ON GENERALIZED BOUNDED VARIATION AND APPROXIMATION OF SDES

RAINER AVIKAINEN

ON GENERALIZED BOUNDED VARIATION AND APPROXIMATION OF SDES

RAINER AVIKAINEN

Acknowledgements

What a journey it was. Although not unlike a common story of a Ph.D. student, I feel priviledged to have experienced all this. Moments of joy, despair, and excitement. Misty morning arrival in Petroskoi station in a Russian night train, winter sun in freezing Mekrijärvi, Mediterranean heat in La Manga, noisy streets of Bucharest. Many meetings and discussions with friends and colleagues. These years have been both the most rewarding and the most challenging times of my life, and leave me with a long list of people I am thankful to.

I wish to express my gratitude to everyone at the Department of Mathematics and Statistics, it has been a pleasure to work with you all. I am especially grateful to my advisor, professor Stefan Geiss, without whom none of this would have happened. It is hard to find words to describe how much I appreciate all the help and stream of ideas he gave me, and his ability to spot my mistakes with a sharp eye. I also wish to thank Christel Geiss for being my co-advisor and for helping me whenever needed, Anni Toivola for sharing the experience of Ph.D. studies and providing a constant supply of chocolate, Kevin Wildrick for advice with language (all mistakes I added after his proofreading!) and occasionally math as well, and Antti Viholainen for friendship and company for coffee breaks. Also, Tuula Blåfield, Eira Henriksson, and Hannele Säntti–Ahomäki deserve special thanks for help in practical matters.

I would like to thank Lasse Leskelä, Andreas Neuenkirch, and Klaus Ritter for valuable comments and corrections that greatly improved the content and presentation of this thesis.

For financial support I am indebted to the Finnish Graduate School in Stochastics and Statistics, the Academy of Finland, and the Ellen and Artturi Nyyssönen Foundation.

I thank my family and friends for support throughout the years. In particular, I thank my dear wife Hanna, for all that she is to me. It is a true blessing to be loved by her. Finally, praise be to the name of God for ever and ever; wisdom and power are his.

Jyväskylä, October 2009

Rainer Avikainen

List of included articles

This dissertation consists of an introductory part and the following publications:

[1] AVIKAINEN R. On irregular functionals of SDEs and the Euler scheme. *Finance Stoch. 13*, 3 (2009), 381–401.

[2] AVIKAINEN R. On generalized bounded variation and approximation of SDEs. Preprint 383, Department of Mathematics and Statistics, University of Jyväskylä, 2009.

In the introductory part these articles will be referred to as [1] and [2], respectively. Moreover, the following appendix referred to in [1] and [2] is attached for completeness:

[A] AVIKAINEN R. Appendix of the paper
Convergence rates for approximations of functionals of SDEs.
arXiv:0712.3635v1 [math.PR] (2007).

Introduction

1. Overview

In this thesis we find optimal estimates for the error $\mathbb{E}|g(X)-g(\widehat{X})|^p$ in terms of moments of $X - \widehat{X}$, where X and $\widehat{X}$ are random variables, and the function g satisfies minimal regularity assumptions. These results, of theoretical interest as such, have natural applications in option pricing and discretization of backward stochastic differential equations. In particular, our results justify the use of the multilevel Monte Carlo method in the case of options with irregular payoff, e.g. the binary option, and make it possible to approximate backward stochastic differential equations with irregular payoff function in the terminal condition.

In the following sections we explain this in more detail. Throughout this introduction we take a fixed 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 $\mathcal{F} = \mathcal{F}_T$.

2. Stochastic Differential Equations

The theory of stochastic differential equations (SDEs) dates back to the 1940's and the work of K. Itô, related to the study of diffusion processes. In the last decades, SDEs have become a common modelling tool in various fields including mathematical finance, physics, and population biology. We consider the Itô SDE

$$\begin{cases} dX_t = \sigma(t, X_t)\, dW_t + b(t, X_t)\, dt, & t \in [0, T], \\ X_0 = x_0, \end{cases} \tag{2.1}$$

with $x_0 \in \mathbb{R}$ and continuous coefficients $\sigma, b : [0, T] \times \mathbb{R} \to \mathbb{R}$. A process $X : [0, T] \times \mathbb{R} \to \mathbb{R}$ is called a strong solution of the SDE (2.1), if X is continuous and adapted, $X_0 = x_0$ and X satisfies the integral equation

$$X_t = X_0 + \int_0^t \sigma(s, X_s)\, dW_s + \int_0^t b(s, X_s)\, ds$$

for all $t \in [0, T]$, almost surely. If the coefficients σ and b satisfy Lipschitz and linear growth assumptions in the state variable, then there exists a unique strong solution X of the SDE (2.1). We advise the reader to look for further information in the wide literature concerning the theory of SDEs. See e.g. [20], [22], [28], and [29].

