Finance Stoch (2009) 13: 381–401
DOI 10.1007/s00780-009-0099-7

On irregular functionals of SDEs and the Euler scheme

Rainer Avikainen

Received: 14 February 2008 / Accepted: 4 September 2008 / Published online: 23 July 2009

Abstract We prove a sharp upper bound for the error $\mathbb{E}|g(X) - g(\hat{X})|^p$ in terms of moments of $X - \hat{X}$, where X and $\hat{X}$ are random variables and the function g is a function of bounded variation. We apply the results to the approximation of a solution to a stochastic differential equation at time T by the Euler scheme, and show that the approximation of the payoff of the binary option has asymptotically sharp strong convergence rate $1/2$. This has consequences for multilevel Monte Carlo methods.

Keywords Stochastic differential equations · Approximation · Rate of convergence · Euler scheme

Mathematics Subject Classification (2000) 60H10 · 41A25 · 26A45 · 65C20 · 65C30

JEL Classification C63 · C65 · G12

1 Introduction

1.1 Motivation

In the theory of mathematical finance, the computation of expected values of payoffs by Monte Carlo methods and the use of backward stochastic differential equations

The author was supported by the Finnish Graduate School in Stochastics and Statistics, the Ellen and Artturi Nyyssönen Foundation, and the Academy of Finland, project #110599.

R. Avikainen (✉)
Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35 (MaD), 40014 Jyväskylä, Finland
e-mail: rainer.avikainen@jyu.fi

(BSDEs) are of particular importance. It turns out that in both areas a certain inequality plays an essential role, namely

$$\left\| g(X_T) - g\left(X_T^\pi\right) \right\|_p^p \le C|\pi|^\gamma, \tag{1.1}$$

where $\gamma > 0$, $1 \le p < \infty$, X is a diffusion, and X_T^π is an approximation of X_T corresponding to a partition π of the interval $[0, T]$ with mesh size $|\pi|$, e.g. the Euler scheme.

The approximation of solutions of SDEs is related to the multilevel Monte Carlo method, which was first introduced by S. Heinrich [14] to approximate parameter dependent integrals in high dimensions. Later on, similar ideas were applied to the SDE setting by A. Kebaier [21] and M. Giles [8, 9]. One purpose of the multilevel Monte Carlo method is to approximate the expected payoff of an option with a small computational cost. Giles' method [9] requires estimates for the variance of $g(X_T) - g(X_T^\pi)$ for possibly non-Lipschitz payoff functions g. Part of the motivation for our work is to investigate in detail the variance in the case of the Euler scheme and the payoff of the binary option.

Consideration of the inequality (1.1) is motivated also by discretization schemes for BSDEs, where the function g appears in the terminal condition. The inequality is responsible for the coupling of the forward and backward parts of some recent numerical algorithms in simulation of BSDEs, and strongly affects the precision of the backward algorithm. This is exploited by C. Geiss et al. [7].

If the function g is Lipschitz, then the inequality (1.1) reduces to the strong convergence rate of the underlying scheme. However, the binary option with payoff function $g(x) = \chi_{[K,\infty)}$ is of importance in both areas mentioned above, and gives a primary example of a situation where an estimate of type (1.1) is needed for a non-Lipschitz function. Our aim is to show that we can get substantial information about (1.1), for a large class of functions g, from existing results on strong convergence of approximation schemes for the solutions of SDEs. This is particularly important in the case that the strong convergence rates are basically the only information available about the scheme.

1.2 Convergence of the underlying scheme

There exists an extensive literature on approximation schemes for stochastic differential equations. P.E. Kloeden and E. Platen [22] show that any order of strong convergence can be achieved by the strong Itô–Taylor approximations, i.e., for any order $\gamma > 0$ there exists a scheme X^π such that

$$\left\| \sup_{0\le t\le T} \left| X_t - X_t^\pi \right| \right\|_1 < C|\pi|^\gamma. \tag{1.2}$$

The most common examples are the Euler scheme and the Milstein scheme, which have the order of strong convergence 1/2 and 1, respectively. Achieving the global error estimate (1.2) requires knowing the complete path of the driving Brownian motion. However, in order to get the same order of convergence for the Euler and Milstein schemes with respect to the pointwise error e.g. at the endpoint T, it suffices to consider the Brownian motion at the discretization points. The errors with respect to both global and pointwise error criteria are considered by N. Hofmann et al. [17, 18], Hofmann and Müller-Gronbach [16], and Müller-Gronbach [23, 25]. The latest result concerning the pointwise error is due to Müller-Gronbach [25], where the author defines certain classes of convergence schemes and finds optimal (adaptive) schemes for each class.

Another point of view is to relax the continuity assumptions of the coefficients σ and b, and consider the convergence of the Euler scheme. I. Gyöngy and N. Krylov [13], Gyöngy [12] and D.J. Higham et al. [15] have presented results in this direction.

1.3 Main results

We develop in Theorem 2.4 a general principle that gives a sharp upper bound for the functional $\mathbb{E}|g(X)-g(\hat{X})|^p$ in terms of moments of $X-\hat{X}$. Here X and $\hat{X}$ are random variables and g is a function of bounded variation, e.g. the payoff of the binary option. The principle implies that if approximations $(X_t^\pi)_{t\in[0,T]}$ satisfy

$$\left\|X_T - X_T^\pi\right\|_p \le C_p^1|\pi|^\gamma \tag{1.3}$$

for some $\gamma > 0$ and all $1 \le p < \infty$, then

$$\left\|g(X_T) - g\left(X_T^\pi\right)\right\|_p^p \le C_p^2|\pi|^{\gamma-\varepsilon} \tag{1.4}$$

for any $0 < \varepsilon < \gamma$ and for any function of bounded variation g. In other words, the convergence result (1.3) automatically gives a convergence rate in (1.4) that is arbitrarily close to the original rate.

For the Euler scheme, where we have $\gamma = 1/2$, and for a sufficiently small mesh size, we show in Theorem 5.4 that in the estimate (1.4), the convergence rate $1/2-\varepsilon$ can be replaced with $1/2-C(-\log|\pi|)^{-1/3}$, which converges to $1/2$ as the mesh size decreases. This we show to be asymptotically sharp in Theorem 7.2, where we obtain a lower bound for the approximation error by considering the geometric Brownian motion. We also apply Theorem 5.4 to the multilevel Monte Carlo method and get an improvement in the mean square error of the multilevel estimator. These results are achieved under certain conditions on the SDE, including the existence of a bounded density for the solution X_T.

Similar results concerning the Euler scheme have been independently obtained by M. Giles et al. [10], who show the convergence rate $1/2-\varepsilon$ for binary options, as well as results for different option types. The estimate for the binary option is now developed further by our Theorem 5.4.

1.4 Organization of the paper

The main result for functions of bounded variation, Theorem 2.4, is presented in Sect. 2 and proved in Sect. 3. The setting for stochastic differential equations and the application to strong Itô-Taylor approximations is presented in Sect. 4. Section 5 contains more specific results obtained for the Euler scheme, and the application to the multilevel MC method follows in Sect. 6. A detailed proof of the application is postponed to Appendix A. A lower bound for the approximation error in the case of the Euler scheme is given in Sect. 7.

This paper is a reduced version of [1], which presents a generalization of (1.4) to a larger class of functions.

2 Functions of bounded variation and moments of random variables

Suppose that we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two random variables $X, \hat{X} : \Omega \to \mathbb{R}$. Consider $\hat{X}$ to be an approximation of X in the L_p-norm. We find an estimate for the functional $\mathbb{E}|g(X) - g(\hat{X})|$ in terms of the pth moment of $X - \hat{X}$, where g is a real-valued function of bounded variation. Let us first recall the definitions of the spaces BV and NBV.

