s\in[t_i,t_{i+1}]$ and by taking $t=s$ or $t=t_i$ we obtain, respectively, that

$$a(s,X(s))-a(s,X(t_i))=\int_{t_i}^{s}\alpha(a,s,u)\mathrm{d}u+\int_{t_i}^{s}\beta(a,s,u)\mathrm{d}W(u), \tag{18}$$

$$b(t_i,X(s))-b(t_i,X(t_i))=\int_{t_i}^{s}\alpha(b,t_i,u)\mathrm{d}u+\int_{t_i}^{s}\beta(b,t_i,u)\mathrm{d}W(u). \tag{19}$$

By Lemma 4 we get that

$$|\alpha(f,t,u)|\le C(1+|X(u)|^2),$$

$$|\beta(f,t,u)|\le C(1+|X(u)|), \tag{20}$$

and

$$|\beta(a,t_1,u)-\beta(a,t_2,u)|\le C(1+|X(u)|^2)\cdot|t_1-t_2|^{\gamma_1},$$

for $f\in\{a,b\}$, $t,t_1,t_2,u\in[t_i,t_{i+1}]$. Then, we can write that

$$\mathbb{E}|\tilde{A}^{RM}_{n,1}(t)|^q\le C\Big(\mathbb{E}|\tilde{M}^{RM}_{n,1}(t)|^q+\mathbb{E}|\tilde{M}^{RM}_{n,2}(t)|^q\Big),$$

where

$$\mathbb{E}|\tilde{M}^{RM}_{n,1}(t)|^q=\mathbb{E}\Big|\int_0^t\sum_{i=0}^{n-1}\Big(\int_{t_i}^{s}\alpha(a,s,u)\mathrm{d}u\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s\Big|^q$$

$$\mathbb{E}|\tilde{M}^{RM}_{n,2}(t)|^q=\mathbb{E}\Big|\int_0^t\sum_{i=0}^{n-1}\Big(\int_{t_i}^{s}\beta(a,s,u)\mathrm{d}W(u)\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s\Big|^q.$$

Note that for almost all $\omega\in\Omega$ the function

$$[t_i,t_{i+1}]\times[t_i,t_{i+1}]\ni(s,u)\to\alpha(a,s,u)(\omega)\in\mathbb{R}$$

is continuous. Hence, parametric indefinite Riemann integral $\left\{\int_{t_i}^{s}\alpha(a,s,u)\,\mathrm{d}u\right\}_{s\in[t_i,t_{i+1}]}$ has almost all trajectories continuous. Moreover, by (15) for all $s\in[t_i,t_{i+1}]$ it holds that

$$\int_{t_i}^{s}\beta(a,s,u)\mathrm{d}W(u)=a(s,X(s))-a(s,X(t_i))-\int_{t_i}^{s}\alpha(a,s,u)\,\mathrm{d}u.$$

Thus parametric indefinite stochastic Itô integral $\left\{\int_{t_i}^{s}\beta(a,s,u)\mathrm{d}W(u)\right\}_{s\in[t_i,t_{i+1}]}$ also has continuous modification. Thereby, $\mathbb{E}|\tilde{M}^{RM}_{n,1}(t)|^q$ and $\mathbb{E}|\tilde{M}^{RM}_{n,2}(t)|^q$ are well defined.

We have that

$$\mathbb{E}|\tilde{M}^{RM}_{n,1}(t)|^q\le T^{q-1}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\Big((s-t_i)^{q-1}\cdot\int_{t_i}^{s}\mathbb{E}|\alpha(a,s,u)|^q\mathrm{d}u\Big)\mathrm{d}s\le Cn^{-q}.$$

Moreover, for any $t \in [0, T]$ there exists $l \in \{0, \dots, n-1\}$ such that $t \in [t_l, t_{l+1}]$ and

$$\begin{aligned}\mathbb{E}|\tilde{M}^{RM}_{n,2}(t)|^q \le C\mathbb{E}\Big|\int_0^{t_l} \sum_{i=0}^{n-1}\Big(\int_{t_i}^s (\beta(a,s,u)-\beta(a,t_i,u))\mathrm{d}W(u)\Big)\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s\Big|^q \\ +C\mathbb{E}\Big|\int_0^{t_l} \sum_{i=0}^{n-1}\Big(\int_{t_i}^s \beta(a,t_i,u)\mathrm{d}W(u)\Big)\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s\Big|^q \\ +C\mathbb{E}\Big|\int_{t_l}^{t}\Big(\int_{t_l}^s \beta(a,s,u)\mathrm{d}W(u)\Big)\,\mathrm{d}s\Big|^q. \end{aligned} \tag{21}$$

By using the Hölder and Burkholder inequalities (see, for example, Theorem 2.2 in [12]), together with Lemma 8, we obtain

$$\begin{aligned}&\mathbb{E}\Big|\int_0^{t_l} \sum_{i=0}^{n-1}\Big(\int_{t_i}^s (\beta(a,s,u)-\beta(a,t_i,u))\mathrm{d}W(u)\Big)\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s\Big|^q \\ &\le T^{q-1}\mathbb{E}\int_0^T\Big(\sum_{i=0}^{n-1}\Big|\int_{t_i}^s (\beta(a,s,u)-\beta(a,t_i,u))\mathrm{d}W(u)\Big|\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\Big)^q\,\mathrm{d}s \\ &= T^{q-1}\mathbb{E}\sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}}\Big(\sum_{i=0}^{n-1}\Big|\int_{t_i}^s (\beta(a,s,u)-\beta(a,t_i,u))\mathrm{d}W(u)\Big|\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\Big)^q\,\mathrm{d}s \\ &= T^{q-1}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\mathbb{E}\Big|\int_{t_i}^s (\beta(a,s,u)-\beta(a,t_i,u))\mathrm{d}W(u)\Big|^q\,\mathrm{d}s \\ &\le T^{q-1}c_q\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\Big((s-t_i)^{(q/2)-1}\cdot\mathbb{E}\int_{t_i}^s|\beta(a,s,u)-\beta(a,t_i,u)|^q du\Big)\,\mathrm{d}s \\ &\le C\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\Big((s-t_i)^{(q/2)-1}\cdot\int_{t_i}^s(1+\mathbb{E}|X(u)|^{2q})\cdot(s-t_i)^{q\gamma_1}\,\mathrm{d}u\Big)\,\mathrm{d}s \\ &\le Cn^{-q(\frac{1}{2}+\gamma_1)}, \end{aligned} \tag{22}$$

and

$$\begin{aligned}&\mathbb{E}\Big|\int_{t_l}^{t}\Big(\int_{t_l}^s \beta(a,s,u)\mathrm{d}W(u)\Big)\,\mathrm{d}s\Big|^q \\ &\le C\Big(\frac{T}{n}\Big)^{q-1}\cdot\int_{t_l}^{t_{l+1}}\Big((s-t_l)^{(q/2)-1}\cdot\mathbb{E}\int_{t_l}^s|\beta(a,s,u)|^q\,\mathrm{d}u\Big)\,\mathrm{d}s \\ &\le Cn^{-q+1}\int_{t_l}^{t_{l+1}}\Big((s-t_l)^{(q/2)-1}\cdot\int_{t_l}^s(1+\mathbb{E}|X(u)|^q)\,\mathrm{d}u\Big)\,\mathrm{d}s \le Cn^{-3q/2}. \end{aligned} \tag{23}$$

Let

$$Y_i = \int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^s \beta(a,t_i,u)\mathrm{d}W(u)\Big)\mathrm{d}s \quad i = 0, 1, \dots, n-1,$$

and

$$Z_k = \sum_{i=0}^{k} Y_i \qquad k = 0, 1, \dots, n-1,$$

where $Z_{-1} := 0$. Therefore,

$$\mathbb{E}\Big|\int_0^{t_l}\sum_{i=0}^{n-1}\Big(\int_{t_i}^s \beta(a,t_i,u)\mathrm{d}W(u)\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s\Big|^q = \mathbb{E}|Z_{l-1}|^q,\ \ l\in\{0,1,\dots,n-1\}. \tag{24}$$

Notice that the process $\Big\{\int_{t_i}^s \beta(a,t_i,u)\mathrm{d}W(u)\Big\}_{s\in[t_i,t_{i+1}]}$ is adapted to the filtration $\{\Sigma_s\}_{s\in[t_i,t_{i+1}]}$ and has continuous paths. Hence, it is progressively measurable. This and Fubini theorem imply that Y_i is $\Sigma_{t_{i+1}}$-measurable. Furthermore, let $\mathcal{G}_i := \Sigma_{t_{i+1}}$, $i \in \{0, 1, \dots, n-1\}$. Then $\{\mathcal{G}_i\}_{i\in\{0,1,\dots,n-1\}}$ is a filtration and Z_k is $\mathcal{G}_k$ measurable for each $k = 0, 1, \dots, n-1$. By using Fubini theorem for conditional expectation (see, for example, [16]) and martingale property of Itô integral we have

$$\mathbb{E}(Z_{k+1}-Z_k|\mathcal{G}_k)=\int_{t_{k+1}}^{t_{k+2}}\mathbb{E}\left(\int_{t_{k+1}}^{s}\beta(a,t_{k+1},u)\mathrm{d}W(u)\middle|\mathcal{G}_k\right)\mathrm{d}s=0,$$

for $k=0,1,\dots,n-2$. This implies that $\{Z_k,\mathcal{G}_k\}_{k\in\{0,1,\dots,n-1\}}$ is a discrete-time martingale. Therefore, by using the discrete version of the Burkholder inequality we have for every $k\in\{0,1,\dots,n-1\}$ that

$$\mathbb{E}|Z_k|^q\le C_q^q\mathbb{E}\Big(\sum_{i=0}^{k}Y_i^2\Big)^{q/2}\le C_q^q n^{q/2-1}\sum_{i=0}^{n-1}\mathbb{E}|Y_i|^q.$$

Moreover, analogously as in (23) we get that

$$\mathbb{E}|Y_i|^q\le Cn^{-3q/2},\quad i=0,1,\dots,n-1.$$

Therefore, for any $k=0,1,\dots,n-1$

$$\mathbb{E}|Z_k|^q\le Cn^{-q}. \tag{25}$$

Combining together (21), (22), (24) and (25), we have

$$\mathbb{E}|\tilde{M}^{RM}_{n,2}(t)|\le Cn^{-q\min\{\frac{1}{2}+\gamma_1,1\}}.$$

Therefore, for any $t\in[0,T]$

$$\mathbb{E}|\tilde{A}^{RM}_{n,1}(t)|^q\le Cn^{-q\min\{\frac{1}{2}+\gamma_1,1\}}. \tag{26}$$

We now bound from above $\sup_{t\in[0,T]}\mathbb{E}|\tilde{A}^{RM}_{n,2}(t)|^q$. The estimation goes analogously as in [17], with some minor adjustments needed in order to include the Hölder regularity. For reader's convenience we present a complete estimation procedure.

