ON IRREGULAR FUNCTIONALS OF SDES AND THE EULER SCHEME

RAINER AVIKAINEN

ABSTRACT. We prove a sharp upper bound for the approximation 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 of 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.

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

$$\mathbb{E}|g(X_T) - g(X_T^\pi)|^p \leq C \, |\pi|^\gamma \,, \qquad (1.1)$$

where $\gamma > 0$, $1 \leq p < \infty$, X_T 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 introduced by M. Giles [7], [8]. One purpose of the multilevel Monte Carlo method is to approximate the expected payoff of an option with a small computational cost. Giles' method 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 part of some recent numerical algorithms in simulation

of BSDEs, and strongly affects the precision of the backward algorithm. This is exploited by C. Geiss and S. Geiss [6].

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 [18] 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

$$\mathbb{E}\left(\sup_{0 \le t \le T} |X_t - X_t^\pi|\right) < C\,|\pi|^\gamma.$$

The most common examples are the Euler scheme and the Milstein scheme, which have the order of strong convergence $1/2$ and 1, respectively. Errors with respect to both global and pointwise error criteria are considered by N. Hofmann, T. Müller-Gronbach, and K. Ritter in [13, 14], Hofmann and Müller-Gronbach in [15], and Müller-Gronbach in [20, 21]. The latest result concerning the pointwise error is due to Müller-Gronbach [21], 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 [10], Gyöngy [11] and D. J. Higham et al. [12] 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

$$\|X_T - X_T^\pi\|_p \le C_p^1\,|\pi|^\gamma \tag{1.2}$$

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

$$\mathbb{E}|g(X_T) - g(X_T^\pi)|^p \le C_p^2\,|\pi|^{\gamma - \varepsilon} \tag{1.3}$$

for any $0 < \varepsilon < \gamma$ and for any function of bounded variation g. In other words, the convergence result (1.2) automatically gives a convergence rate in (1.3) 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 4.4 that in the estimate (1.3), 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 6.2, where we obtain a lower bound for the approximation error by considering the geometric Brownian motion. We also apply Theorem 4.4 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, D. J. Higham and X. Mao [9], 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 4.4.

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

This paper is a reduced version of [1], which presents a generalization of (1.3) 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

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

where the supremum is taken over N and all partitions $-\infty < x_0 < x_1 < \ldots < x_N = x$, 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 and $\hat{X}$ are random variables and X has a bounded density f_X. If $g \in BV$ and $1 \le q < \infty$, then for any $1 \le p < \infty$ we have*

$$\mathbb{E}|g(X) - g(\hat{X})|^q \le 3^{q+1}V(g)^q (\sup f_X)^{\frac{p}{p+1}} \left\| X - \hat{X} \right\|_p^{\frac{p}{p+1}}.$$

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

$$\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| \le C(X,K,p,r) \left\| X - \hat{X} \right\|_p^r$$

for all random variables X and $\hat{X}$, then $r = \frac{p}{p+1}$.

(iii) *Let X be a random variable. If there exist $p_0 > 0$ and $B_X > 0$ such that*

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

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

Proof. The proof is presented in Appendix A. $\qquad\square$

3. Setting for SDEs

The results of Section 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

$$\begin{cases} dX_t = \sigma(t, X_t)\, dW_t + b(t, X_t)\, dt, \\ X_0 = x_0 \end{cases} \tag{3.1}$$

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) $|f(t,x)| \le C_T(1 + |x|)$,

(ii) $|f(t,x) - f(t,y)| \le C_T|x - y|$

(iii) $|f(t,x) - f(s,x)| \le C_T(1 + |x|)|t - s|^\alpha$.

Assumptions (i) and (ii) imply the existence of a unique adapted strong solution X of the SDE (3.1), see e.g. [17]. Note also that (ii) and (iii) imply (i). Moreover, we assume that

(iv) X_T has a bounded density.