Definition 2.1 Let

$$T_f(x) := \sup \sum_{j=1}^{N} \bigl|f(x_j) - f(x_{j-1})\bigr|,$$

where the supremum is taken over N and all partitions

$$-\infty < x_0 < x_1 < \cdots < x_N = x,$$

be the *total variation function* of f. Then we say that f is a function of *bounded variation*, $f \in BV$, if

$$V(f) := \lim_{x \to \infty} T_f(x)$$

is finite, and call $V(f)$ the *(total) variation* of f.

Definition 2.2 Let NBV, where N stands for normalized, be the set of functions $f \in BV$ such that f is left-continuous and $f(x) \to 0$ as $x \to -\infty$.

Example 2.3 Let $g = \chi_{[K,\infty)}$ be the payoff function of the binary option. Then $g \in BV$, $T_g = g$, and $V(g) = 1$.

Theorem 2.4 *The following assertions hold*:

(i) *Suppose that X has a bounded density f_X. If $g \in BV$ and $1 \le q < \infty$, then for every $1 \le p < \infty$ we have*

$$\bigl\|g(X) - g(\hat{X})\bigr\|_q^q \le 3^{q+1} V(g)^q (\sup f_X)^{\frac{p}{p+1}} \|X - \hat{X}\|_p^{\frac{p}{p+1}}.$$

(ii) *The power $\frac{p}{p+1}$ of the L_p-norm is optimal, i.e., if*

$$\bigl\|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})\bigr\|_1 \le C(X, K, p, r)\|X - \hat{X}\|_p^r$$

for all random variables X with a bounded density, then $r \le \frac{p}{p+1}$.

(iii) *If there exist* $p_0 > 0$ *and* $B_X > 0$ *such that*

$$\left\| \chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X}) \right\|_1 \leq B_X \|X - \hat{X}\|_p^{\frac{p}{p+1}}$$

for all $p_0 \leq p < \infty$, *all* $K \in \mathbb{R}$ *and all random variables* $\hat{X}$, *then* X *has a bounded density.*

3 Proof of Theorem 2.4

The proof of Theorem 2.4 exploits the non-increasing rearrangement of random variables, which we recall first. Using this, we formulate in Lemma 3.4 a statement for indicator functions that is analogous to Theorem 2.4(i). Then we proceed with the proof of Theorem 2.4(i), which is based on Lemma 3.4 and the measure representation of a function of bounded variation.

As in Sect. 2, let X and $\hat{X}$ be random variables defined on a common probability space.

Definition 3.1 The *non-increasing rearrangement* of X, denoted by $X^* : [0,1] \to \mathbb{R} \cup \{+\infty, -\infty\}$, is defined by

$$X^*(s) := \inf\bigl\{c \in \mathbb{R} : \mathbb{P}(X > c) \leq s\bigr\}.$$

Here we use the convention that $\inf \emptyset = \infty$.

Remark 3.2 Definition 3.1 is slightly different from the standard non-increasing rearrangement as defined e.g. in [3, Chap. 2, Definition 1.5], where the absolute value of the function X is taken, and in fact defines the $(1-s)$-quantile of X. However, by analogous arguments we can show the following properties:

1. $X^*(1) = -\infty$ always, $X^*(0) = \infty$ if X is not essentially bounded, and $X^*(s) \in \mathbb{R}$ for $s \in (0,1)$,
2. X^* is right-continuous,
3. X^* has the same distribution as X with respect to the Lebesgue measure on $[0,1]$.

Definition 3.3 Denote the minimal slope of the function X^* from the level K by $d_X : \mathbb{R} \to [0, \infty)$,

$$d_X(K) := \inf_{\substack{s \in [0,1] \\ s \neq \alpha(K)}} \left\{ \frac{|X^*(s) - K|}{|s - \alpha(K)|} \right\},$$

where

$$\alpha(K) = \mathbb{P}(X \geq K).$$

Lemma 3.4 *If* X *has a bounded density* f_X, *then for all* $K \in \mathbb{R}$, *all random variables* $\hat{X}$ *and all* $0 < p < \infty$ *we have*

$$\left\| \chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X}) \right\|_1 \leq 3 D_X(K)^{\frac{p}{p+1}} \|X - \hat{X}\|_p^{\frac{p}{p+1}},$$

where

$$D_X(K) := \frac{1}{d_X(K)} \in (0, \sup f_X].$$

Remark 3.5 Corresponding results for the functions $\chi_{(K,\infty)}$, $\chi_{(-\infty,K]}$ and $\chi_{(-\infty,K)}$ are obtained by considering complements of the intervals in the indicator functions and the random variables $-X$ and $-\hat{X}$.

Remark 3.6 We shall only use the upper bound of the constant $D_X(K)$. However, the information that $D_X(K)$ contains about K can be exploited. This could be an issue of further investigation.

Proof of Lemma 3.4 Fix $K \in \mathbb{R}$ and $0 < p < \infty$, and let $\hat{X}$ be a random variable such that

$$\left\| \chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X}) \right\|_1 = \varepsilon$$

for some $\varepsilon \in (0, 1]$. Define $\varepsilon_1 := \mathbb{P}(X \geq K, \hat{X} < K)$ and $\varepsilon_2 := \mathbb{P}(X < K, \hat{X} \geq K)$, so that $\varepsilon = \varepsilon_1 + \varepsilon_2$. Set $\alpha := \alpha(K) = \mathbb{P}(X \geq K)$. Notice that $\alpha - \varepsilon_1 \geq 0$ and $\alpha + \varepsilon_2 \leq 1$. Now

$$\begin{aligned}
\mathbb{E}|X - \hat{X}|^p &\geq \mathbb{E}|X - \hat{X}|^p \chi_{\{X \geq K, \hat{X} < K\} \cup \{X < K, \hat{X} \geq K\}} \\
&\geq \mathbb{E}|X - K|^p \chi_{\{X \geq K, \hat{X} < K\} \cup \{X < K, \hat{X} \geq K\}} \\
&= \mathbb{E}|X - K|^p \chi_{\{X \geq K, \hat{X} < K\}} + \mathbb{E}|X - K|^p \chi_{\{X < K, \hat{X} \geq K\}}.
\end{aligned}$$

Since X has a bounded density, we can find a number $c_0 \in [K, \infty]$ such that $\mathbb{P}(K \leq X < c_0) = \varepsilon_1$, and so $|\{K \leq X^* < c_0\}| = \varepsilon_1$, where $|\cdot|$ denotes Lebesgue measure. Note that c_0 need not be unique. However, $\{K \leq X < c_0\}$ is a set of probability ε_1 where $\mathbb{E}|X - K|^p \chi_A$ is minimized over all $A \subset \{X \geq K\}$ with $\mathbb{P}(A) = \varepsilon_1$, which implies that

$$\begin{aligned}
\mathbb{E}|X - K|^p \chi_{\{X \geq K, \hat{X} < K\}} &\geq \mathbb{E}|X - K|^p \chi_{[K,c_0)}(X) \\
&= \int_{[0,1]} \left|X^*(s) - K\right|^p \chi_{[K,c_0)}\big(X^*(s)\big)\, ds \\
&= \int_{\alpha - \varepsilon_1}^{\alpha} \left|X^*(s) - K\right|^p ds \\
&\geq \int_0^{\varepsilon_1} \left|d_X(K)s\right|^p ds = \frac{d_X(K)^p \varepsilon_1^{p+1}}{p+1}.
\end{aligned}$$

Similar arguments show that

$$\mathbb{E}|X - K|^p \chi_{\{X < K, \hat{X} \geq K\}} \geq \int_{\alpha}^{\alpha + \varepsilon_2} \left|X^*(s) - K\right|^p ds \geq \frac{d_X(K)^p \varepsilon_2^{p+1}}{p+1}.$$

Thus

$$\mathbb{E}|X-\hat{X}|^p \geq \frac{d_X(K)^p(\varepsilon_1^{p+1}+\varepsilon_2^{p+1})}{p+1} \geq \frac{d_X(K)^p\varepsilon^{p+1}}{2^p(p+1)}. \tag{3.1}$$