We denote by

$$i(t)=\sup\{i=0,1,\dots,n\mid iT/n\le t\},$$

$$\zeta(t)=i(t)\frac{T}{n},$$

for $t\in[0,T]$. Now we can write that

$$\mathbb{E}|\tilde{A}^{RM}_{n,2}(t)|^q\le 2^{q-1}\Big(\mathbb{E}|\tilde{A}^{RM}_{n,21}(t)|^q+\mathbb{E}|\tilde{A}^{RM}_{n,22}(t)|^q\Big), \tag{27}$$

with

$$\mathbb{E}|\tilde{A}^{RM}_{n,21}(t)|^q=\mathbb{E}\Big|\sum_{k=0}^{i(t)-1}\int_{t_k}^{t_{k+1}}\Big(a(s,X(t_k))-a(\xi_k,X(t_k))\Big)\,\mathrm{d}s\Big|^q, \tag{28}$$

$$\mathbb{E}|\tilde{A}^{RM}_{n,22}(t)|^q=\mathbb{E}\Big|\int_{\zeta(t)}^{t}\Big(a(s,X(\zeta(t)))-a(\xi_{i(t)},X(\zeta(t)))\Big)\,\mathrm{d}s\Big|^q, \tag{29}$$

for all $t\in[0,T]$, where we take $\mathbb{E}|\tilde{A}^{RM}_{n,22}(T)|^q=0$. Moreover, let

$$\tilde{Y}_k=\int_{t_k}^{t_{k+1}}\Big(a(s,X(t_k))-a(\xi_k,X(t_k))\Big)\,\mathrm{d}s,\ \ k=0,1,\dots,n-1, \tag{30}$$

and

$$\tilde{Z}_j=\sum_{k=0}^{j}\tilde{Y}_k,\ \ j=0,1,\dots,n-1,$$

where we set $\tilde{Z}_{-1}:=0$. Note that

$$|\tilde{Y}_k|\le K\,\Big(1+\sup_{0\le t\le T}|X(t)|\Big)\,(T/n)^{\gamma_1+1},$$

and conditioned on Σ_∞ the random variables $(\tilde{Y}_k)_{k=0}^{n-1}$ are zero mean, independent, and bounded by $K\Big(1+\sup_{0\le t\le T}|X(t)|\Big)(T/n)^{\gamma_1+1}$. Therefore, by applying Theorem 4 from [18] and Lemma 8 we have for all $t\in[0,T]$ that

$$\begin{aligned}\mathbb{E}|\tilde{A}^{RM}_{n,21}(t)|^q&=\mathbb{E}|\tilde{Z}_{i(t)-1}|^q\le\mathbb{E}\left[\mathbb{E}\left(\max_{0\le j\le n-1}|\tilde{Z}_j|^q\mid\Sigma_\infty\right)\right]\\&\le C_2(T/n)^{q(\gamma_1+1)}\cdot n^{q/2}\cdot\mathbb{E}\Big(1+\sup_{t\in[0,T]}|X(t)|\Big)^q\le C_3n^{-q(\gamma_1+\frac{1}{2})},\end{aligned} \tag{31}$$

where $C_2, C_3 > 0$ depend only on the parameters of the class $\mathcal{F}(\gamma_1, \gamma_2, q, K)$ and q. Moreover, due to the fact that Σ_∞ and $\sigma(\xi_0, \xi_1, \ldots, \xi_{n-1})$ are independent σ-fields, we get for all $t \in [0, T)$ that

$$\begin{aligned}\mathbb{E}|\tilde{A}^{RM}_{n,22}(t)|^q &\le (t-\zeta(t))^{q-1} \cdot \int_{\zeta(t)}^t \mathbb{E}|a(s, X(\zeta(t))) - a(\xi_{i(t)}, X(\zeta(t)))|^q \,\mathrm{d}s \\ &\le C_4(t-\zeta(t))^{q-1} \cdot \Big(1 + \sup_{t\in[0,T]} \mathbb{E}|X(t)|^q\Big) \cdot \mathbb{E}\int_{\zeta(t)}^t |s-\xi_{i(t)}|^{q\gamma_1}\,\mathrm{d}s,\end{aligned}$$

and $\mathbb{E}|\tilde{A}^{RM}_{n,22}(T)|^q = 0$. Note that for $t \in [0, T) = \bigcup_{i=0}^{n-1} [t_i, t_{i+1})$ we have that $\zeta(t) \le t < \zeta(t) + h$. In addition, for $t \in [0, T)$ we have that $\xi_{i(t)}$ is uniformly distributed on $[\zeta(t), \zeta(t) + h]$. Hence, $|s - \xi_{i(t)}| \le h$ for all $t \in [0, T)$ and $s \in [\zeta(t), t] \subset [\zeta(t), \zeta(t) + h)$, which gives for all $t \in [0, T)$

$$\mathbb{E}\int_{\zeta(t)}^t |s-\xi_{i(t)}|^{q\gamma_1}\,\mathrm{d}s \le (t-\zeta(t)) \cdot h^{q\gamma_1}.$$

Therefore,

$$\mathbb{E}|\tilde{A}^{RM}_{n,22}(t)|^q \le C_5 n^{-q(1+\gamma_1)}. \tag{32}$$

Using (27), (31) and (32) we obtain

$$\mathbb{E}|\tilde{A}^{RM}_{n,2}(t)|^q \le C_6 n^{-q(\gamma_1+\frac{1}{2})}, \tag{33}$$

for all $t \in [0, T]$. Combining (13), (26) and (33) we get

$$\mathbb{E}|A(t) - \tilde{A}^{RM}_n(t)|^q \le C_1 \int_0^t \sum_{i=0}^{n-1} \mathbb{E}|X(t_i) - \tilde{X}^{RM}_n(t_i)|^q \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s + C_2 n^{-q\min\{\frac{1}{2}+\gamma_1, 1\}}. \tag{34}$$

The analysis of the diffusion part is as follows. For all $t \in [0, T]$

$$\mathbb{E}|B(t) - \tilde{B}^{RM}_n(t)|^q \le C\sum_{k=1}^3 \mathbb{E}|\tilde{B}^{RM}_{n,k}(t)|^q,$$

where

$$\mathbb{E}|\tilde{B}^{RM}_{n,1}(t)|^q = \mathbb{E}\Big|\int_0^t \sum_{i=0}^{n-1} (b(s, X(s)) - b(t_i, X(s))) \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}W(s)\Big|^q, \tag{35}$$

$$\mathbb{E}|\tilde{B}^{RM}_{n,2}(t)|^q = \mathbb{E}\Big|\int_0^t \sum_{i=0}^{n-1} \Big(b(t_i, X(s)) - b(t_i, X(t_i)) - \int_{t_i}^s L_1 b(U_i)\mathrm{d}W(u)\Big) \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}W(s)\Big|^q, \tag{36}$$

$$\mathbb{E}|\tilde{B}^{RM}_{n,3}(t)|^q = \mathbb{E}\Big|\int_0^t \sum_{i=0}^{n-1} (b(t_i, X(t_i)) - b(U_i)) \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}W(s)\Big|^q. \tag{37}$$

By Burkholder inequality and Lemma 8 we have for every $t \in [0, T]$ that

$$\begin{aligned}\mathbb{E}|\tilde{B}^{RM}_{n,1}(t)|^q &\le C\int_0^T \mathbb{E}\sum_{i=0}^{n-1} |b(s, X(s)) - b(t_i, X(s))|^q \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s \\ &\le C\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} \mathbb{E}(1+|X(s)|)^q \cdot (s-t_i)^{q\gamma_2}\mathrm{d}s \le Cn^{-q\gamma_2},\end{aligned}$$

and

$$\mathbb{E}|\tilde{B}^{RM}_{n,3}(t)|^q \le C_2\int_0^t \sum_{i=0}^{n-1} \mathbb{E}|X(t_i) - \tilde{X}^{RM}_n(t_i)|^q \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s.$$

From (15) we get for $s \in [t_i, t_{i+1}]$ that

$$b(t_i, X(s)) - b(t_i, X(t_i)) - \int_{t_i}^s L_1 b(U_i)\mathrm{d}W(u) = \int_{t_i}^s \alpha(b, t_i, u)\mathrm{d}u + \int_{t_i}^s \Big(\beta(b, t_i, u) - L_1 b(U_i)\Big)\mathrm{d}W(u). \tag{38}$$