3. OPTIONS AND OPTION PRICING

In mathematical finance, stochastic differential equations are an essential tool in modelling stock prices and, consequently, appear in option pricing. Suppose that a price of a share is given by $(X_t)_{t\in[0,T]}$, a solution of the SDE (2.1), and X_T is the price of the share at the time T. The term 'option' commonly refers to a contract that gives its holder the right, but not the obligation, to buy or sell a particular asset, e.g. a share, at an agreed time and price. Some 'exotic' options do not necessarily involve buying or selling the underlying asset, but can be more complex financial derivatives. We consider European options, which can be exercised only at the maturity time T, and the profit gained by the holder is given by $g(X_T)$, where g is called the payoff function.

An important example in this thesis is a binary call option with a strike price $K > 0$, i.e. an option that gives its holder a fixed amount of cash, say 1 euro, if the price of the share X_T exceeds K at time T, and otherwise gives nothing. The payoff is then

$$g(x) = \begin{cases} 1, & x \geq K, \\ 0, & x < K, \end{cases}$$

which is not continuous at the point of the strike price K.

It is natural to ask for a fair price of such an option. At time T the price should obviously be $g(X_T)$, as the value X_T is known. At any time $0 \leq t \leq T$, a natural price would be $\mathbb{E}_{\mathbb{Q}}(g(X_T) \mid \mathcal{F}_t)$, the conditional expectation of the payoff given the information known at time t, with respect to an equivalent martingale measure $\mathbb{Q}$. This can be written as a solution of a certain PDE according to Feynman–Kac theory. In particular, at time $t = 0$ the price should be $\mathbb{E}_{\mathbb{Q}} g(X_T)$. To compute this, we can also use Monte Carlo methods. However, as the SDE (2.1) often can not be solved explicitly, and thus the distribution of X_T is not available, we need to find numerical approximations for X_T.

4. APPROXIMATION OF SDES

A natural way to approximate X_T is to take a partition $0 = t_0 < t_i < \cdots < t_n = T$ of the interval $[0, T]$, and call it π with mesh size $|\pi| = \max_{0 \leq i \leq n-1}(t_{i+1} - t_i)$. Then we construct an approximation X_T^π of X_T by using simulated values of the driving Brownian motion and the information given by the coefficients σ and b of the SDE (2.1) at the points of the partition π. We look for a numerical scheme X_T^π that converges to X_T under a certain convergence criterion, as the mesh size goes to zero. However, there is no uniquely optimal way to do the construction, as the nature of the approximation problem affects the choice of the convergence criterion. Two most common criteria are the

strong and the weak convergence. In the following, we consider the standard L_p–spaces equipped with the norm $\|\cdot\|_p = (\mathbb{E}|\cdot|^p)^{1/p}$.

We say that an approximation X_T^π converges strongly to X_T in L_p with order $\gamma > 0$, if there exists a constant $C_p > 0$ such that

$$\|X_T - X_T^\pi\|_p \le C_p |\pi|^\gamma$$

for all mesh sizes $|\pi|$. We say that X_T^π converges weakly to X_T with order $\gamma > 0$, if, for any g in a class $\mathcal{C}$ of test functions, there exists a constant $C(g) > 0$ such that

$$|\mathbb{E}g(X_T) - \mathbb{E}g(X_T^\pi)| \le C(g)|\pi|^\gamma$$

for all mesh sizes $|\pi|$. For the time being, we choose $\mathcal{C}$ to be the class of all polynomials.

The essential difference in the criteria is that the strong convergence requires us to generate a pathwise approximation of the solution X_T, whereas for the weak convergence it is enough to approximate the distribution of X_T.

The simplest example of a numerical scheme is the Euler scheme, sometimes called Euler-Maruyama scheme. It is a generalization of the deterministic Euler scheme and was first introduced by G. Maruyama [23] in 1953. Let us first fix a partition π of the interval $[0, T]$. Then we define the Euler scheme at $t = 0$ to be $X_0^E = x_0$, and recursively for $i = 0, \ldots, n-1$ we define

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

This gives a random variable X_T^E approximating X_T. The Euler scheme has strong order $1/2$, given that the coefficients of the SDE are Hölder continuous with an exponent $\alpha \ge 1/2$ with respect to the time variable, and weak order 1, under additional regularity assumptions on the co-efficients. Sometimes the order can be improved in a subclass of SDEs. For example, we can reach strong order 1 for SDEs with additive noise, i.e., SDEs such that the diffusion coefficient σ does not depend on the state variable x.

We can find numerical schemes with higher convergence rates than the Euler scheme, at the cost of simplicity of the scheme. A method to generate such schemes is to add higher order terms of the Itô-Taylor expansion to the approximation. Adding the second order term gives the Milstein scheme, introduced by Milstein [24], which still has weak order 1, but the strong order is improved to 1. Further improvements are possible, although the higher order schemes have less practical value as they tend to contain multiple stochastic integrals that are difficult to simulate.