Remark 3.1. Assumption (iv) 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 [5, p. 263]. Another sufficient condition is given by Caballero et al. [4, 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 $0 = t_0 < t_1 < \ldots < t_n = T$ 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 3.2. *Let X be the solution of the equation (3.1), $1 \le q < \infty$, 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*

$$\|X_T - X_T^\pi\|_p \le C_p |\pi|^\gamma \tag{3.2}$$

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

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

Remark 3.3. Assuming that (3.2) holds for all $1 \le p < \infty$, Corollary 3.2 gives the asymptotic convergence rate $\gamma - \varepsilon$ for any $\varepsilon > 0$ and is applicable to all appropriate strong Taylor approximation schemes, see [18]. 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 4.3. Under certain assumptions, for the Milstein scheme it is $\gamma = 1$ [19, p. 140].

4. EULER SCHEME

In case of the Euler scheme, Corollary 3.2 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 4.1 (Euler scheme). Let X^E be the Euler scheme relative to π, i.e. $X_0^E = x_0$, and for $i = 0, \ldots, n-1$,

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

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

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

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

Remark 4.2. 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 3.2 in the case of the Euler scheme is based on the following statement.

Theorem 4.3. *If the assumptions (i) and (ii) in Section 3 hold, and $1 \leq p < \infty$, then*

$$\left\| \sup_{0 \leq t \leq T} |X_t - X_t^E| \right\|_p \leq e^{Mp^2} |\pi|^{\frac{1}{2}},$$

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 [2, pp. 275-276], where the result is stated without computing the constant explicitly. See also [1, Theorem A.1]. $\square$

Using the information about the constant in Theorem 4.3, we can write an extended version of Corollary 3.2 for the Euler scheme:

Theorem 4.4. *Let $1 \leq p < \infty$ and $g \in BV$. Then there exists $m \in (0,1)$ such that for $|\pi| < m$ we have*

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

where the constant $M = M(x_0, T, C_T)$ is taken from Theorem 4.3.

Proof. By Theorem 2.4, Theorem 4.3, and the formula $a^{\frac{q}{q+1}} \leq a \vee \sqrt{a}$ for $a > 0$ and $q \geq 1$, we get that for all $1 \leq q < \infty$ and $p \geq 1$,

$$\mathbb{E}|g(X_T) - g(X_T^E)|^p \leq 3^{p+1}(\sup f_{X_T} \vee \sqrt{\sup f_{X_T}})V(g)^p e^{Mq^2 \cdot \frac{q}{q+1}} |\pi|^{\frac{q}{2(q+1)}}$$

$$\leq 3^{p+1}(\sup f_{X_T} \vee \sqrt{\sup f_{X_T}})V(g)^p e^{Mq^2} |\pi|^{\frac{q}{2(q+1)}}.$$

Now choose q such that

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

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

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

and

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

Using the above we get

$$3e^{Mq^2} |\pi|^{\frac{q}{2(q+1)}} = 3e^{Mq^2} |\pi|^{\frac{1}{2} - \frac{1}{2(q+1)}} \leq 3|\pi|^{\frac{1}{2} - \frac{1+M}{(-\log|\pi|)^{1/3}}} \leq |\pi|^{\frac{1}{2} - \frac{2+M}{(-\log|\pi|)^{1/3}}},$$

where in the last step we used the inequality

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

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

$$\mathbb{E}|g(X_T) - g(X_T^E)|^p \leq 3^p(\sup f_{X_T} \vee \sqrt{\sup f_{X_T}})V(g)^p |\pi|^{\frac{1}{2} - \frac{2+M}{(-\log|\pi|)^{1/3}}}.$$

$\square$

Remark 4.5. 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 4.3 with an explicit constant.

5. Application to the Multilevel Monte Carlo Method

The multilevel Monte Carlo method introduced by Giles [7] requires a variance estimate for the difference of the payoff and its approximation. Corollary 3.2 gives the parameter $\beta = \gamma - \varepsilon$ in [7, Theorem 3.1 iii)] in the case of a payoff of bounded variation, especially for the binary option, and any approximation scheme satisfying the moment estimate (3.2). We now show how our estimate on the Euler scheme, Theorem 4.4, applies in this setting. In the following, we strictly keep the notation of Giles and refer the reader to [7] for details.