Hence, from (36), (38), and by Burkholder inequality we get

$$\mathbb{E}|\tilde{B}^{RM}_{n,2}(t)|^q \leq C_1\mathbb{E}\int_0^t\sum_{i=0}^{n-1}\Big|\int_{t_i}^s \alpha(b,t_i,u)\mathrm{d}u\Big|^q\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s$$
$$+C_2\mathbb{E}\int_0^t\sum_{i=0}^{n-1}\Big|\int_{t_i}^s (\beta(b,t_i,u)-L_1b(U_i))\,\mathrm{d}W(u)\Big|^q\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s, \tag{39}$$

where

$$\mathbb{E}\int_0^t\sum_{i=0}^{n-1}\Big|\int_{t_i}^s \alpha(b,t_i,u)\mathrm{d}u\Big|^q\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s \leq \sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}(s-t_i)^{q-1}\int_{t_i}^s\mathbb{E}|\alpha(b,t_i,u)|^q\mathrm{d}u\mathrm{d}s$$
$$\leq C\cdot\Big(1+\sup_{t\in[0,T]}\mathbb{E}|X(t)|^{2q}\Big)\cdot\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}(s-t_i)^q\mathrm{d}s\leq Cn^{-q}, \tag{40}$$

$$\mathbb{E}\int_0^t\sum_{i=0}^{n-1}\Big|\int_{t_i}^s (\beta(b,t_i,u)-L_1b(U_i))\,\mathrm{d}W(u)\Big|^q\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s$$
$$\leq C\cdot n^{-(q/2)+1}\cdot\int_0^t\sum_{i=0}^{n-1}\Big(\int_{t_i}^s\mathbb{E}|\beta(b,t_i,u)-L_1b(U_i)|^q\mathrm{d}u\Big)\cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s. \tag{41}$$

Note that

$$|\beta(b,t_i,u)-L_1b(U_i)|\leq K^2(1+|X(u)|)\cdot|u-t_i|^{\gamma_2}+K|X(u)-X(t_i)|+K|X(t_i)-\tilde{X}^{RM}_n(t_i)| \tag{42}$$

and, therefore, for any $s\in[t_i,t_{i+1}]$ we have

$$\int_{t_i}^s\mathbb{E}|\beta(b,t_i,u)-L_1b(U_i)|^q\mathrm{d}u\leq\tilde{C}_1n^{-q\gamma_2-1}+C_2n^{-(q/2)-1}+C_3n^{-1}\mathbb{E}|X(t_i)-\tilde{X}^{RM}_n(t_i)|^q. \tag{43}$$

From (39), (40), (41) and (43) we obtain that

$$\mathbb{E}|\tilde{B}^{RM}_{n,2}(t)|^q\leq C_1n^{-q}+C_2n^{-q(\frac{1}{2}+\gamma_2)}+C_3\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|X(t_i)-\tilde{X}^{RM}_n(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s.$$

Hence, for any $t\in[0,T]$ we have

$$\mathbb{E}|B(t)-\tilde{B}^{RM}_n(t)|^q\leq K_1n^{-q\gamma_2}+K_2\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|X(t_i)-\tilde{X}^{RM}_n(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s. \tag{44}$$

By (34) and (44) we get for all $t\in[0,T]$ that

$$\mathbb{E}|X(t)-\tilde{X}^{RM}_n(t)|^q\leq C_1n^{-q\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}+C_2\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|X(t_i)-\tilde{X}^{RM}_n(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s,$$

which implies for all $t\in[0,T]$ that

$$\sup_{0\leq s\leq t}\mathbb{E}|X(s)-\tilde{X}^{RM}_n(s)|^q\leq C_1n^{-q\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}+C_2\int_0^t\sup_{0\leq u\leq s}\mathbb{E}|X(u)-\tilde{X}^{RM}_n(u)|^q\mathrm{d}s.$$

Finally, by using Gronwall's inequality we arrive at (12), which ends the proof. ■

Remark 5. The idea of time-randomization applied in construction of the randomized Milstein algorithm $\mathcal{A}^{RM}_n$, is analogous to that used for Monte Carlo approximation of Lebesgue integrals of scalar functions (see, for example, [19]) and similar to [17], where the authors analogously defined randomized Euler scheme. We also refer to [12], where a two-stage randomized Milstein scheme was constructed and its error was investigated. In particular, if $\gamma_2=\min\{\frac{1}{2}+\gamma_1,1\}$ then

$$\sup_{t\in[0,T]}\|X(t)-\tilde{X}^{RM}_n(t)\|_q\leq Cn^{-\min\{\frac{1}{2}+\gamma_1,1\}},$$

which recovers the result from [12].

Remark 6. We compare errors of classical Euler method $\mathcal{A}_n^E$, randomized Euler algorithm $\mathcal{A}_n^{RE}$, classical Milstein scheme $\mathcal{A}_n^M$, and randomized Milstein algorithm $\mathcal{A}_n^{RM}$ in the class $\mathcal{F}(\gamma_1,\gamma_2,q,K)$. Namely, in the case of exact information about a and b, we have that

$$
\begin{aligned}
&e^{(q)}(\mathcal{A}_n^E,\mathcal{F}(\gamma_1,\gamma_2,q,K),W,V^i,0,0)=O(n^{-\min\{\gamma_1,\gamma_2,1/2\}}),\\
&e^{(q)}(\mathcal{A}_n^{RE},\mathcal{F}(\gamma_1,\gamma_2,q,K),W,V^i,0,0)=O(n^{-\min\{1/2,\gamma_2\}}),\\
&e^{(q)}(\mathcal{A}_n^M,\mathcal{F}(\gamma_1,\gamma_2,q,K),W,V^i,0,0)=O(n^{-\min\{\gamma_1,\gamma_2\}}),\\
&e^{(q)}(\mathcal{A}_n^{RM},\mathcal{F}(\gamma_1,\gamma_2,q,K),W,V^i,0,0)=O(n^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}).
\end{aligned}
\tag{45}
$$

Hence, if $\gamma\in(0,1/2]$ and $\gamma_2\in(0,1]$ then $\mathcal{A}_n^E$ and $\mathcal{A}_n^M$ have the same error $O(n^{-\min\{\gamma_1,\gamma_2\}})$. Moreover, for $\gamma_1\in(0,1]$ and $\gamma_2\in(0,1/2]$ the methods $\mathcal{A}_n^{RE}$ and $\mathcal{A}_n^{RM}$ have the same error $O(n^{-\gamma_2})$. Finally, for $\gamma_1\in(1/2,1)$ and $\gamma_2\in(1/2,1]$ the randomized Milstein algorithm $\mathcal{A}_n^{RM}$ outperforms $\mathcal{A}_n^E$, $\mathcal{A}_n^{RE}$, and $\mathcal{A}_n^M$.

3.2. Performance of randomized derivative-free Milstein algorithm for exact information

In this section we analyze the error of the algorithm $\mathcal{A}_n^{df-RM}$ in the case of exact information. Recall that its time-continuous version is denoted by $\tilde{X}_n^{df-RM}=\{\tilde{X}_n^{df-RM}(t)\}_{t\in[0,T]}$.

We now give proof of the following results.

Proposition 2. *There exists a positive constant C, depending only on the parameters of the class $\mathcal{F}(\gamma_1,\gamma_2,q,K)$, such that for all $n\in\mathbb{N}$ and all $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$ we have*

$$\sup_{t\in[0,T]}\|X(t)-\tilde{X}_n^{df-RM}(t)\|_q\le Cn^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}},\tag{46}$$

and, in particular,

$$\|X(T)-\mathcal{A}_n^{df-RM}(a,b,\eta,W,0,0)\|_q\le Cn^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}.$$

Proof. By (12) we have that

$$\sup_{t\in[0,T]}\|X(t)-\tilde{X}_n^{df-RM}(t)\|_q\le Cn^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}+\sup_{t\in[0,T]}\|\tilde{X}_n^{RM}(t)-\tilde{X}_n^{df-RM}(t)\|_q.\tag{47}$$

Hence, we only need to estimate

$$\sup_{t\in[0,T]}\|\tilde{X}_n^{RM}(t)-\tilde{X}_n^{df-RM}(t)\|_q.$$

Recall that

$$U_i=(t_i,\tilde{X}_n^{RM}(t_i)),\quad V_i=(\xi_i,\tilde{X}_n^{RM}(t_i)).$$

In addition, let us denote by

$$U_i^{df}=(t_i,\tilde{X}_n^{df-RM}(t_i)),\quad V_i^{df}=(\xi_i,\tilde{X}_n^{df-RM}(t_i)).$$

We have that for all $t\in[0,T]$

$$
\begin{aligned}
&\tilde{X}_n^{df-RM}(t)=\eta+\tilde{A}_n^{df-RM}(t)+\tilde{B}_n^{df-RM}(t),\\
&\tilde{A}_n^{df-RM}(t)=\int_0^t\sum_{i=0}^{n-1}a(V_i^{df})\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s,\\
&\tilde{B}_n^{df-RM}(t)=\int_0^t\sum_{i=0}^{n-1}\Big(b(U_i^{df})+\int_{t_i}^s\mathcal{L}_{1,h}b(U_i^{df})\mathrm{d}W(u)\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}W(s).
\end{aligned}
$$

Then

$$\mathbb{E}|\tilde{A}_n^{RM}(t)-\tilde{A}_n^{df-RM}(t)|^q\le C_1\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|\tilde{X}_n^{RM}(t_i)-\tilde{X}_n^{df-RM}(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s,\tag{48}$$

and

$$
\begin{aligned}
\mathbb{E}|\tilde{B}_n^{RM}(t)-\tilde{B}_n^{df-RM}(t)|^q\le{}&C\mathbb{E}\Big|\int_0^t\sum_{i=0}^{n-1}(b(U_i)-b(U_i^{df}))\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}W(s)\Big|^q\\
&+C\mathbb{E}\Big|\int_0^t\sum_{i=0}^{n-1}\Big(\int_{t_i}^s\Big(L_1b(U_i)-\mathcal{L}_{1,h}b(U_i^{df})\Big)\,\mathrm{d}W(u)\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}W(s)\Big|^q.
\end{aligned}
$$