An important aspect of weak convergence is to consider enlarging the function class $\mathcal{C}$ in the definition. A common choice of $\mathcal{C}$ is polynomials, possibly with some additional regularity properties. In applications, such as option pricing, we often encounter a situation where

we need to compute the quantity $\mathbb{E}g(X_T)$ with a function g that is irregular or exceeds polynomial growth. We would then like to know whether the weak error still converges to zero, and whether the rate γ is affected. In the case of the Euler scheme, much work has been done, e.g. in [30, 3, 17], to prove that the order $\gamma = 1$ is unchanged with quite mild assumptions on the function g, but on the other hand, stronger assumptions on the SDE. In [3] the result is proved for bounded and measurable functions g, under hypoellipticity assumption on the SDE. In [17] this is extended to measurable functions with exponential growth, with ellipticity assumption on the SDE.

Many of these results as well as a comprehensive discussion about approximation of SDEs can be found in [22]. See also [21] and [27] for a more compact presentation and a survey of the development of the field.

5. Monte Carlo Methods for SDEs

We can approximate the expected value of the payoff, $\mathbb{E}g(X_T)$, using the classical Monte Carlo method, i.e. we define an estimator

$$\widehat{Y}_n = \frac{1}{n} \sum_{i=1}^{n} g(X_T^E(i)),$$

where we have n independent samples $X_T^E(1), \ldots, X_T^E(n)$ of X_T^E. Then the error of the method splits into two parts,

$$\mathbb{E}g(X_T) - \widehat{Y}_n = \mathbb{E}[g(X_T) - g(X_T^E)] + \mathbb{E}g(X_T^E) - \widehat{Y}_n.$$

The first part is the discretization error caused by the Euler scheme, and can be estimated using the weak convergence results. The second part is the statistical error of the Monte Carlo method, and it converges to zero by the strong law of large numbers. A common error measure to use, combining both sources of error above, is the mean square error

$$MSE(\widehat{Y}_n) = \mathbb{E}(\mathbb{E}g(X_T) - \widehat{Y}_n)^2.$$

To measure the efficiency of an algorithm, we use the notion of computational complexity, $C(\widehat{Y}_n)$, which is the number of units of computer time the algorithm $\widehat{Y}_n$ needs to achieve a given precision. We may consider one unit of computer time to be the time needed to complete a certain computational operation, e.g. simulation of an increment of the Brownian motion. This time depends on the system resources available. However, the computational complexity expressed in abstract time units gives a worst-case estimate that can be applied in any given system. For the classical Monte Carlo estimator, Duffie and Glynn [8] have showed that, if we require $MSE(\widehat{Y}_n) < \varepsilon^2$, then there exists $c > 0$ such that the computational complexity of $\widehat{Y}_n$ satisfies $C(\widehat{Y}_n) \leq c\varepsilon^{-3}$.

The efficiency of the classical Monte Carlo method can be significantly improved using the multilevel Monte Carlo method. The idea

is due to Heinrich [18], who approximates parameter dependent integrals in high dimensions. In 2008, Giles [14] introduced the multilevel method in option pricing to compute the expected value of the payoff more efficiently. Take $M \geq 2$ and $L \geq 0$, and let $h_l = T/M^l$, $0 \leq l \leq L$, be a sequence of timesteps. We denote by X_T^{E,h_l} the Euler scheme related to the partition of the interval $[0, T]$ using the timestep h_l. The classical method is to write the Monte Carlo estimator of $\mathbb{E}g(X_T^{E,h_L})$. Instead, we write the telescoping sum

$$\mathbb{E}g(X_T^{E,h_L}) = \mathbb{E}g(X_T^{E,h_0}) + \sum_{l=1}^{L} \mathbb{E}[g(X_T^{E,h_l}) - g(X_T^{E,h_{l-1}})].$$

We estimate $\mathbb{E}g(X_T^{E,h_0})$ with a Monte Carlo estimator $\widehat{Y}_0$ with N_0 independent samples, i.e.

$$\widehat{Y}_0 = \frac{1}{N_0} \sum_{i=1}^{N_0} g(X_T^{E,h_0}(i)), \tag{5.1}$$

and we estimate each of the summands $\mathbb{E}[g(X_T^{E,h_l}) - g(X_T^{E,h_{l-1}})]$ with a Monte Carlo estimator $\widehat{Y}_l$ with N_l independent samples of a Brownian motion path, i.e.

$$\widehat{Y}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} [g(X_T^{E,h_l}(i)) - g(X_T^{E,h_{l-1}}(i))]. \tag{5.2}$$