For the Euler scheme (and for simplicity for $T = 1$) and a fixed step size parameter $M \geq 2$ we can replace condition iii) in [7, Theorem 3.1] by

iii) $V[\widehat{Y}_l] \leq c_2 N_l^{-1} M^{-\frac{l}{2} + \frac{Al}{((l \log M) \vee B)^{1/3}}}$,

where $A, B > 0$ depend at most on the diffusion X and M. Using the computation in the proof of [7, Theorem 3.1] for $\alpha = 1$ and $\beta = 1/2$, we obtain that the computational complexity is bounded by $c_4 \varepsilon^{-\frac{5}{2}}$, whereas the mean square error of the multilevel estimator can be bounded from above as follows:

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

Here

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

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

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

with $c_1 > 0$ taken from [7]. We note that for all $\delta > 0$,

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

This follows from the definition of Φ and the inequality (6) in [7], which says that

$$\frac{\varepsilon}{\sqrt{2}M} < c_1 M^{-L(\varepsilon)} \leq \frac{\varepsilon}{\sqrt{2}}.$$

6. Lower bound

In this section we find a solution X_1 (i.e. $T = 1$) of an SDE of the type (3.1) such that it gives a lower bound for the approximation error of the Euler scheme in Theorem 4.4. This is achieved by choosing $X_t = S_t$, 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 4.1 corresponding to the equidistant partition of $[0, 1]$, i.e. $\pi = (i/n)_{i=0}^n$.

Lemma 6.1. *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*

$$U_t = \int_0^t U_s \, dW_s - \frac{1}{\sqrt{2}} \int_0^t S_s \, dB_s \tag{6.1}$$

and B is a standard Brownian motion independent of W.

Proof. The statement is an immediate consequence of a result by Jacod and Protter [16, Corollary 5.4]. □

Theorem 6.2. *There exists $K_0 > 0$ such that*

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

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

Remark 6.3. Theorem 6.2 states that the convergence rate in Theorem 4.4 is optimal up to the logarithmic term.

Proof of Theorem 6.2. Let us consider the setting of Lemma 6.1 and the process U defined by the equation (6.1). 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

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

The case $\mathbb{P}(U_1 < 0) > 0$ can be treated in a similar way by changing the condition $U_1 > \varepsilon$ to $U_1 < -\varepsilon$. By Lemma 6.1 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

$$\liminf_{n\to\infty} \mathbb{P}\left(S_1 \in [K-1, K), \sqrt{n}[S_1^E - S_1] > \varepsilon\right)$$
$$= \liminf_{n\to\infty} \mathbb{P}\left(S_1 \in (K-1, K), \sqrt{n}\, U_1^n > \varepsilon\right)$$
$$\geq \mathbb{P}\left(S_1 \in (K-1, K), U_1 > \varepsilon\right)$$
$$= \mathbb{P}\left(S_1 \in [K-1, K), U_1 > \varepsilon\right).$$

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

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

Now choose the partition $(K_l^m)_{l=1}^m$ to be equidistant with

$$\frac{1}{m} \leq \frac{\varepsilon}{\sqrt{n}}. \tag{6.2}$$

Then there exists $l_0 \in \{1,\dots,m\}$ such that

$$\frac{\delta}{2m} \leq \mathbb{P}\left(S_1 \in [K_{l_0-1}^m, K_{l_0}^m), \ S_1^E > S_1 + \frac{\varepsilon}{\sqrt{n}}\right)$$
$$\leq \mathbb{P}\left(S_1 < K_{l_0}^m, \ S_1^E \geq K_{l_0}^m\right).$$

Let $m = \lceil \sqrt{n}/\varepsilon \rceil$, which satisfies the condition (6.2) for the mesh size. Thus

$$\frac{\delta}{2\lceil \sqrt{n}/\varepsilon \rceil} \leq \mathbb{P}\left(S_1 < K_{l_0}^m, \ S_1^E \geq K_{l_0}^m\right)$$

$$\leq \mathbb{E}|\chi_{[K_{l_0}^m,\infty)}(S_1) - \chi_{[K_{l_0}^m,\infty)}(S_1^E)|.$$