Furthermore, by Burkholder inequality and Lemma 5

$$\mathbb{E}\Big|\int_0^t\sum_{i=0}^{n-1}(b(U_i)-b(U_i^{df}))\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}W(s)\Big|^q\leq C\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|\tilde{X}_n^{RM}(t_i)-\tilde{X}_n^{df-RM}(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s,$$

$$\mathbb{E}\Big|\int_0^t\sum_{i=0}^{n-1}\Big(\int_{t_i}^s\Big(L_1b(U_i)-\mathcal{L}_{1,h}b(U_i^{df})\Big)\,\mathrm{d}W(u)\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}W(s)\Big|^q$$
$$\leq C\int_0^t\sum_{i=0}^{n-1}\Big((s-t_i)^{q/2-1}\cdot\int_{t_i}^s\mathbb{E}|L_1b(U_i)-\mathcal{L}_{1,h}b(U_i^{df})|^q\,\mathrm{d}u\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s$$
$$\leq C_1\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|\tilde{X}_n^{RM}(t_i)-\tilde{X}_n^{df-RM}(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s+C_2n^{-3q/2}\Big(1+\sup_{t\in[0,T]}\mathbb{E}|\tilde{X}_n^{df-RM}(t)|^q\Big).$$

Therefore,

$$\mathbb{E}|\tilde{B}_n^{RM}(t)-\tilde{B}_n^{df-RM}(t)|^q\leq C_1\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|\tilde{X}_n^{RM}(t_i)-\tilde{X}_n^{df-RM}(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s$$
$$+\,C_2\Big(1+\sup_{t\in[0,T]}\mathbb{E}|\tilde{X}_n^{df-RM}(t)|^q\Big)n^{-3q/2}. \tag{49}$$

Hence, from (48) and (49) we get for all $t\in[0,T]$

$$\mathbb{E}|\tilde{X}_n^{RM}(t)-\tilde{X}_n^{df-RM}(t)|^q\leq C_1\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|\tilde{X}_n^{RM}(t_i)-\tilde{X}_n^{df-RM}(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s$$
$$+\,C_2\Big(1+\sup_{t\in[0,T]}\mathbb{E}|\tilde{X}_n^{df-RM}(t)|^q\Big)n^{-3q/2}.$$

Hence, by Gronwall's lemma we obtain

$$\mathbb{E}|\tilde{X}_n^{RM}(t)-\tilde{X}_n^{df-RM}(t)|^q\leq C\Big(1+\sup_{t\in[0,T]}\mathbb{E}|\tilde{X}_n^{df-RM}(t)|^q\Big)n^{-3q/2}. \tag{50}$$

Therefore, by (47), (50) and Lemma 7

$$\sup_{0\leq t\leq T}\|X(t)-\tilde{X}_n^{df-RM}(t)\|_q\leq C_1\Big(1+\sup_{t\in[0,T]}\|\tilde{X}_n^{df-RM}(t)\|_q\Big)n^{-3/2}+C_2n^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}\leq Cn^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}. \tag{51}$$

This ends the proof. ■

Having Proposition 2 we are ready to prove Theorem 1.

3.3. Proof of Theorem 1

We set

$$\bar{U}_i^{df}=(t_i,\tilde{\tilde{X}}_n^{df-RM}(t_i)),\qquad \bar{V}_i^{df}=(\xi_i,\tilde{\tilde{X}}_n^{df-RM}(t_i)).$$

The process $\{\tilde{\tilde{X}}_n^{df-RM}(t)\}_{t\in[0,T]}$ can be decomposed as follows

$$\tilde{\tilde{X}}_n^{df-RM}(t)=\eta+\tilde{\tilde{A}}_n^{df-RM}(t)+\tilde{\tilde{B}}_n^{df-RM}(t),$$

where

$$\tilde{\tilde{A}}_n^{df-RM}(t)=\int_0^t\sum_{i=0}^{n-1}\tilde{a}(\bar{V}_i^{df})\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s,$$
$$\tilde{\tilde{B}}_n^{df-RM}(t)=\int_0^t\sum_{i=0}^{n-1}\Big(\tilde{b}(\bar{U}_i^{df})+\int_{t_i}^s\mathcal{L}_{1,h}\tilde{b}(\bar{U}_i^{df})\mathrm{d}W(u)\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}W(s). \tag{52}$$

Due to Lemma 1 the process

$$\Big\{\sum_{i=0}^{n-1}\Big(\tilde{b}(\bar{U}_i^{df})+\int_{t_i}^{s}\mathcal{L}_{1,h}\tilde{b}(\bar{U}_i^{df})\mathrm{d}W(u)\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\Big\}_{s\in[0,T]}$$

is adapted to $\{\tilde{\Sigma}_t^n\}_{t\in[0,T]}$ and has cádlág paths. Hence, the Itô integral in (52) is well-defined.

From (46) we have that

$$\sup_{t\in[0,T]}\|X(t)-\tilde{\tilde{X}}_n^{df-RM}(t)\|_q\le Cn^{-\min\{\frac{1}{2}+\gamma_1,\gamma_2\}}+\sup_{t\in[0,T]}\|\tilde{X}_n^{df-RM}(t)-\tilde{\tilde{X}}_n^{df-RM}(t)\|_q,\tag{53}$$

and we only need to estimate $\sup_{t\in[0,T]}\|\tilde{X}_n^{df-RM}(t)-\tilde{\tilde{X}}_n^{df-RM}(t)\|_q$. We have that

$$\begin{aligned}\mathbb{E}|\tilde{A}_n^{df-RM}(t)-\tilde{\tilde{A}}_n^{df-RM}(t)|^q&\le C\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|\tilde{X}_n^{df-RM}(t_i)-\tilde{\tilde{X}}_n^{df-RM}(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\mathrm{d}s\\&+C\Big(1+\sup_{0\le t\le T}\mathbb{E}|\tilde{\tilde{X}}_n^{df-RM}(t)|^q\Big)\cdot\delta_1^q,\end{aligned}\tag{54}$$

and, by the Burkholder inequality,

$$\begin{aligned}\mathbb{E}|\tilde{B}_n^{df-RM}(t)-\tilde{\tilde{B}}_n^{df-RM}(t)|^q&\le C\mathbb{E}\Big|\int_0^t\sum_{i=0}^{n-1}\Big(b(U_i^{df})-\tilde{b}(\bar{U}_i^{df})\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}W(s)\Big|^q\\&+C\mathbb{E}\Big|\int_0^t\sum_{i=0}^{n-1}\Big(\int_{t_i}^{s}\Big(\mathcal{L}_{1,h}b(U_i^{df})-\mathcal{L}_{1,h}\tilde{b}(\bar{U}_i^{df})\Big)\,\mathrm{d}W(u)\Big)\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}W(s)\Big|^q\\&\le C\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|b(U_i^{df})-\tilde{b}(\bar{U}_i^{df})|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s\\&+C\int_0^t\sum_{i=0}^{n-1}(s-t_i)^{q/2}\cdot\mathbb{E}|\mathcal{L}_{1,h}b(U_i^{df})-\mathcal{L}_{1,h}\tilde{b}(\bar{U}_i^{df})|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s.\end{aligned}\tag{55}$$

Note that

$$\mathbb{E}|b(U_i^{df})-\tilde{b}(\bar{U}_i^{df})|^q\le C\mathbb{E}|\tilde{X}_n^{df-RM}(t_i)-\tilde{\tilde{X}}_n^{df-RM}(t_i)|^q+C\delta_2^q\Big(1+\sup_{0\le t\le T}\mathbb{E}|\tilde{\tilde{X}}_n^{df-RM}(t)|^q\Big)$$

and, by Lemma 6,

$$\begin{aligned}\mathbb{E}|\mathcal{L}_{1,h}b(U_i^{df})-\mathcal{L}_{1,h}\tilde{b}(\bar{U}_i^{df})|^q&\le C\Big(1+\sup_{0\le t\le T}\mathbb{E}|\tilde{X}_n^{df-RM}(t)|^q+\sup_{0\le t\le T}\mathbb{E}|\tilde{\tilde{X}}_n^{df-RM}(t)|^q\Big)\cdot h^q\\&+K\mathbb{E}|\tilde{X}_n^{df-RM}(t_i)-\tilde{\tilde{X}}_n^{df-RM}(t_i)|^q\\&+C\Big(1+\sup_{0\le t\le T}\mathbb{E}|\tilde{\tilde{X}}_n^{df-RM}(t)|^q\Big)\cdot(1+\delta_2^q)\cdot\begin{cases}\delta_2^q, & \text{if } p_b\in\mathcal{K}_{\mathrm{Lip}}^1\\(\delta_2h^{-1})^q, & \text{if } p_b\in\mathcal{K}^2.\end{cases}\end{aligned}\tag{56}$$

Therefore, we get that for all $t\in[0,T]$

$$\begin{aligned}\mathbb{E}|\tilde{B}_n^{df-RM}(t)-\tilde{\tilde{B}}_n^{df-RM}(t)|^q&\le C\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|\tilde{X}_n^{df-RM}(t_i)-\tilde{\tilde{X}}_n^{df-RM}(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s\\&+C\Big(1+\sup_{0\le t\le T}\mathbb{E}|\tilde{\tilde{X}}_n^{df-RM}(t)|^q\Big)\cdot\delta_2^q\\&+C\Big(1+\sup_{0\le t\le T}\mathbb{E}|\tilde{X}_n^{df-RM}(t)|^q+\sup_{0\le t\le T}\mathbb{E}|\tilde{\tilde{X}}_n^{df-RM}(t)|^q\Big)\cdot h^{3q/2}\\&+C\Big(1+\sup_{0\le t\le T}\mathbb{E}|\tilde{\tilde{X}}_n^{df-RM}(t)|^q\Big)\cdot(1+\delta_2^q)\cdot\begin{cases}(h^{1/2}\delta_2)^q, & \text{if } p_b\in\mathcal{K}_{\mathrm{Lip}}^1\\(h^{-1/2}\delta_2)^q, & \text{if } p_b\in\mathcal{K}^2.\end{cases}\end{aligned}\tag{57}$$