Equation (3.1) gives

$$\left\|\chi_{[K,\infty)}(X)-\chi_{[K,\infty)}(\hat{X})\right\|_1 = \varepsilon \leq 2^{\frac{p}{p+1}}(p+1)^{\frac{1}{p+1}}\left(\frac{1}{d_X(K)}\right)^{\frac{p}{p+1}}\left(\mathbb{E}|X-\hat{X}|^p\right)^{\frac{1}{p+1}}.$$

By elementary computations we can show that

$$2^{\frac{p}{p+1}}(p+1)^{\frac{1}{p+1}} \leq 2e^{\frac{1}{2e}} \leq 3.$$

Recalling the definition of D_X from Lemma 3.4, we may write

$$\left\|\chi_{[K,\infty)}(X)-\chi_{[K,\infty)}(\hat{X})\right\|_1 \leq 3D_X(K)^{\frac{p}{p+1}}\|X-\hat{X}\|_p^{\frac{p}{p+1}}.$$

Using the definition of X^* and the boundedness assumption for the density of X we see that $1/d_X(K) \leq \sup f_X$. □

Proof of Theorem 2.4 (i) First we show the result for functions $g \in NBV$. By [26, Theorem 8.14] there is a unique signed measure μ such that

$$g(x) = \mu((-\infty,x)) \text{ and } |\mu|((-\infty,x)) = T_g(x),$$

where $|\mu|$ is the total variation measure of μ, and T_g was defined in Definition 2.1. We consider the Jordan decomposition of μ, i.e., $\mu=\mu_1-\mu_2$, where $\mu_1=\frac{1}{2}(|\mu|+\mu)$ and $\mu_2=\frac{1}{2}(|\mu|-\mu)$ are positive measures. Then $|\mu|=\mu_1+\mu_2$, and all three measures $|\mu|$, μ_1, and μ_2 are finite since $|\mu|(\mathbb{R})=V(g)<\infty$. Thus we have

$$g(x)=\mu((-\infty,x))=\int_{\mathbb{R}}\chi_{(-\infty,x)}(z)\,d\mu(z)=\int_{\mathbb{R}}\chi_{(z,\infty)}(x)\,d\mu(z).$$

By Lemma 3.4 and Remark 3.5,

$$\begin{aligned}
\left\|g(X)-g(\hat{X})\right\|_q &= \left\|\int_{\mathbb{R}}\chi_{(z,\infty)}(X)\,d\mu(z)-\int_{\mathbb{R}}\chi_{(z,\infty)}(\hat{X})\,d\mu(z)\right\|_q \\
&= \left\|\int_{\mathbb{R}}\left[\chi_{(z,\infty)}(X)-\chi_{(z,\infty)}(\hat{X})\right]d\mu(z)\right\|_q \\
&\leq \left\|\int_{\mathbb{R}}\left|\chi_{(z,\infty)}(X)-\chi_{(z,\infty)}(\hat{X})\right|d|\mu|(z)\right\|_q \\
&\leq \int_{\mathbb{R}}\left\|\chi_{(z,\infty)}(X)-\chi_{(z,\infty)}(\hat{X})\right\|_q d|\mu|(z) \\
&\leq 3^{\frac{1}{q}}(\sup f_X)^{\frac{p}{q(p+1)}}V(g)\|X-\hat{X}\|_p^{\frac{p}{q(p+1)}},
\end{aligned}$$

which completes the proof for functions in NBV.

Next, let g be an arbitrary function in BV. By [26, Theorem 8.13], there exists a unique function $\tilde{g} \in NBV$ and a unique constant $c \in \mathbb{R}$ such that $g(x) = \tilde{g}(x) + c$ at all points of continuity of g, and moreover, $V(\tilde{g}) \le V(g)$ and g can have only countably many points of discontinuity. Define $\bigcup_{j\in J}\{a_j\}$ to be the set of these points and let $\lambda_j := g(a_j) - \tilde{g}(a_j) - c$. Then we can write

$$g(x) = \tilde{g}(x) + c + \Delta(x),$$

where

$$\Delta(x) := \sum_{j\in J} \lambda_j \chi_{\{a_j\}}(x) = \sum_{j\in J} \lambda_j \big(\chi_{(-\infty,a_j]}(x) - \chi_{(-\infty,a_j)}(x)\big).$$

We define a measure

$$\nu = \sum_{j\in J} \lambda_j \delta_{a_j},$$

where δ_a is the Dirac measure in a. It follows from [26], Theorem 8.13., that $g(a_j-)$ exists. Thus we have $\tilde{g}(a_j) + c = g(a_j-)$ and

$$|\nu|(\mathbb{R}) = \sum_{j\in J} |\lambda_j| = \sum_{j\in J} \big|g(a_j) - g(a_j-)\big| \le V(g).$$

Now we may write

$$\Delta(x) = \int_{\mathbb{R}} \big(\chi_{(-\infty,z]}(x) - \chi_{(-\infty,z)}(x)\big)\, d\nu(z)$$

and compute, similarly to the NBV case, that

$$\begin{aligned}
\big\|\Delta(X) - \Delta(\hat{X})\big\|_q &\le \left\| \int_{\mathbb{R}} \big|\chi_{(-\infty,z]}(X) - \chi_{(-\infty,z]}(\hat{X})\big|\, d|\nu|(z) \right\|_q \\
&\quad + \left\| \int_{\mathbb{R}} \big|\chi_{(-\infty,z)}(X) - \chi_{(-\infty,z)}(\hat{X})\big|\, d|\nu|(z) \right\|_q \\
&\le \int_{\mathbb{R}} \big\|\chi_{(-\infty,z]}(X) - \chi_{(-\infty,z]}(\hat{X})\big\|_q\, d|\nu|(z) \\
&\quad + \int_{\mathbb{R}} \big\|\chi_{(-\infty,z)}(X) - \chi_{(-\infty,z)}(\hat{X})\big\|_q\, d|\nu|(z) \\
&\le 2 \cdot 3^{\frac{1}{q}} (\sup f_X)^{\frac{p}{q(p+1)}} V(g) \|X - \hat{X}\|_p^{\frac{p}{q(p+1)}}.
\end{aligned}$$

This, combined with the NBV result, implies that

$$\begin{aligned}
\big\|g(X) - g(\hat{X})\big\|_q &= \big\|\tilde{g}(X) - \tilde{g}(\hat{X}) + \Delta(X) - \Delta(\hat{X})\big\|_q \\
&\le \big\|\tilde{g}(X) - \tilde{g}(\hat{X})\big\|_q + \big\|\Delta(X) - \Delta(\hat{X})\big\|_q \\
&\le 3 \cdot 3^{\frac{1}{q}} (\sup f_X)^{\frac{p}{q(p+1)}} V(g) \|X - \hat{X}\|_p^{\frac{p}{q(p+1)}},
\end{aligned}$$

which gives the statement.

(ii) To see the optimality of the power, we construct an example where the lower bound given by (3.1) is achieved. Suppose that $\Omega = [0,1]$ is equipped with Lebesgue measure, $K = \frac{1}{2}$ and $\varepsilon < 1$. Letting $X(\omega) = \omega$ we see that X has a bounded density and $d_X(\frac{1}{2}) = 1$. Define

$$\hat{X} = \begin{cases} X, & \text{if } \omega \in [0, \frac{1}{2} - \frac{\varepsilon}{2}) \cup (\frac{1}{2} + \frac{\varepsilon}{2}, 1], \\ X + \frac{\varepsilon}{2}, & \text{if } \omega \in [\frac{1}{2} - \frac{\varepsilon}{2}, \frac{1}{2}], \\ X - \frac{\varepsilon}{2}, & \text{if } \omega \in (\frac{1}{2}, \frac{1}{2} + \frac{\varepsilon}{2}]. \end{cases}$$

Then

$$\mathbb{E}|X - \hat{X}|^p = \mathbb{E}\left|\frac{\varepsilon}{2}\right|^p \chi_{[\frac{1}{2}-\frac{\varepsilon}{2}, \frac{1}{2}+\frac{\varepsilon}{2}]}(X) = \frac{\varepsilon^{p+1}}{2^p},$$

so by the assumption we have, for all $0 < \varepsilon < 1$, that

$$\varepsilon = \left\| \chi_{[\frac{1}{2},\infty)}(X) - \chi_{[\frac{1}{2},\infty)}(\hat{X}) \right\|_1 \le C(X, 1/2, p, r) \left(\frac{\varepsilon^{\frac{p+1}{p}}}{2} \right)^r,$$

which implies $r \le \frac{p}{p+1}$.