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

$$\mathbb{E}|\chi_{[K_{l_0}^m,\infty)}(S_1) - \chi_{[K_{l_0}^m,\infty)}(S_1^E)| \geq \frac{\delta}{2\lceil \sqrt{n}/\varepsilon \rceil} \geq \frac{\delta\varepsilon}{4\sqrt{n}}.$$

Therefore

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

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

APPENDIX A. 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 A.4 a statement for indicator functions that is analoguous to Theorem 2.4 (i). Then we proceed with the proof of Theorem 2.4 (i), which is based on Lemma A.4 and the measure representation of a function of bounded variation.

Definition A.1. The non-increasing rearrangement of a random variable X, notated by $X^* : [0,1] \to \mathbb{R} \cup \{+\infty, -\infty\}$, is defined

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

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

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

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

Definition A.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 A.4. *Suppose that X is a random variable. Then the following assertion holds: 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*

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

where

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

Remark A.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 A.6. We will 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 A.4. Fix $K \in \mathbb{R}$ and $0 < p < \infty$, and let $\hat{X}$ be a random variable such that

$$\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| = \varepsilon$$

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 $|\{K \leq X^* < c_0\}| = \varepsilon_1$. Note that c_0 may 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

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

Similar arguments show that

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

Thus

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

The equation (A.1) gives

$$\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| = \varepsilon$$

$$\leq 2^{\frac{p}{p+1}}(p+1)^{\frac{1}{p+1}}\left(\frac{1}{d_X(K)}\right)^{\frac{p}{p+1}}\left(\mathbb{E}|X - \hat{X}|^p\right)^{\frac{1}{p+1}}.$$

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 A.4, we may write

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

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 [22, Thm. 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 A.4 and Remark A.5,

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

which completes the proof for functions in NBV.

Next, let g be an arbitrary function in BV. By [22, Thm. 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, $V(\tilde{g}) \leq V(g)$ and g can have only countably many points of discontinuity. Define $\cup_{j \in J}\{a_j\}$ 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

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

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 [22] that $g(a_j-)$ exists. Thus we have $\tilde{g}(a_j) + c = g(a_j-)$ and

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

Now we may write

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

and compute, similarly to the NBV case, that

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

This, combined with the NBV result, implies that

$$\left\| g(X) - g(\hat{X}) \right\|_q = \left\| \tilde{g}(X) - \tilde{g}(\hat{X}) + \Delta(X) - \Delta(\hat{X}) \right\|_q$$
$$\leq \left\| \tilde{g}(X) - \tilde{g}(\hat{X}) \right\|_q + \left\| \Delta(X) - \Delta(\hat{X}) \right\|_q$$
$$\leq 3 \cdot 3^{\frac{1}{q}} (\sup f_X)^{\frac{p}{q(p+1)}} V(g) \left\| X - \hat{X} \right\|_p^{\frac{p}{q(p+1)}},$$

which gives the statement. $\qquad \square$

Proof of Theorem 2.4 (ii). To see the optimality of the power, we construct an example where the lower bound given by equation (A.1) is achieved. Suppose that $\Omega = [0,1]$ is equipped with the 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

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

which implies $r = \frac{p}{p+1}$. $\qquad\square$

Proof of Theorem 2.4 (iii). Let $\delta > 0$ and choose $\hat{X} = X - \delta$. Then

$$\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})|$$
$$= \mathbb{P}(X \geq K, \ X - \delta < K) + \mathbb{P}(X < K, \ X - \delta \geq K)$$
$$= \mathbb{P}(K \leq X < K + \delta),$$

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

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

We let p tend to infinity and conclude that

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

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

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

and

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

This implies that $\mathcal{L}_X(N) = 0$, so $\mathcal{L}_X$ is absolutely continuous with respect to the Lebesgue measure. By the Radon-Nikodym 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 [22, Thm. 8.17], we have that $\Phi'(t) = f(t)$ a.e. in $\mathbb{R}$. On the other hand, we have that

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

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

References

[1] AVIKAINEN, R., *Convergence Rates for Approximations of Functionals of SDEs.* arXiv:0712.3635v1 [math.PR], 12/2007.