From (54) and (57) we get

$$\begin{aligned}\mathbb{E}|\tilde{X}_n^{df-RM}(t)-\tilde{\bar{X}}_n^{df-RM}(t)|^q &\leq C\int_0^t\sum_{i=0}^{n-1}\mathbb{E}|\tilde{X}_n^{df-RM}(t_i)-\tilde{\bar{X}}_n^{df-RM}(t_i)|^q\cdot\mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s\\ &+C\Big(1+\sup_{0\leq t\leq T}\mathbb{E}|\tilde{\bar{X}}_n^{df-RM}(t)|^q\Big)\cdot(\delta_1^q+\delta_2^q)\\ +C\Big(1+\sup_{0\leq t\leq T}\mathbb{E}|\tilde{X}_n^{df-RM}(t)|^q&+\sup_{0\leq t\leq T}\mathbb{E}|\tilde{\bar{X}}_n^{df-RM}(t)|^q\Big)\cdot h^{3q/2}\\ +C\Big(1+\sup_{0\leq t\leq T}\mathbb{E}|\tilde{\bar{X}}_n^{df-RM}(t)|^q\Big)&\cdot(1+\delta_2^q)\cdot\begin{cases}(h^{1/2}\delta_2)^q, & \text{if } p_b\in\mathcal{K}^1_{\text{Lip}}\\ (h^{-1/2}\delta_2)^q, & \text{if } p_b\in\mathcal{K}^2.\end{cases}\end{aligned}\tag{58}$$

Thereby, the Gronwall's lemma implies

$$\begin{aligned}&\sup_{0\leq t\leq T}\|\tilde{X}_n^{df-RM}(t)-\tilde{\bar{X}}_n^{df-RM}(t)\|_q\leq C_1\Big(1+\sup_{0\leq t\leq T}\|\tilde{\bar{X}}_n^{df-RM}(t)\|_q\Big)\cdot(\delta_1+\delta_2)\\ &+C_2\Big(1+\sup_{0\leq t\leq T}\|\tilde{X}_n^{df-RM}(t)\|_q+\sup_{0\leq t\leq T}\|\tilde{\bar{X}}_n^{df-RM}(t)\|_q\Big)\cdot h^{3/2}\\ &+C_3\Big(1+\sup_{0\leq t\leq T}\|\tilde{\bar{X}}_n^{df-RM}(t)\|_q\Big)\cdot(1+\delta_2)\cdot\begin{cases}h^{1/2}\delta_2, & \text{if } p_b\in\mathcal{K}^1_{\text{Lip}}\\ h^{-1/2}\delta_2, & \text{if } p_b\in\mathcal{K}^2.\end{cases}\end{aligned}\tag{59}$$

Combining (53), (59), and Lemma 7 we get the hypothesis. ■

4. Lower bounds and optimality of randomized derivative-free Milstein algorithm

This section is dedicated to establishing the lower bounds on the worst-case error of an arbitrary algorithm from Φ_n and to prove that the randomized derivative-free Milstein algorithm $\bar{X}_n^{df-RM}$ is asymptotically optimal.

Lemma 3. *Let $q\in[2,+\infty)$, $\gamma_1,\gamma_2\in(0,1]$, $K\in(0,+\infty)$, then*

$$e_n^{(q)}(\mathcal{F}(\gamma_1,\gamma_2,q,K),W,V^i,\delta_1,\delta_2)=\Omega(\max\{n^{-\min\{1/2+\gamma_1,\gamma_2\}},\delta_1,\delta_2\}),\tag{60}$$

for $i=1,2$ as $n\to+\infty$, $\max\{\delta_1,\delta_2\}\to 0+$.

Proof. The proof is similar to that presented in [10], however, for reader's convenience we provide the details.

Let us consider the following subclasses of $\mathcal{F}(\gamma_1,\gamma_2,q,K)$:

$$\mathcal{G}_1(\gamma_1,\gamma_2,q,K)=\bar{\mathcal{A}}_K^{\gamma_1}\times\{0\}\times\{0\},\tag{61}$$

$$\mathcal{G}_2(\gamma_1,\gamma_2,q,K)=\{0\}\times\bar{\mathcal{B}}_K^{\gamma_2}\times\{0\},\tag{62}$$

where

$$\begin{aligned}\bar{\mathcal{A}}_K^{\gamma_1}&=\{a\in\mathcal{A}_K^{\gamma_1}\mid a(t,y)=a(t,0)\text{ for all } t\in[0,T], y\in\mathbb{R}\},\\ \bar{\mathcal{B}}_K^{\gamma_2}&=\{b\in\mathcal{B}_K^{\gamma_2}\mid b(t,y)=b(t,0)\text{ for all } t\in[0,T], y\in\mathbb{R}\}.\end{aligned}\tag{63}$$

In the class $\mathcal{G}_1(\gamma_1,\gamma_2,q,K)$ the approximation of $X(T)$ is equivalent to the problem of approximating the Lebesgue integral $X(T)=\int_0^T a(t,0)dt$, while in $\mathcal{G}_2(\gamma_1,\gamma_2,q,K)$ the problem reduces to approximation of scalar stochastic Itô integral $X(T)=\int_0^T b(t,0)dW(t)$.

Since $\{a\}\times\{b\}=V_a(0)\times V_b^i(0)\subset V_a(\delta_1)\times V_b^i(\delta_2)$ for any $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$ and $i=1,2$, we get, by considering the subclasses (61), (62) of $\mathcal{F}(\gamma_1,\gamma_2,q,K)$, that

$$\begin{aligned}&e_n^{(q)}(\mathcal{F}(\gamma_1,\gamma_2,q,K),W,V^i,\delta_1,\delta_2)\geq e_n^{(q)}(\mathcal{F}(\gamma_1,\gamma_2,q,K),W,V^i,0,0)\\ &\geq\max\{e_n^{(q)}(\mathcal{G}_1(\gamma_1,\gamma_2,q,K),W,V^i,0,0),e_n^{(q)}(\mathcal{G}_2(\gamma_1,\gamma_2,q,K),W,V^i,0,0)\}.\end{aligned}\tag{64}$$

We now recall known results on the lower bounds in the case of exact information, i.e. $\delta_1=\delta_2=0$. For Lebesgue integration of Hölder continuous functions under randomized standard information the following lower bound follows from [19]

$$e_n^{(q)}(\mathcal{G}_1(\gamma_1,\gamma_2,q,K),W,V^i,0,0)=\Omega(n^{-(1/2+\gamma_1)}),\tag{65}$$

for $i=1,2$. Furthermore, in [17] and [20] the following lower bound was established for Itô integration

$$e_n^{(q)}(\mathcal{G}_2(\gamma_1,\gamma_2,q,K),W,V^i,0,0)=\Omega(n^{-\gamma_2}),\tag{66}$$

for $i = 1, 2$. (The lower bound (66) holds also in case when the evaluation points for W are chosen in an adaptive way, see [20] for details.) By (64), (65), and (66) we arrive at

$$e_n^{(q)}(\mathcal{F}(\gamma_1, \gamma_2, q, K), W, V^i, \delta_1, \delta_2) = \Omega(n^{-\min\{\frac{1}{2}+\gamma_1, \gamma_2\}}). \tag{67}$$

Let us assume that $\delta_1, \delta_2 \in [0, \min\{K, 1\}]$. Since $\mathcal{G}_1(\gamma_1, \gamma_2, q, K)$, $\mathcal{G}_2(\gamma_1, \gamma_2, q, K) \subset \mathcal{F}(\gamma_1, \gamma_2, q, K)$, $V_a(0) \times V_b^i(\delta_2) \subset V_a(\delta_1) \times V_b^i(\delta_2)$, and $V_a(\delta_1) \times V_b^i(0) \subset V_a(\delta_1) \times V_b^i(\delta_2)$, we have that

$$\begin{aligned} &e_n^{(q)}(\mathcal{F}(\gamma_1, \gamma_2, q, K), W, V^i, \delta_1, \delta_2) \\ &\geq \max\{e_n^{(q)}(\mathcal{G}_1(\gamma_1, \gamma_2, q, K), W, V^i, \delta_1, 0), e_n^{(q)}(\mathcal{G}_2(\gamma_1, \gamma_2, q, K), W, V^i, 0, \delta_2)\}, \end{aligned} \tag{68}$$

and we need to establish the lower bounds for $e_n^{(q)}(\mathcal{G}_1(\gamma_1, \gamma_2, q, K), W, V^i, \delta_1, 0)$, $e_n^{(q)}(\mathcal{G}_2(\gamma_1, \gamma_2, q, K), W, V^i, 0, \delta_2)$. To do this we need the following auxiliary inequality, that is a direct consequence of the triangle inequality and the definition of the worst-case error (3). Let $\mathcal{G}$ be a subclass of $\mathcal{F}(\gamma_1, \gamma_2, q, K)$. Then for any algorithm $\mathcal{A} \in \Phi_n$ and any (a_1, b_1, η), $(a_2, b_2, \eta) \in \mathcal{G}$, such that

$$\left(V_{a_1}(\delta_1) \times V_{b_1}^i(\delta_2)\right) \cap \left(V_{a_2}(\delta_1) \times V_{b_2}^i(\delta_2)\right) \neq \emptyset, \tag{69}$$

it holds

$$e^{(q)}(\mathcal{A}, \mathcal{G}, W, V^i, \delta_1, \delta_2) \geq \frac{1}{2}\|X(a_1, b_1, \eta)(T) - X(a_2, b_2, \eta)(T)\|_q, \tag{70}$$

where $i = 1, 2$. Since $(\delta_1, 0, 0)$, $(-\delta_1, 0, 0) \in \mathcal{G}_1(\gamma_1, \gamma_2, q, K)$ and $(0, 0) \in \left(V_{\delta_1}(\delta_1) \times \{0\}\right) \cap \left(V_{-\delta_1}(\delta_1) \times \{0\}\right)$, we get by (70) that

$$e^{(q)}(\mathcal{A}, \mathcal{G}_1(\gamma_1, \gamma_2, q, K), W, V^i, \delta_1, 0) \geq \frac{1}{2}\|X(\delta_1, 0, 0)(T) - X(-\delta_1, 0, 0)(T)\|_q = T\delta_1. \tag{71}$$