(iii) Let $\delta > 0$ and choose $\hat{X} = X - \delta$. Then

$$\begin{aligned} \mathbb{E}\left|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})\right| &= \mathbb{P}(X \ge K,\ X - \delta < K) + \mathbb{P}(X < K,\ X - \delta \ge K) \\ &= \mathbb{P}(K \le X < K + \delta), \end{aligned}$$

so that by assumption, for $p > p_0$, we have

$$\mathbb{P}(K \le X < K + \delta) \le B_X \delta^{\frac{p}{p+1}}.$$

We let p tend to infinity and conclude that

$$\mathbb{P}(K \le X < K + \delta) \le B_X \delta.$$

Let $N \subset \mathbb{R}$ be a null set with respect to Lebesgue measure and let $\varepsilon > 0$. We can find a sequence (I_j) of open intervals such that $N \subset \bigcup I_j$ and $\sum |I_j| \le \varepsilon$. Let $\mathcal{L}_X$ be the law of X. Then we have

$$\mathcal{L}_X((a,b)) \le \mathcal{L}_X([a,b)) \le B_X |b - a|$$

and

$$\mathcal{L}_X(N) \le \mathcal{L}_X\left(\bigcup_j I_j\right) \le \sum_j \mathcal{L}_X(I_j) \le B_X \sum_j |I_j| \le B_X \varepsilon.$$

This implies that $\mathcal{L}_X(N) = 0$, so $\mathcal{L}_X$ is absolutely continuous with respect to Lebesgue measure. By the Radon–Nikodým theorem there exists a probability density $f : \mathbb{R} \to [0, \infty)$ such that

$$\mathcal{L}_X(M) = \int_M f(x)\,dx$$

for all measurable $M \subseteq \mathbb{R}$. Define a distribution function $\Phi : \mathbb{R} \to [0, 1]$ such that

$$\Phi(t) = \int_{-\infty}^{t} f(x)\, dx.$$

Then by [26, Theorem 8.17], we have that $\Phi'(t) = f(t)$ a.e. in $\mathbb{R}$. On the other hand, we have that

$$\Phi'(t) = \lim_{h\to 0} \frac{\Phi(t+h) - \Phi(t)}{h} \le \lim_{h\to 0} \frac{B_X h}{h} = B_X \quad \text{a.e. in } \mathbb{R},$$

because $\Phi(t+h) - \Phi(t) = \mathcal{L}_X((t, t+h))$. Therefore we conclude that $f(t) \le B_X$ a.e. in $\mathbb{R}$. □

4 Setting for SDEs

The results of Sect. 2 can be applied directly to the pointwise approximation of solutions of stochastic differential equations. We start by defining the setting. We fix a terminal time $T > 0$ and suppose that $(W_t)_{t\in[0,T]}$ is a standard one-dimensional Brownian motion defined on a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in[0,T]})$, where the filtration is the augmentation of the natural filtration of W and $\mathcal{F} = \mathcal{F}_T$.

We consider a diffusion process X, which is a solution to

$$\begin{aligned} dX_t &= \sigma(t, X_t)\, dW_t + b(t, X_t)\, dt, \\ X_0 &= x_0 \end{aligned} \tag{4.1}$$

with $x_0 \in \mathbb{R}$ and continuous coefficients $\sigma, b : [0, T] \times \mathbb{R} \to \mathbb{R}$. We assume that for $f \in \{\sigma, b\}$ there exist constants C_T and $\alpha \ge \frac{1}{2}$ such that

(i) $|f(t, x) - f(t, y)| \le C_T |x - y|$,
(ii) $|f(t, x) - f(s, x)| \le C_T (1 + |x|)|t - s|^{\alpha}$.

Assumptions (i) and (ii) imply the existence of a unique adapted strong solution X of the SDE (4.1), see e.g. [20, Section 5.2 B]. Moreover, we assume that

(iii) X_T has a bounded density.

Remark 4.1 Assumption (iii) is satisfied if we assume that $\sigma, b \in C_b^{\infty}([0, T] \times \mathbb{R})$ and σ satisfies the uniform ellipticity condition, i.e., there exists a constant β such that

$$\sigma(t, x) \ge \beta > 0 \text{ for all } (t, x) \in [0, T] \times \mathbb{R}.$$

See [6, Chap. 9, Theorem 8]. Another sufficient condition is given by Caballero et al. [5, Theorem 2]. They assume that σ and b are C^2 in x, the second derivatives have polynomial growth, the functions $|\sigma(0, x)|$, $|\sigma_x(t, x)|$, $|b(0, x)|$ and $|b_x(t, x)|$ are bounded, and

$$\mathbb{E}\left(\left| \int_0^t \sigma(s, X_s)^2\, ds \right|^{-p_0/2} \right) < \infty$$

for some $p_0 > 2$ and for all $t \in (0, T]$. Then there exists a continuous density f_{X_t} of X_t such that for all $p > 1$

$$f_{X_t}(x) \le C_p \left\| \left(\int_0^t \sigma(s, X_s)^2 \, ds \right)^{-1/2} \right\|_p$$

for some constant $C_p > 0$.

Denote by π a partition $0 = t_0 < t_1 < \cdots < t_n = T$ of the interval $[0, T]$, and let

$$|\pi| = \max_{0 \le i < n} |t_{i+1} - t_i|$$

be the mesh size of π. Moreover, denote an approximation of X corresponding to π by X^π. As an immediate consequence of Theorem 2.4, we can derive

Corollary 4.2 *Let X be the solution of* (4.1), $1 \le q < \infty$, *and* $g \in BV$. *Suppose that X_T has a bounded density,* $1 \le p < \infty$, *and X_T^π is an approximation of X_T such that*

$$\left\| X_T - X_T^\pi \right\|_p \le C_p |\pi|^\gamma \tag{4.2}$$

for some $\gamma > 0$ and some constant $C_p \ge 0$. Then

$$\left\| g(X_T) - g\left(X_T^\pi\right) \right\|_q^q \le 3^{q+1} (\sup f_{X_T})^{\frac{p}{p+1}} V(g)^q C_p^{\frac{p}{p+1}} |\pi|^{\frac{\gamma p}{p+1}}.$$

Remark 4.3 Assuming that (4.2) holds for all $1 \le p < \infty$, Corollary 4.2 gives the asymptotic convergence rate $\gamma - \varepsilon$ for any $\varepsilon > 0$ and is applicable to all appropriate strong Taylor approximation schemes, see [22]. Two such schemes are the well-known Euler and Milstein schemes. For the Euler scheme the rate of strong convergence is $\gamma = 1/2$, which is later given in Theorem 5.3. Under certain assumptions, for the Milstein scheme it is $\gamma = 1$ [24, Chap. V, Proposition 1].