[2] BOULEAU, N., LÉPINGLE, D., *Numerical Methods for Stochastic Processes.* Wiley, 1994.

[3] BENNETT, C., SHARPLEY, R., *Interpolation of Operators.* Academic Press, 1988.

[4] CABALLERO, M.E., FERNÁNDEZ, B, NUALART, D., *Estimation of Densities and Applications.* J. Theoret. Probab. 11 (1998), no. 3, 831 – 851.

[5] FRIEDMAN, A., *Partial Differential Equations of Parabolic Type.* Prentice-Hall, 1964.

[6] GEISS, C., GEISS, S., *On Discrete-time Approximation of BSDEs and Besov Spaces.* Preprint, 2007.

[7] GILES, M.B., *Multilevel Monte Carlo Path Simulation.* Technical Report NA06/03, Oxford University Computing Laboratory, 2006.

[8] GILES, M.B., *Improved Multilevel Monte Carlo Convergence Using the Milstein Scheme.* Technical Report NA06/22, Oxford University Computing Laboratory, 2006.

[9] GILES, M.B., HIGHAM, D.J., MAO, X., *Analyzing Multi-level Monte Carlo for Options with Non-globally Lipschitz Payoff.* Preprint, 2008.

[10] GYÖNGY, I., KRYLOV, N., *Existence of Strong Solutions for Itô's Stochastic Equations via Approximations.* Probab. Theory Related Fields 105 (1996), 143 – 158.

[11] GYÖNGY, I., *A Note on Euler's Approximations.* Potential Anal. 8 (1998), 205 – 216.

[12] HIGHAM, D.J., MAO, X., STUART, A.M., *Strong Convergence of Euler-type Methods for Nonlinear Stochastic Differential Equations.* Siam J. Numer. Anal. Vol. 40 No. 3, 1041 – 1063.

[13] HOFMANN, N., MÜLLER-GRONBACH, T., RITTER, K., *The Optimal Discretization of Stochastic Differential Equations.* J. Complexity 17 (2001), No. 1, 117 – 153.

[14] HOFMANN, N., MÜLLER-GRONBACH, T., RITTER, K., *Linear vs. Standard Information for Scalar Stochastic Differential Equations.* J. Complexity 18 (2002), 394 – 414.

[15] HOFMANN, N., MÜLLER-GRONBACH, T., *On the Global Error of Itô-Taylor Schemes for Strong Approximation of Scalar Stochastic Differential Equations.* J. Complexity 20 (2004), 732 – 752.

[16] JACOD, J., PROTTER, P., *Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations.* Ann. Prob. 26 (1998), no. 1, 267–307.

[17] KARATZAS, I., SHREVE, S.E., *Brownian Motion and Stochastic Calculus, Second Edition*, Springer-Verlag, 1991.

[18] KLOEDEN, P.E., PLATEN, E., *Numerical Solutions of Stochastic Differential Equations.* Springer-Verlag, 1992.

[19] MÜLLER-GRONBACH, T., *Strong Approximation of Systems of Stochastic Differential Equations.* Habilitation thesis. Darmstadt, 2002.

[20] MÜLLER-GRONBACH, T., *The Optimal Uniform Approximation of Systems of Stochastic Differential Equations.* Ann. Appl. Probab. 12 (2002), no.2, 664 – 690.

[21] MÜLLER-GRONBACH, T., *Optimal Pointwise Approximation of SDEs Based on Brownian Motion at Discrete Points.* Ann. Appl. Probab. 14 (2004), no. 4, 1605–1642.

[22] RUDIN, W., *Real and Complex Analysis, Second Edition.* McGraw-Hill, 1966, 1974.

DEPARTMENT OF MATHEMATICS AND STATISTICS, P. O. BOX 35 (MAD), FIN-40014 UNIVERSITY OF JYVÄSKYLÄ, FINLAND.

E-mail address: `avikain@maths.jyu.fi`