Moreover, $(0, \delta_2, 0)$, $(0, -\delta_2, 0) \in \mathcal{G}_2(\gamma_1, \gamma_2, q, K)$ and $(0, 0) \in \left(\{0\} \times V_{\delta_2}^i(\delta_2)\right) \cap \left(\{0\} \times V_{-\delta_2}^i(\delta_2)\right)$. Therefore, by (70)

$$e^{(q)}(\mathcal{A}, \mathcal{G}_2(\gamma_1, \gamma_2, q, K), W, V^i, 0, \delta_2) \geq \frac{1}{2}\|X(0, \delta_2, 0)(T) - X(0, -\delta_2, 0)(T)\|_q = m_q T^{1/2}\delta_2, \tag{72}$$

where $m_q = \|Z\|_q$ and Z is normally distributed random variable with mean zero and variance equal to 1. Hence, (68), (71), and (72) imply

$$e_n^{(q)}(\mathcal{F}(\gamma_1, \gamma_2, q, K), W, V^i, \delta_1, \delta_2) = \Omega(\max\{\delta_1, \delta_2\}) \tag{73}$$

for $i = 1, 2$. Finally, from (67) and (73) we get (60). $\square$

The following theorem is the main result of the paper and establishes optimality of randomized derivative-free Milstein algorithm.

Theorem 2. *Let $q \in [2, +\infty)$, $\gamma_1, \gamma_2 \in (0, 1]$, $K \in (0, +\infty)$, then*

$$e_n^{(q)}(\mathcal{F}(\gamma_1, \gamma_2, q, K), W, V^1, \delta_1, \delta_2) = \Theta(\max\{n^{-\min\{1/2+\gamma_1, \gamma_2\}}, \delta_1, \delta_2\}),$$

as $n \to +\infty$, $\max\{\delta_1, \delta_2\} \to 0+$. An algorithm of optimal order is the randomized derivative-free Milstein algorithm $\bar{X}_n^{df-RM}$.

Sharp bounds for the class V^2 in the case when $\delta_2 > 0$ remain as an open problem.

5. Numerical experiments

We present numerical results for randomized derivative-free Milstein algorithm $\bar{X}_n^{df-RM}$ for the following problem

$$\begin{cases} dX(t) = \sin(M \cdot X(t) \cdot t^{\gamma_1})\,\mathrm{d}t + \cos(M \cdot X(t) \cdot t^{\gamma_2})\,\mathrm{d}W(t), & t \in [0, T], \\ X(0) = 1.0, \end{cases} \tag{74}$$

where $M = 100$, $\gamma_1 \in (0, 1]$, $\gamma_2 = \min\{\gamma_1 + 0.5, 1\}$. Drift and diffusion coefficients are Hölder continuous functions with Hölder exponents γ_1 and γ_2, respectively (see Remark 1). The expected theoretical convergence rate for this problem, according to Theorem 1, is $n^{-\gamma_2}$ as n tends to $+\infty$, and δ_1, δ_2 tend to zero.

Note that the exact solution of (74) is not known. Hence, in the simulations we computed in parallel the approximation of the solution for mesh of cardinality n and $1000n$, treating the one on dense mesh as the exact. The rule of thumb for such a choice is as follows. The projected convergence rate is at least $n^{-0.5}$, so the error for $1000n$ should be at least an order of magnitude lower than the error on n points, hence,

$$\|\bar{X}_{1000n}^{df-RM}(T) - \bar{X}_n^{df-RM}(T)\|_2 \approx \|X(T) - \bar{X}_n^{df-RM}(T)\|_2.$$

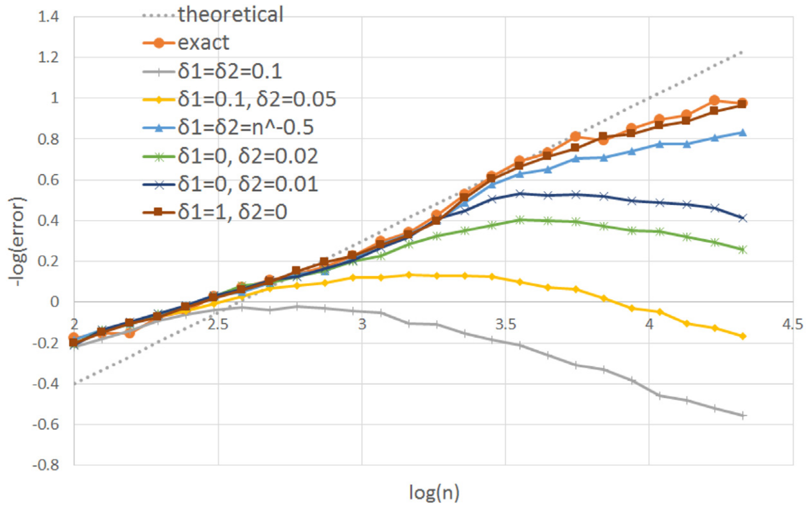

Fig. 1. Error for exact/noisy information for the case $\gamma_1 = 0.2$, $\gamma_2 = 0.7$.

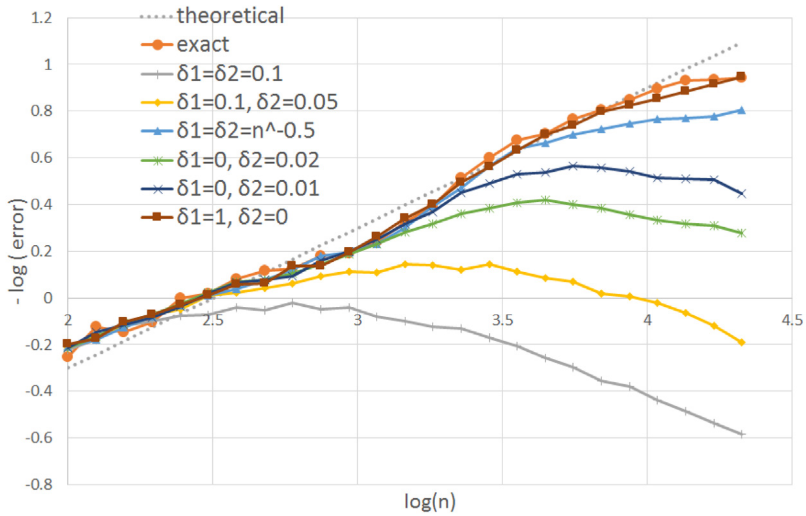

Fig. 2. Error for exact/noisy information for the case $\gamma_1 = 0.1$, $\gamma_2 = 0.6$.

The expectation is estimated as an average taken over $K = 10^4$ trajectories of the driving Wiener process. The informational noise the coefficients a and b is simulated as follows. We assume that the corrupting functions $p(t, y)$ for drift and diffusion coefficients are bounded, i.e. $|p_a(t, y)| \le \delta_1$ and $|p_b(t, y)| \le \delta_2$. The noising procedure was simulated as a realization of a random variable uniformly distributed on [0, 1], scaled by the respective precision level δ_1 or δ_2. Each corruption was generated independently. The obtained results are presented in Figs. 1 and 2. The plots present minus logarithm of the approximation error based on the logarithm of number of discretization points, hence, the theoretical error should form a line with the slope corresponding to the theoretical rate of convergence. For the obtained numerical results, the empirical convergence rate was also computed (through the linear regression of the $\log n$ vs $-\log$ error curve, where log denotes the logarithm with base 10), the summary of those can be find in Table 1.

The obtained numerical results confirm the theoretical results. The most surprising might be the fact that for a set precision on diffusion coefficient and with increasing number of discretization points, the error grows. That indicates that it is likely that the theoretical upper bound for error estimate for analyzed method is sharp with respect to the factor of $\delta_2 n^{1/2}$ in Theorem 1. This behavior is not observed for the set precision δ_1 on the drift coefficient and increasing number of discretization points. The results also prove that with precision levels tending to zero with the theoretical convergence rate of the method, the observed convergence rate behaves similarly as by the exact information.

Moreover, the obtained results clearly indicate that this method cannot be optimal, as we can simply omit part of the information used, not letting the noise (coming from the corrupted diffusion coefficient) to increase the overall error of the method. (We can see from Figs. 1 and 2 that for a given precision level δ_2 and for a larger number of evaluations used

Table 1
Empirical convergence rates for various precision levels.

	$\gamma_1 = 0.1,\ \gamma_2 = 0.6$	$\gamma_1 = 0.2,\ \gamma_2 = 0.7$
Theoretical	0.6	0.7
Exact	0.54	0.55
$\delta_1 = \delta_2 = 0.1$	−0.20	−0.19
$\delta_1 = 0.1, \delta_2 = 0.05$	0.01	0.00
$\delta_1 = \delta_2 = n^{-0.5}$	0.48	0.48
$\delta_1 = 0, \delta_2 = 0.02$	0.25	0.23
$\delta_1 = 0, \delta_2 = 0.01$	0.33	0.31
$\delta_1 = 1, \delta_2 = 0$	0.52	0.54

the error is higher than for the algorithm using fewer number of evaluations.) We believe that it is possible to propose an adaptive procedure that chooses an optimal number of discretization points according to a given precision level. However, we leave it as an open problem. Furthermore, in this paper we considered only noisy information about drift and diffusion coefficients. In case when also the evaluations of the Wiener process are corrupted direct application of the technique used in this paper is not possible. Hence, further extension of research on the subject is needed both for the lower and the upper bounds on the error.