5 Euler scheme

In the case of the Euler scheme, Corollary 4.2 gives the convergence rate $\frac{1}{2} - \varepsilon$. In this section we improve it by replacing ε by an explicit formula in terms of $\log |\pi|$ for small mesh size $|\pi|$. First recall the definition:

Definition 5.1 (Euler scheme) Let X^E be the Euler scheme relative to π, i.e., $X_0^E = x_0$, and for $i = 0, \ldots, n-1$,

$$X_{t_{i+1}}^E = X_{t_i}^E + \sigma\left(t_i, X_{t_i}^E\right)(W_{t_{i+1}} - W_{t_i}) + b\left(t_i, X_{t_i}^E\right)(t_{i+1} - t_i).$$

Given the values at the partition points, we also define the Euler scheme in continuous time by setting

$$X_t^E = X_{t_k}^E + \sigma\left(t_k, X_{t_k}^E\right)(W_t - W_{t_k}) + b\left(t_k, X_{t_k}^E\right)(t - t_k)$$

for $t \in (t_k, t_{k+1})$.

Remark 5.2 The Euler approximation of X_T, denoted X_T^E, always depends on the corresponding partition π. This is omitted from the notation for simplicity.

The improvement of Corollary 4.2 in the case of the Euler scheme is based on the following statement.

Theorem 5.3 *If the assumptions* (i) *and* (ii) *in Sect.* 4 *hold, and* $1 \le p < \infty$, *then*

$$\Big\| \sup_{0\le t\le T} \big|X_t - X_t^E\big| \Big\|_p \le e^{Mp^2} |\pi|^{\frac{1}{2}},$$

where the constant $M > 0$ *depends at most on* x_0, T *and* C_T.

Proof We omit the proof, and refer the reader to [4, Chap. 5, Theorem B.1.4.], where the result is stated without computing the constant explicitly. See also [1, Theorem A.1]. □

Using the information about the constant in Theorem 5.3, we can write an extended version of Corollary 4.2 for the Euler scheme:

Theorem 5.4 *Let* $1 \le p < \infty$ *and* $g \in BV$. *Then there exists* $m \in (0,1)$ *such that for* $|\pi| < m$ *we have*

$$\big\|g(X_T) - g\big(X_T^E\big)\big\|_p^p \le 3^p (\sup f_{X_T} \vee \sqrt{\sup f_{X_T}}) V(g)^p |\pi|^{\frac{1}{2} - \frac{2+M}{(-\log|\pi|)^{1/3}}},$$

where the constant $M = M(x_0, T, C_T)$ *is taken from Theorem* 5.3.

Proof By Theorem 2.4, Theorem 5.3, and the formula $a^{\frac{p}{p+1}} \le a \vee \sqrt{a}$ for $a > 0$ and $p \ge 1$, we get that for all $1 \le p < \infty$ and $q \ge 1$,

$$\begin{aligned}\big\|g(X_T) - g\big(X_T^E\big)\big\|_q^q &\le 3^{q+1} (\sup f_{X_T} \vee \sqrt{\sup f_{X_T}}) V(g)^q e^{Mp^2 \cdot \frac{p}{p+1}} |\pi|^{\frac{p}{2(p+1)}} \\ &\le 3^{q+1} (\sup f_{X_T} \vee \sqrt{\sup f_{X_T}}) V(g)^q e^{Mp^2} |\pi|^{\frac{p}{2(p+1)}}.\end{aligned}$$

Now choose p such that

$$4p(p+1)^2 = -\log|\pi|$$

for $|\pi| \le m$ with $m = e^{-16}$. This gives $p^3 \le -\log|\pi|$ and $p^2 \le (-\log|\pi|)^{2/3}$. Thus we have

$$e^{Mp^2} \le e^{M(-\log|\pi|)^{2/3}} = |\pi|^{-M(-\log|\pi|)^{-1/3}}$$

and

$$\frac{1}{2(p+1)} = \sqrt{\frac{p}{-\log|\pi|}} \le \sqrt{\big(-\log|\pi|\big)^{1/3-1}} = \big(-\log|\pi|\big)^{-1/3}.$$

Using the above we get

$$3e^{Mp^2} |\pi|^{\frac{p}{2(p+1)}} = 3e^{Mp^2} |\pi|^{\frac{1}{2} - \frac{1}{2(p+1)}} \le 3|\pi|^{\frac{1}{2} - \frac{1+M}{(-\log|\pi|)^{1/3}}} \le |\pi|^{\frac{1}{2} - \frac{2+M}{(-\log|\pi|)^{1/3}}},$$

where in the last step we used the inequality

$$3|\pi|^{\frac{1}{(-\log|\pi|)^{1/3}}} \le 1$$

for $|\pi| < m$. We conclude that

$$\left\| g(X_T) - g\left(X_T^E\right)\right\|_q^q \le 3^q \left(\sup f_{X_T} \vee \sqrt{\sup f_{X_T}}\right) V(g)^q |\pi|^{\frac{1}{2} - \frac{2+M}{(-\log|\pi|)^{1/3}}}.$$

□

Remark 5.5 The power of $|\pi|$ in Theorem 5.4 may be negative for some mesh sizes $|\pi| < m$, but it is positive if we make a more restrictive choice of m, i.e., $m = e^{-2^3(2+M)^3}$.

Remark 5.6 We could apply a similar technique to the Milstein scheme or any other strong Taylor approximation, if we proved the L_p-estimate corresponding to Theorem 5.3 with an explicit constant.

6 Application to the multilevel Monte Carlo method

The multilevel Monte Carlo method introduced by Giles [9] requires a variance estimate for the difference of the payoff and its approximation. Corollary 4.2 gives the parameter $\beta = \gamma - \varepsilon$ in the variance assumption (iii) of [9, Theorem 3.1] for a payoff of bounded variation, in particular for the binary option, and any approximation scheme satisfying the moment estimate (4.2) for all $1 \le p < \infty$. We now show how the estimate on the Euler scheme given by Theorem 5.4 applies in this setting. In the following, we use the notation of Giles and refer the reader to [9] for details.

Giles considers a multilevel estimator

$$\widehat{Y} = \sum_{\ell=0}^{L} \widehat{Y}_\ell$$

of the expected value of the payoff, $\mathbb{E}P$, on L levels of refining discretizations with timesteps $h_\ell = M^{-\ell}T$. Let $\widehat{P}_\ell$ be an approximation of P using a numerical discretization with timestep h_ℓ. On each level the estimator $\widehat{Y}_\ell$ of $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$, or of $\mathbb{E}\widehat{P}_0$ for $\ell = 0$, uses N_ℓ Monte Carlo samples. The variance of $\widehat{Y}_\ell$ for $\ell > 0$ has the form $V(\widehat{Y}_\ell) = N_\ell^{-1} V_\ell$, where $V_\ell = V(\widehat{P}_\ell - \widehat{P}_{\ell-1})$ is the variance of a single sample, and $\widehat{P}_{\ell-1}$ uses the same realization as $\widehat{P}_\ell$ with timestep $h_{\ell-1}$. Abusing notation, the differences $(\widehat{P}_\ell - \widehat{P}_{\ell-1})_{\ell=1}^L$ and the estimators $(\widehat{Y}_\ell)_{\ell=0}^L$ are independent. To be proper, one should write $\widehat{P}_\ell - \widehat{P}_{\ell-1} = \widehat{P}^\ell_\ell - \widehat{P}^\ell_{\ell-1}$. Let $\alpha \geq 1/2$ be the parameter in assumption (i) of [9, Theorem 3.1], and choose $T = 1$ for simplicity. For the Euler scheme and a fixed step size parameter $M \geq 2$ we can replace assumption (iii) in [9, Theorem 3.1] by

(iii′) $V(\widehat{Y}_\ell) \leq c N_\ell^{-1} M^{-\frac{\ell}{2}+\frac{A\,\ell}{((\ell \log M)\vee B)^{1/3}}}$,

where $c, A, B > 0$ depend at most on the diffusion X, g, and M. Applying this to Giles' method gives the following result, which is proved in Appendix A.