Acknowledgments

This research was partly supported by the National Science Centre, Poland, under project 2017/25/B/ST1/00945.

We would like to thank two anonymous referees for their valuable comments and suggestions that helped to improve the presentation of the results and quality of this paper.

Appendix

The proofs of the following two lemmas are straightforward and will be omitted.

Lemma 4. *If $f \in F_K^{\gamma}$, $\gamma \in \{\gamma_1, \gamma_2\}$, then for all $(t, y) \in [0, T] \times \mathbb{R}$*

$$|f(t, y)| \leq K_1(1 + |y|),$$

$$\left|\frac{\partial^j f}{\partial y^j}(t, y)\right| \leq K, \quad j = 1, 2,$$

where $K_1 = K(1 + \max\{T^{\gamma_1}, T^{\gamma_2}\})$.

Lemma 5. *For all $n \in \mathbb{N}$, $b \in \mathcal{B}_K^{\gamma_2}$, and all $t \in [0, T]$, $y, z \in \mathbb{R}$ it holds*

$$|L_1 b(t, y)| \leq KK_1(1 + |y|),$$

$$|\mathcal{L}_{1,h} b(t, y)| \leq KK_1(1 + |y|),$$

$$|L_1 b(t, y) - \mathcal{L}_{1,h} b(t, z)| \leq K|y - z| + KK_1(1 + |z|)h,$$

where $h = T/n$ and $K_1 = K(1 + \max\{T^{\gamma_1}, T^{\gamma_2}\})$.

In the following lemma we investigate behavior of difference operator $\mathcal{L}_{1,h}$ in the case of inexact information about b.

Lemma 6. *There exists a positive constant C, such that for all $n \in \mathbb{N}$, $\delta_1, \delta_2 \in [0, 1]$, $(a, b) \in \mathcal{A}_K^{\gamma_1} \times \mathcal{B}_K^{\gamma_2}$, $(\tilde{a}, \tilde{b}) \in V_a(\delta_1) \times (V_b^1(\delta_2) \cup V_b^2(\delta_2))$, and all $t \in [0, T]$, $y, z \in \mathbb{R}$ it holds*

$$|\tilde{a}(t, y)| \leq C(1 + \delta_1)(1 + |y|), \tag{75}$$

$$|\tilde{b}(t, y)| \leq C(1 + \delta_2)(1 + |y|), \tag{76}$$

$$|\mathcal{L}_{1,h}\tilde{b}(t, y)| \leq C(1 + \delta_2)(1 + |y|) \cdot \begin{cases} 1 + \delta_2, & \text{if } p_b \in \mathcal{K}^1_{\mathrm{Lip}}, \\ 1 + \delta_2 h^{-1}, & \text{if } p_b \in \mathcal{K}^2, \end{cases} \tag{77}$$

$$\begin{aligned} |\mathcal{L}_{1,h} b(t, y) - \mathcal{L}_{1,h}\tilde{b}(t, z)| &\leq C(1 + |y| + |z|)h + K|y - z| \\ &\quad + C(1 + |z|) \cdot (1 + \delta_2) \cdot \begin{cases} \delta_2, & \text{if } p_b \in \mathcal{K}^1_{\mathrm{Lip}} \\ \delta_2 h^{-1}, & \text{if } p_b \in \mathcal{K}^2. \end{cases} \end{aligned} \tag{78}$$

Proof. The proof of (75) and (76) is straightforward.

We have that

$$|\mathcal{L}_{1,h}b(t,y)-\mathcal{L}_{1,h}\tilde{b}(t,z)|\leq|\mathcal{L}_{1,h}b(t,y)-\mathcal{L}_{1,h}b(t,z)|+|\mathcal{L}_{1,h}b(t,z)-\mathcal{L}_{1,h}\tilde{b}(t,z)|. \tag{79}$$

From Lemma 5 we get that

$$\begin{aligned}|\mathcal{L}_{1,h}b(t,y)-\mathcal{L}_{1,h}b(t,z)| &\leq |\mathcal{L}_{1,h}b(t,y)-L_1b(t,z)|+|L_1b(t,z)-\mathcal{L}_{1,h}b(t,z)|\\ &\leq K|y-z|+K_1K(1+|y|)h+K_1K(1+|z|)h\\ &\leq C(1+|y|+|z|)h+K|y-z|.\end{aligned} \tag{80}$$

Furthermore,

$$\begin{aligned}|\mathcal{L}_{1,h}b(t,z)-\mathcal{L}_{1,h}\tilde{b}(t,z)| &\leq |b(t,z)|\cdot|\Delta_h b(t,z)-\Delta_h\tilde{b}(t,z)|+\delta_2\cdot|p_b(t,z)|\cdot|\Delta_h\tilde{b}(t,z)|\\ &\leq K_1(1+|z|)\cdot|\Delta_h b(t,z)-\Delta_h\tilde{b}(t,z)|+\delta_2\cdot(1+|z|)\cdot|\Delta_h\tilde{b}(t,z)|.\end{aligned} \tag{81}$$

Note that

$$|\Delta_h\tilde{b}(t,z)|\leq|\Delta_h b(t,z)|+\delta_2\cdot|\Delta_h p_b(t,z)|\leq K+\delta_2\cdot|\Delta_h p_b(t,z)|. \tag{82}$$

Moreover,

$$|\Delta_h b(t,z)-\Delta_h\tilde{b}(t,z)|=\delta_2\cdot|\Delta_h p_b(t,z)|,$$

and

$$|\Delta_h p_b(t,z)|\leq\begin{cases}1, & \text{if } p_b\in\mathcal{K}^1_{\mathrm{Lip}}\\ 2h^{-1}, & \text{if } p_b\in\mathcal{K}^2.\end{cases} \tag{83}$$

Hence,

$$\begin{aligned}|\mathcal{L}_{1,h}b(t,z)-\mathcal{L}_{1,h}\tilde{b}(t,z)| &\leq (K_1+\delta_2)\cdot\delta_2\cdot(1+|z|)\cdot|\Delta_h p_b(t,z)|+K\cdot\delta_2\cdot(1+|z|)\\ &\leq C(1+|z|)\cdot(1+\delta_2)\cdot\begin{cases}\delta_2, & \text{if } p_b\in\mathcal{K}^1_{\mathrm{Lip}}\\ \delta_2\cdot h^{-1}, & \text{if } p_b\in\mathcal{K}^2.\end{cases}\end{aligned} \tag{84}$$

Combining (79), (80), and (84) we get (78). Finally, by (76), (82), (83), and

$$|\mathcal{L}_{1,h}\tilde{b}(t,y)|\leq C(1+\delta_2)\cdot(1+|y|)\cdot(K+|\Delta_h p_b(t,y)|) \tag{85}$$

the result (77) follows. ■

Lemma 7.

(i) *There exists a positive constant C, depending only on the parameters of the class $\mathcal{F}(\gamma_1,\gamma_2,q,K)$, such that for all $n\in\mathbb{N}$, $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$, we have*

$$\sup_{t\in[0,T]}\mathbb{E}|\tilde{X}^{RM}_n(t)|^q\leq C, \tag{86}$$

$$\sup_{t\in[0,T]}\mathbb{E}|\tilde{X}^{df-RM}_n(t)|^q\leq C. \tag{87}$$

(ii) *There exists a positive constant C, depending only on the parameters of the class $\mathcal{F}(\gamma_1,\gamma_2,q,K)$, such that for all $n\in\mathbb{N}$, $\delta_1,\delta_2\in[0,1]$, $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$, $(\tilde{a},\tilde{b})\in V_a(\delta_1)\times V^1_b(\delta_2)$, we have*

$$\sup_{t\in[0,T]}\mathbb{E}|\tilde{\tilde{X}}^{df-RM}_n(t)|^q\leq C(1+\delta_1^q+\delta_2^q+\delta_2^{2q})\,e^{CT(1+\delta_1^q+\delta_2^q+\delta_2^{2q})}. \tag{88}$$

(iii) *There exists a positive constant C, depending only on the parameters of the class $\mathcal{F}(\gamma_1,\gamma_2,q,K)$ and q, such that for all $n\in\mathbb{N}$, $\delta_1,\delta_2\in[0,1]$, $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$, $(\tilde{a},\tilde{b})\in V_a(\delta_1)\times V^2_b(\delta_2)$, we have*

$$\sup_{t\in[0,T]}\mathbb{E}|\tilde{\tilde{X}}^{df-RM}_n(t)|^q\leq C(1+\delta_1^q+\delta_2^q+(1+\delta_2^q)\delta_2^q n^{q/2})\,e^{CT(1+\delta_1^q+\delta_2^q+(1+\delta_2^q)\delta_2^q n^{q/2})}. \tag{89}$$

Proof. We only show (ii) and (iii), since the proof of (i) is analogous.