Theorem 6.1 *Let $\alpha \geq 1/2$ and $\varepsilon < \min\{\sqrt{2}c_1, 1/e\}$, with $c_1 > 0$ taken from assumption* (i) *in* [9, *Theorem* 3.1], *and let the underlying numerical discretization be the Euler scheme. If the assumptions of* [9, *Theorem* 3.1] *hold except for* (iii), *then the inequality* (iii′) *holds, and the computational complexity of the multilevel estimator $\widehat{Y}$ is bounded by $c_4' \varepsilon^{-2-\frac{1}{2\alpha}}$, whereas the mean square error of $\widehat{Y}$ satisfies the upper bound*

$$MSE(\widehat{Y}) := \mathbb{E}\big[\big(\widehat{Y} - \mathbb{E}[P]\big)^2\big] \leq \frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2}\Phi(\varepsilon).$$

Here

$$\Phi(\varepsilon) := \big(1 - M^{-\frac{1}{4}}\big) M^{-\frac{L(\varepsilon)}{4}} \sum_{\ell=0}^{L(\varepsilon)} M^{\frac{\ell}{4}+(D\ell^{\frac{2}{3}}\vee E)},$$

where $D, E > 0$ depend at most on the diffusion X and M, and

$$L(\varepsilon) := \left\lceil \frac{\log(\sqrt{2}c_1\varepsilon^{-1})}{\alpha \log M} \right\rceil.$$

Moreover, for all $\delta > 0$,

$$\lim_{\varepsilon \downarrow 0} \Phi(\varepsilon)\varepsilon^\delta = 0.$$

Remark 6.2 The weak convergence parameter $\alpha \geq 1/2$ can be determined. We refer to the results on weak convergence in [27], [2], and [11], which together show that $\alpha = 1$ for a rather large class of functions, and thus the complexity is of order $\varepsilon^{-2-1/2\alpha} = \varepsilon^{-5/2}$. Our results are for functions of bounded variation, which form a subspace of the space of measurable and bounded functions considered in [2]. We should remark that the weak convergence results require stronger assumptions, e.g. hypoellipticity in [2] and uniform ellipticity in [11].

7 Lower bound

In this section we find a solution X_1 (i.e., $T = 1$) of an SDE of the type (4.1) such that it gives a lower bound for the approximation error of the Euler scheme in Theorem 5.4. This is achieved by choosing $X = S$, the geometric Brownian motion. Let

$S_t = e^{W_t - t/2}$ for $t \in [0, 1]$, so that S is a solution of

$$S_t = 1 + \int_0^t S_s \, dW_s,$$

and let $U^n := S^E - S$, where S^E is the Euler scheme as defined in Definition 5.1 corresponding to the equidistant partition of $[0, 1]$, i.e., $\pi = (i/n)_{i=0}^n$.

Lemma 7.1 *We have $(W, \sqrt{n}U^n) \Longrightarrow (W, U)$ in the Skorohod topology, where U is the unique strong L_2-solution of the equation*

$$U_t = \int_0^t U_s \, dW_s - \frac{1}{\sqrt{2}} \int_0^t S_s \, dB_s \tag{7.1}$$

and B is a standard Brownian motion independent of W.

Proof The statement is an immediate consequence of [19, Corollary 5.4]. □

The following theorem states that the convergence rate in Theorem 5.4 is optimal up to the logarithmic term.

Theorem 7.2 *There exists $K_0 > 0$ such that*

$$\liminf_{n\to\infty} \sqrt{n} \sup_{K \geq K_0} \left\| \chi_{[K,\infty)}(S_1) - \chi_{[K,\infty)}\left(S_1^E\right) \right\|_1 > 0,$$

where S_1^E is the equidistant Euler approximation of S_1.

Proof Consider the setting of Lemma 7.1 and the process U defined by (7.1). If $U_1 = 0$ a.s., then for all $t \in [0, 1]$ we have $U_t = 0$ a.s., which leads to a contradiction. Therefore $\mathbb{P}(U_1 > 0) > 0$ or $\mathbb{P}(U_1 < 0) > 0$. If $\mathbb{P}(U_1 > 0) > 0$, then there exist $\varepsilon \in (0, 1]$, $\delta > 0$ and $K \geq 1 + K_0$ with $K_0 > 0$ such that

$$\mathbb{P}\big(S_1 \in [K-1, K),\ U_1 > \varepsilon\big) = \delta.$$

The case $\mathbb{P}(U_1 < 0) > 0$ can be treated in a similar way by changing the condition that $U_1 > \varepsilon$ to $U_1 < -\varepsilon$. By Lemma 7.1 we know that $(W, \sqrt{n}U^n) \Rightarrow (W, U)$ in the Skorohod topology. This implies that $(W_1, \sqrt{n}U_1^n) \Rightarrow (W_1, U_1)$, since the projection mapping π_1, i.e., the mapping $\alpha \mapsto \alpha(1)$ for a process α, is continuous in the Skorohod topology. Because the function $e^{x-\frac{t}{2}}$ is continuous, we have $(S_1, \sqrt{n}U_1^n) \Rightarrow (S_1, U_1)$. Therefore

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{P}\big(S_1 \in [K-1, K), \sqrt{n}[S_1^E - S_1] > \varepsilon\big) \\
&\quad = \liminf_{n\to\infty} \mathbb{P}\big(S_1 \in (K-1, K), \sqrt{n}\, U_1^n > \varepsilon\big) \\
&\quad \geq \mathbb{P}\big(S_1 \in (K-1, K), U_1 > \varepsilon\big) \\
&\quad = \mathbb{P}\big(S_1 \in [K-1, K), U_1 > \varepsilon\big).
\end{aligned}$$

We see that there exists $n_0 \geq 1$ such that for all $n \geq n_0$,

$$\mathbb{P}\left(S_1 \in [K-1, K),\ S_1^E - S_1 > \frac{\varepsilon}{\sqrt{n}}\right) \geq \frac{\delta}{2}.$$

For $m \geq 1$ and any partition $K - 1 = K_0^m < K_1^m < \cdots < K_m^m = K$, we have

$$\sup_{\ell=1,\ldots,m} \mathbb{P}\left(S_1 \in \left[K_{\ell-1}^m, K_\ell^m\right),\ S_1^E - S_1 > \frac{\varepsilon}{\sqrt{n}}\right) \geq \frac{\delta}{2m}.$$

Now choose the partition $(K_\ell^m)_{\ell=0}^m$ to be equidistant with

$$\frac{1}{m} \leq \frac{\varepsilon}{\sqrt{n}}. \tag{7.2}$$

Then there exists $\ell_0 \in \{1, \ldots, m\}$ such that

$$\frac{\delta}{2m} \leq \mathbb{P}\left(S_1 \in \left[K_{\ell_0-1}^m, K_{\ell_0}^m\right),\ S_1^E > S_1 + \frac{\varepsilon}{\sqrt{n}}\right) \leq \mathbb{P}\left(S_1 < K_{\ell_0}^m,\ S_1^E \geq K_{\ell_0}^m\right).$$

Let $m = \lceil \sqrt{n}/\varepsilon \rceil$, which satisfies the condition (7.2) for the mesh size. Thus

$$\frac{\delta}{2\lceil \sqrt{n}/\varepsilon \rceil} \leq \mathbb{P}\left(S_1 < K_{\ell_0}^m,\ S_1^E \geq K_{\ell_0}^m\right) \leq \left\| \chi_{[K_{\ell_0}^m,\infty)}(S_1) - \chi_{[K_{\ell_0}^m,\infty)}\left(S_1^E\right) \right\|_1.$$

Since $\lceil \sqrt{n}/\varepsilon \rceil \leq 2\sqrt{n}/\varepsilon$, we have

$$\left\| \chi_{[K_{\ell_0}^m,\infty)}(S_1) - \chi_{[K_{\ell_0}^m,\infty)}\left(S_1^E\right) \right\|_1 \geq \frac{\delta}{2\lceil \sqrt{n}/\varepsilon \rceil} \geq \frac{\delta\varepsilon}{4\sqrt{n}}.$$

Therefore

$$\sqrt{n} \sup_{K \geq K_0} \left\| \chi_{[K,\infty)}(S_1) - \chi_{[K,\infty)}\left(S_1^E\right) \right\|_1 \geq \frac{\delta\varepsilon}{4}$$

for all $n \geq n_0$, which implies the assertion. □

Acknowledgements The author wishes to thank his advisor Professor Stefan Geiss for numerous invaluable discussions during the preparation of the manuscript, and the referee for helpful comments.