Take $(a,b,\eta)\in\mathcal{F}(\gamma_1,\gamma_2,q,K)$, $(\tilde{a},\tilde{b})\in V_a(\delta_1)\times(V^1_b(\delta_2)\cup V^2_b(\delta_2))$. By Lemma 1 we have that the random variables $\tilde{b}(t_i,\tilde{\tilde{X}}^{df-RM}_n(t_i))$, $\mathcal{L}_{1,h}\tilde{b}(t_i,\tilde{\tilde{X}}^{df-RM}_n(t_i))$ are $\tilde{\Sigma}^n_{t_i}$-measurable, while the increment $W(t)-W(t_i)$ is independent of $\tilde{\Sigma}^n_{t_i}$ for all $i=0,1,\ldots,n-1$ and $t\in[t_i,t_{i+1}]$. Additionally, $\|W(t)-W(t_i)\|_q=m_q\cdot(t-t_i)^{1/2}$, and $\|\mathcal{I}_{t_i,t}(W,W)\|_q\leq\frac{1}{2}(m^2_{2q}+1)(t-t_i)$

for $t \in [t_i, t_{i+1}]$, where m_q is the qth root of the qth absolute moment of a normal variable with zero mean and variance equal to 1. This and Lemma 6 give, for all $i = 0, 1, \ldots, n-1$ and $t \in [t_i, t_{i+1}]$, that

$$\begin{aligned}\|\tilde{\bar{X}}_n^{df-RM}(t) - \tilde{\bar{X}}_n^{df-RM}(t_i)\|_q &\le \|\tilde{a}(\xi_i, \tilde{\bar{X}}_n^{df-RM}(t_i))\|_q \cdot (t-t_i) \\ &\quad + \|\tilde{b}(t_i, \tilde{\bar{X}}_n^{df-RM}(t_i))\|_q \cdot \|W(t) - W(t_i)\|_q \\ &\quad + \|\mathcal{L}_{1,h}\tilde{b}(t_i, \tilde{\bar{X}}_n^{df-RM}(t_i))\|_q \cdot \|\mathcal{I}_{t_i,t}(W,W)\|_q \\ &\le C \cdot (1+\delta_1+\delta_2)\cdot(1+\delta_2\cdot\max\{1,h^{-1}\})\cdot(1+\|\tilde{\bar{X}}_n^{df-RM}(t_i)\|_q)\cdot(t-t_i)^{1/2}, \end{aligned} \tag{90}$$

where $C > 0$ depends only on the parameters of the class $\mathcal{F}(\gamma_1, \gamma_2, q, K)$. Since $\|\tilde{\bar{X}}_n^{df-RM}(0)\|_q = \|\eta\|_q \le \|\eta\|_{2q} \le K$, we get by (90) and induction that

$$\max_{i\in\{0,1,\ldots,n\}} \|\tilde{\bar{X}}_n^{df-RM}(t_i)\|_q < +\infty. \tag{91}$$

From (90) and (91) we get that $\sup_{t\in[0,T]} \|\tilde{\bar{X}}_n^{df-RM}(t)\|_q < +\infty$. Therefore, the function $[0,T] \ni t \mapsto \sup_{0\le u\le t} \mathbb{E}|\tilde{\bar{X}}_n^{df-RM}(u)|^q \in \mathbb{R}_+ \cup \{0\}$ is Borel measurable (as a nondecreasing function) and bounded. We now show that we can bound this mapping from above by a finite number that depends only on the parameters of the class $\mathcal{F}(\gamma_1, \gamma_2, q, K)$, δ_1, and δ_2.

We have that for all $t \in [0, T]$

$$\mathbb{E}|\tilde{\bar{X}}_n^{df-RM}(t)|^q \le C(\mathbb{E}|\eta|^q + \mathbb{E}|\tilde{\bar{A}}_n^{df-RM}(t)|^q + \mathbb{E}|\tilde{\bar{B}}_n^{df-RM}(t)|^q).$$

From the Hölder inequality we obtain that

$$\mathbb{E}|\tilde{\bar{A}}_n^{df-RM}(t)|^q \le C_1(1+\delta_1^q) + C_2(1+\delta_1^q)\int_0^t \sum_{i=0}^{n-1} \mathbb{E}|\tilde{\bar{X}}_n^{df-RM}(t_i)|^q \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s.$$

Moreover, by Burkholder inequality

$$\begin{aligned}\mathbb{E}|\tilde{\bar{B}}_n^{df-RM}(t)|^q &\le C_3(1+\delta_2^q) + C_4(1+\delta_2^q)\int_0^t \sum_{i=0}^{n-1} \mathbb{E}|\tilde{\bar{X}}_n^{df-RM}(t_i)|^q \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s \\ &+ C_5 h^{q/2}\int_0^t \sum_{i=0}^{n-1} \mathbb{E}|\mathcal{L}_{1,h}\tilde{b}(\bar{U}_i^{df})|^q \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s,\end{aligned} \tag{92}$$

where, by Lemma 6, we have

$$|\mathcal{L}_{1,h}\tilde{b}(\bar{U}_i^{df})|^q \le C_1\cdot(1+\delta_2^q)\cdot(1+|\tilde{\bar{X}}_n^{df-RM}(t_i)|^q)\cdot \begin{cases} 1+\delta_2^q, & \text{if } p_b \in \mathcal{K}^1_{\text{Lip}} \\ 1+\delta_2^q\cdot h^{-q}, & \text{if } p_b \in \mathcal{K}^2. \end{cases} \tag{93}$$

Therefore, if $p_b \in \mathcal{K}^1_{\text{Lip}}$ we get

$$\begin{aligned}\mathbb{E}|\tilde{\bar{B}}_n^{df-RM}(t)|^q &\le C_1(1+\delta_2^q) + C_2(1+\delta_2^{2q}) \\ &+ C_3(1+\delta_2^q+\delta_2^{2q})\cdot\int_0^t \sum_{i=0}^{n-1} \mathbb{E}|\tilde{\bar{X}}_n^{df-RM}(t_i)|^q \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s\end{aligned} \tag{94}$$

while for $p_b \in \mathcal{K}^2$ it holds that

$$\begin{aligned}\mathbb{E}|\tilde{\bar{B}}_n^{df-RM}(t)|^q &\le C_1(1+\delta_2^q)\cdot(1+\delta_2^q\cdot h^{-q/2}) \\ &+ C_3(1+\delta_2^q)\cdot(1+\delta_2^q\cdot h^{-q/2})\cdot\int_0^t \sum_{i=0}^{n-1} \mathbb{E}|\tilde{\bar{X}}_n^{df-RM}(t_i)|^q \cdot \mathbb{1}_{[t_i,t_{i+1})}(s)\,\mathrm{d}s.\end{aligned} \tag{95}$$

By applying Gronwall's lemma we get the thesis in (ii) and (iii). ■

Finally, we recall the well-known bound on the absolute L^{2q}-moment of the solution X of (1). The following lemma is a direct consequence of Theorems 4.3 and 4.4 in Chapter 2 in [15].

Lemma 8. *There exists a positive constant C, depending only on the parameters of the class $\mathcal{F}(\gamma_1, \gamma_2, q, K)$, such that for all $(a, b, \eta) \in \mathcal{F}(\gamma_1, \gamma_2, q, K)$, we have*

$$\Big\| \sup_{t\in[0,T]} |X(t)| \Big\|_{2q} \le C.$$

References

[1] P.E. Kloeden, E. Platen, Numerical Solution of Stochastic Differential Equations, Springer-Verlag Berlin, Heidelber, 1992.
[2] M. Milanese, A. Vicino, Optimal estimation theory for dynamic systems with set membership uncertainty: an overview, Automatica 27 (1991) 997–1009.
[3] L. Plaskota, Noisy Information and Computational Complexity, Cambridge Univ. Press, Cambridge, 1996.
[4] B. Kacewicz, M. Milanese, A. Vicino, Conditionally optimal algorithms and estimation of reduced order models, J. Complexity 4 (1988) 73–85.
[5] B. Kacewicz, L. Plaskota, On the minimal cost of approximating linear problems based on information with deterministic noise, Numer. Funct. Anal. Optim. 11 (1990) 511–528.
[6] P.M. Morkisz, L. Plaskota, Approximation of piecewise Hölder functions from inexact information, J. Complexity 32 (2016) 122–136.
[7] B. Kacewicz, P. Przybyłowicz, On the optimal robust solution of IVPs with noisy information, Numer. Algorithms 71 (2016) 505–518.
[8] A.G. Werschulz, The complexity of definite elliptic problems with noisy data, J. Complexity 12 (1996) 440–473.
[9] A.G. Werschulz, The complexity of indefinite elliptic problems with noisy data, J. Complexity 13 (1997) 457–479.
[10] P.M. Morkisz, P. Przybyłowicz, Optimal pointwise approximation of SDE's from inexact information, J. Comput. Appl. Math. 324 (2017) 85–100.
[11] A. Kałuza, P.M. Morkisz, P. Przybyłowicz, Optimal approximation of stochastic integrals in analytic noise model, Appl. Math. Comput. 356 (2019) 74–91.
[12] R. Kruse, Y. Wu, A randomized Milstein method for stochastic differential equations with non-differentiable drift coefficients, Discrete Contin. Dyn. Syst. Ser. B 24 (2019) 3475–3502.
[13] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, second ed., Springer - Verlag, New York, 1991.
[14] J.F. Traub, G.W. Wasilkowski, H. Woźniakowski, Information-Based Complexity, Academic Press, New York, 1988.
[15] X. Mao, Stochastic Differential Equations and Applications, second ed., Woodhead Publishing, Cambridge, 2011.
[16] R.A. Brooks, Conditional expectations associated with stochastic processes, Pacific J. Math. 41 (1972) 33–42.
[17] P. Przybyłowicz, P. Morkisz, Strong approximation of solutions of stochastic differential equations with time-irregular coefficients via randomized Euler algorithm, Appl. Numer. Math. 78 (2014) 80–94.
[18] A. Jentzen, A. Neuenkirch, A random Euler scheme for Carathéodory differential equations, J. Comput. Appl. Math. 224 (2009) 346–359.
[19] E. Novak, Deterministic and Stochastic Error Bounds in Numerical Analysis, in: Lecture Notes in Mathematics, vol. 1349, Springer–Verlag, New York, 1988.
[20] S. Heinrich, Lower complexity bounds for parametric stochastic Itô integration, J. Math. Anal. Appl. 476 (2019) 177–195.