Appendix A: Proof of Theorem 6.1

We point out that the proof of Theorem 6.1 follows the proof of Giles [9, Theorem 3.1], and advise the reader to look for notation and details in [9].

Proof We can divide the MSE in two parts by computing

$$MSE(\widehat{Y}) = \mathbb{E}\left[(\widehat{Y} - \mathbb{E}P)^2\right] = (\mathbb{E}\widehat{Y} - \mathbb{E}P)^2 + \mathbb{E}\left[(\widehat{Y} - \mathbb{E}\widehat{Y})^2\right] = Bias^2(\widehat{Y}) + V(\widehat{Y}),$$

where the second equality is valid because of orthogonality.

Let us first consider the bias term. Following the proof of Giles [9], we see that if we set

$$L(\varepsilon) = \left\lceil \frac{\log(\sqrt{2}c_1\varepsilon^{-1})}{\alpha \log M} \right\rceil, \tag{A.1}$$

then, as [9, (6)] states,

$$\frac{1}{\sqrt{2}} M^{-\alpha}\varepsilon < c_1 h_L^{\alpha} \leq \frac{1}{\sqrt{2}}\varepsilon. \tag{A.2}$$

Then by the properties (ii) and (i) of Giles [9, Theorem 3.1] and (A.2),

$$(\mathbb{E}\widehat{Y} - \mathbb{E}P)^2 = (\mathbb{E}\widehat{P}_L - \mathbb{E}P)^2 \leq \big(c_1 h_L^{\alpha}\big)^2 \leq \varepsilon^2/2.$$

This proves the estimate for the bias term.

Let us then consider the variance term. Let $1 \leq p < \infty$. Theorem 5.4 and its proof show that if $|\pi| < m = e^{-16}$, then

$$\big\| g(X_T) - g\big(X_T^E\big)\big\|_p^p \leq C_1(p, T, X, g)|\pi|^{\frac{1}{2} - \frac{C_2(x_0, T, C_T)}{(-\log|\pi|)^{1/3}}}.$$

On the other hand, Corollary 4.2 implies that, for any $\delta > 0$,

$$\big\| g(X_T) - g\big(X_T^E\big)\big\|_p^p \leq C_3(p, T, X, g, \delta)|\pi|^{\frac{1}{2} - \delta}$$

for all mesh sizes $|\pi| > 0$. We choose

$$\delta = \frac{C_2}{(-\log m)^{1/3}} = \frac{C_2}{16^{1/3}}.$$

As $|\pi| \leq m = e^{-16}$ implies $-\log|\pi| \geq -\log m = 16$, this implies that for all mesh sizes $|\pi| > 0$,

$$\big\| g(X_T) - g\big(X_T^E\big)\big\|_p^p \leq C_5(p, T, X, g, x_0, C_T)|\pi|^{\frac{1}{2} - \frac{C_2(x_0, T, C_T)}{(-\log|\pi| \vee 16)^{1/3}}}.$$

By definition, $|\pi| = h_\ell = M^{-\ell}$. We plug this into the estimate and get

$$\big\| g(X_T) - g\big(X_T^E\big)\big\|_p^p \leq C_5(p, T, X, g, x_0, C_T)\big(M^{-\ell}\big)^{\frac{1}{2} - \frac{C_2(x_0, T, C_T)}{(\ell \log M \vee 16)^{1/3}}} =: \psi(\ell). \tag{A.3}$$

Let us now assume that $V(\widehat{Y}_\ell) = N_\ell^{-1} V_\ell$, where V_ℓ is the variance of a single sample. Then by Minkowski's inequality, for $\ell \geq 1$,

$$V_\ell = V(\widehat{P}_\ell - \widehat{P}_{\ell-1}) \leq \Big(\sqrt{V(\widehat{P}_\ell - P)} + \sqrt{V(\widehat{P}_{\ell-1} - P)}\Big)^2,$$

where both of the variance terms on the right-hand side can be bounded from above by $\psi(\ell)$. Indeed, first,

$$V(\widehat{P}_\ell - P) \leq \mathbb{E}(\widehat{P}_\ell - P)^2 \leq \psi(\ell),$$

where we apply the result (A.3) for $p=2$. Similarly, $V(\widehat{P}_{\ell-1}-P)\le\psi(\ell-1)$, but here we would like to have $\psi(\ell)$ instead of $\psi(\ell-1)$. Now

$$\begin{aligned}\psi(\ell-1)&=C_5\big(M^{-\ell+1}\big)^{\frac{1}{2}-\frac{C_2}{((\ell-1)\log M\vee 16)^{1/3}}}\\&=C_5\big(M^{-\ell}\big)^{\frac{1}{2}-\frac{C_2}{((\ell-1)\log M\vee 16)^{1/3}}}M^{\frac{1}{2}-\frac{C_2}{((\ell-1)\log M\vee 16)^{1/3}}}\\&\le C_5\big(M^{-\ell}\big)^{\frac{1}{2}-\frac{C_2}{((\ell\log M-\log M)\vee 16)^{1/3}}}\cdot M\\&\le C_6(p,T,X,g,x_0,C_T,M)\big(M^{-\ell}\big)^{\frac{1}{2}-\frac{C_2}{(\ell\frac{\log M}{2}\vee 16)^{1/3}}},\end{aligned}$$

where the last inequality follows from the fact that for $\ell\ge 2$,

$$\ell\log M-\log M\ge\ell\frac{\log M}{2},$$

and for $\ell=1$ we can increase the constant C_6 if $(\log M)/2\ge 16$, and otherwise we could use 16 in the estimate. Collecting the above results, we get that

$$V_\ell\le C_7(p,T,X,g,x_0,C_T,M)\big(M^{-\ell}\big)^{\frac{1}{2}-\frac{C_2}{(\ell\frac{\log M}{2}\vee 16)^{1/3}}}.$$

Note that by adjusting the constant C_7, the term $V_0:=V(\widehat{P}_0)\le\sup_{x\in\mathbb{R}}g(x)^2<\infty$ also satisfies the above estimate. Therefore

$$V(\widehat{Y}_\ell)=N_\ell^{-1}V_\ell\le C_7N_\ell^{-1}h_\ell^{\frac{1}{2}-\frac{C_2}{(l\frac{\log M}{2}\vee 16)^{1/3}}},$$

and

$$V(\widehat{Y})\le\sum_{\ell=0}^{L}C_7N_\ell^{-1}h_\ell^{\frac{1}{2}-\frac{C_2}{(\ell\frac{\log M}{2}\vee 16)^{1/3}}}.$$

Now we need to choose N_ℓ. We make the same choice as Giles in [9, proof of Theorem 3.1, c)] with $\beta=1/2$ (except for the constant; Giles has c_2, we have C_7), i.e.,

$$N_\ell=\big\lceil 2\varepsilon^{-2}C_7h_L^{-1/4}\big(1-M^{-1/4}\big)^{-1}h_\ell^{3/4}\big\rceil.$$

We plug this into the variance estimate to get

$$\begin{aligned}V(\widehat{Y})&\le\frac{1}{2}\varepsilon^2h_L^{1/4}\big(1-M^{-1/4}\big)\sum_{\ell=0}^{L}h_\ell^{-3/4}h_\ell^{\frac{1}{2}-\frac{C_2}{(\ell\frac{\log M}{2}\vee 16)^{1/3}}}\\&\le\frac{1}{2}\varepsilon^2\underbrace{\big(1-M^{-1/4}\big)M^{-L/4}\sum_{\ell=0}^{L}M^{\frac{\ell}{4}+(D\ell^{2/3}\vee E)}}_{:=\Phi(M^{-L})},\end{aligned}$$

where

$$D = \frac{C_2}{(\frac{\log M}{2})^{1/3}}, \qquad E = \frac{C_2 l_0}{16^{1/3}},$$

and

$$\ell_0 = \max\left\{\ell \geq 0 : \frac{\ell \log M}{2} < 16\right\}.$$

We would like to ensure that $\Phi(M^{-L})$ does not grow too rapidly as L goes to infinity. Let $\delta > 0$. Then for some $\gamma > 0$ we get

$$\Phi\big(M^{-L}\big)M^{-\alpha L\delta} = \big(1 - M^{-1/4}\big)M^{-L/4}M^{-\alpha L\delta}\sum_{\ell=0}^{L} M^{\frac{\ell}{4}+\gamma\ell}M^{-\gamma\ell+(D\ell^{2/3}\vee E)}.$$

Since

$$M^{-\gamma\ell+(D\ell^{2/3}\vee E)} \leq F(M,\gamma,D,E),$$

where F is a constant with respect to ℓ, we see that for $\gamma < \alpha\delta$,

$$\begin{aligned}
\Phi\big(M^{-L}\big)M^{-\alpha L\delta} &\leq \big(1 - M^{-1/4}\big)M^{-L/4}M^{-\alpha L\delta}F\sum_{\ell=0}^{L} M^{\frac{\ell}{4}+\gamma\ell} \\
&\leq \big(1 - M^{-1/4}\big)M^{-L/4}M^{-\alpha L\delta}F(L+1)M^{L/4}M^{\gamma L} \\
&= \big(1 - M^{-1/4}\big)F(L+1)M^{(\gamma-\alpha\delta)L} \to 0 \quad \text{as} \quad L \to \infty.
\end{aligned}$$

Since we can always choose $\gamma < \alpha\delta$, we deduce that

$$\Phi\big(M^{-L}\big)M^{-\alpha L\delta} \to 0 \quad \text{as} \quad L \to \infty$$

for any $\delta > 0$.

Recall the choice of $L(\varepsilon)$ in (A.1). Solving for ε in the inequality (A.2) gives

$$\sqrt{2}c_1 M^{-\alpha L} \leq \varepsilon < \sqrt{2}c_1 M^{\alpha} \cdot M^{-\alpha L}.$$

We define

$$\Phi(\varepsilon) := \Phi\big(M^{-L(\varepsilon)}\big),$$

and conclude that

$$\Phi(\varepsilon)\varepsilon^{\delta} = \Phi\big(M^{-L(\varepsilon)}\big)\varepsilon^{\delta} < \big(\sqrt{2}c_1 M^{\alpha}\big)^{\delta}\Phi\big(M^{-L}\big)M^{-\alpha L\delta} \longrightarrow 0$$

for any $\delta > 0$ as $L \to \infty$ or, equivalently, $\varepsilon \to 0$.

So we have analyzed also the variance term, and thus proved that

$$MSE(\widehat{Y}) \leq \frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2}\Phi(\varepsilon),$$

where $\Phi(\varepsilon)$ satisfies

$$\lim_{\varepsilon \downarrow 0} \Phi(\varepsilon)\varepsilon^{\delta} = 0$$

for all $\delta > 0$.

It remains to show the complexity estimate

$$C(\widehat{Y}) \leq c_4' \varepsilon^{-2-\frac{1}{2\alpha}}, y$$

where $c_4'(p, T, X, g, x_0, C_T, M, c_1, c_3, \alpha)$ is the constant in [9] with c_2 replaced by C_7. As we made the same choice of N_ℓ as Giles did, with $\beta = 1/2$ and C_7 instead of c_2, except for these two changes, our proof is exactly the same as in the latter part of Theorem 3.1(c) in [9]. □

References

1. Avikainen, R.: Convergence rates for approximations of functionals of SDEs. arXiv:0712.3635v1 [math.PR] (2007)
2. Bally, V., Talay, D.: The law of the Euler scheme for stochastic differential equations, I: convergence rate of the distribution function. Probab. Theory Relat. Fields **104**, 43–60 (1996)
3. Bennett, C., Sharpley, R.: Interpolation of Operators. Academic Press, New York (1988)
4. Bouleau, N., Lépingle, D.: Numerical Methods for Stochastic Processes. Wiley, New York (1994)
5. Caballero, M.E., Fernández, B., Nualart, D.: Estimation of densities and applications. J. Theor. Probab. **11**, 831–851 (1998)
6. Friedman, A.: Partial Differential Equations of Parabolic Type. Prentice-Hall, New York (1964)
7. Geiss, C., Geiss, S., Gobet, E.: Generalized fractional smoothness and L_p-variation of BSDEs with non-Lipschitz terminal condition (2008, in preparation)
8. Giles, M.B.: Improved multilevel Monte Carlo convergence using the Milstein scheme. In: Keller, A., Heinrich, S., Niederreiter, H. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2006, pp. 343–358. Springer, Berlin (2008)
9. Giles, M.B.: Multilevel Monte Carlo path simulation. Oper. Res. **56**, 607–617 (2008)
10. Giles, M.B., Higham, D.J., Mao, X.: Analyzing multi-level Monte Carlo for options with non-globally Lipschitz payoff. Finance Stoch. **13**, 403–414 (2009)
11. Guyon, J.: Euler scheme and tempered distributions. Stoch. Process. Appl. **116**, 877–904 (2006)
12. Gyöngy, I.: A note on Euler's approximations. Potential Anal. **8**, 205–216 (1998)
13. Gyöngy, I., Krylov, N.: Existence of strong solutions for Itô's stochastic equations via approximations. Probab. Theory Relat. Fields **105**, 143–158 (1996)
14. Heinrich, S.: Multilevel Monte Carlo methods. In: Margenov, S., et al. (eds.) Large-Scale Scientific Computing, Third International Conference. Lecture Notes in Computer Science, vol. 2179, pp. 58–67. Springer, Berlin (2001)
15. Higham, D.J., Mao, X., Stuart, A.M.: Strong convergence of Euler-type methods for nonlinear stochastic differential equations. SIAM J. Numer. Anal. **40**, 1041–1063 (2002)
16. Hofmann, N., Müller-Gronbach, T.: On the global error of Itô–Taylor schemes for strong approximation of scalar stochastic differential equations. J. Complex. **20**, 732–752 (2004)
17. Hofmann, N., Müller-Gronbach, T., Ritter, K.: The optimal discretization of stochastic differential equations. J. Complex. **17**, 117–153 (2001)
18. Hofmann, N., Müller-Gronbach, T., Ritter, K.: Linear vs. standard information for scalar stochastic differential equations. J. Complex. **18**, 394–414 (2002)
19. Jacod, J., Protter, P.: Asymptotic error distributions for the Euler method for stochastic differential equations. Ann. Probab. **26**, 267–307 (1998)
20. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus, 2nd edn. Springer, Berlin (1991)
21. Kebaier, A.: Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. Ann. Appl. Probab. **15**, 2681–2705 (2005)

22. Kloeden, P.E., Platen, E.: Numerical Solutions of Stochastic Differential Equations. Springer, Berlin (1992)
23. Müller-Gronbach, T.: The optimal uniform approximation of systems of stochastic differential equations. Ann. Appl. Probab. **12**, 664–690 (2002)
24. Müller-Gronbach, T.: Strong approximation of systems of stochastic differential equations. Habilitation thesis, TU Darmstadt (2002)
25. Müller-Gronbach, T.: Optimal pointwise approximation of SDEs based on Brownian motion at discrete points. Ann. Appl. Probab. **14**, 1605–1642 (2004)
26. Rudin, W.: Real and Complex Analysis, 2nd edn. McGraw-Hill, New York (1974)
27. Talay, D., Tubaro, L.: Expansion of the global error for numerical schemes solving stochastic differential equations. Stoch. Anal. Appl. **8**, 483–509 (1990)