In each simulation in $\widehat{Y}_l$, we use the simulated Brownian motion path with step size h_l to compute the path with step size h_{l-1} by summing up the additional increments of the finer partition. This provides simulations of both X_T^{E,h_l} and $X_T^{E,h_{l-1}}$ from the same simulation of the Brownian motion path. By construction, the estimators $\widehat{Y}_l$ are independent. Then we approximate $\mathbb{E}g(X_T)$ by the combined estimator

$$\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l$$

and get the following, slightly more general, result:

Theorem 1 ([14, Theorem 3.1]). *If there exist independent estimators $\widehat{Y}_l$ based on N_l Monte Carlo samples, and positive constants $\alpha \geq 1/2$, β, c_1, c_2, c_3 such that*

(i) $\left| \mathbb{E}[g(X_T) - g(X_T^{E,h_l})] \right| \leq c_1 h_l^\alpha,$

(ii) $\mathbb{E}\widehat{Y}_l = \begin{cases} \mathbb{E}g(X_T^{E,h_0}), & l = 0, \\ \mathbb{E}[g(X_T^{E,h_l}) - g(X_T^{E,h_{l-1}})], & l > 0, \end{cases}$

(iii) $Var(\widehat{Y_l}) \leq c_2 N_l^{-1} h_l^\beta$, and

(iv) $C(\widehat{Y_l}) \leq c_3 N_l h_l^{-1}$,

then there exists a positive constant c_4 such that for any $\varepsilon < e^{-1}$, there are values L and N_l for which the multilevel estimator $\widehat{Y}$ satisfies

$$MSE(\widehat{Y}) \leq \varepsilon^2$$

with computational complexity

$$C(\widehat{Y}) \leq \begin{cases} c_4 \varepsilon^{-2}, & \beta > 1, \\ c_4 \varepsilon^{-2}(\log \varepsilon)^2, & \beta = 1, \\ c_4 \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

If the function g is Lipschitz and the coefficients σ and b of the SDE (2.1) satisfy certain regularity properties, then using the classical convergence results of the Euler scheme it is easy to show that the estimators (5.1) and (5.2) satisfy the assumptions of Theorem 1 with $\alpha = 1$ and $\beta = 1$. For a non-Lipschitz g, we can still use weak convergence results to determine the parameter α, and get that $\alpha = 1$ for a large class of functions as discussed in Section 4. However, the condition (iii) for the parameter β reduces to

$$Var\left(g(X_T^{E,h_l}) - g(X_T^{E,h_{l-1}})\right)$$
$$\leq \left(\sqrt{Var\left(g(X_T^{E,h_l}) - g(X_T^E)\right)} + \sqrt{Var\left(g(X_T^{E,h_{l-1}}) - g(X_T^E)\right)}\right)^2$$
$$\leq \left(\sqrt{\mathbb{E}|g(X_T^{E,h_l}) - g(X_T^E)|^2} + \sqrt{\mathbb{E}|g(X_T^{E,h_{l-1}}) - g(X_T^E)|^2}\right)^2.$$

Therefore we need an estimate for the error $\mathbb{E}|g(X_T^E) - g(X_T^{E,h_l})|^2$, a question to which this thesis provides answers.

For further reading about Monte Carlo methods for SDEs, see [15]. The survey paper [25] offers an overview of results concerning the multilevel method.

6. BACKWARD STOCHASTIC DIFFERENTIAL EQUATIONS

The theory of backward stochastic differential equations (BSDEs) originates in the stochastic optimal control theory from the 70's [5, 19]. If X is a solution of the forward equation (2.1), then the backward equation is

$$Y_t = g(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)\, ds - \int_t^T Z_s\, dW_s,\ 0 \leq t \leq T.$$

A solution of a BSDE is a pair $(Y_t, Z_t)_{t \in [0,T]}$ of adapted processes satisfying the backward equation, see [26]. BSDEs are extensively studied

for example in stochastic finance [9], and there is a need of numerical schemes to solve BSDEs and of theoretical investigations of discretization schemes and their convergence rates.

The discretization error splits into the error of the forward and the backward approximation. The error of the forward approximation is of the form $\|g(X_T) - g(X_T^\pi)\|_p$, where X^π is a forward approximation of the diffusion X, e.g. the Euler scheme. The function g is usually assumed to be Lipschitz, see [6, 31] and the references therein. The results of this thesis open a way to consider functions g, which are not Lipschitz.

Moreover, inequalities involving $\left\|g(X_T) - g(\tilde{X}_T)\right\|_p$ are used in [11] to determine the L_p-variation of the solution of the BSDE. The L_p-variation is mainly responsible for the convergence properties of the approximation schemes for the backward component. Results in this direction are also presented in [16].

7. Relation to fractional smoothness

The approximation problem considered in this thesis is related to the fractional smoothness of the indicator function $\chi_{[K,\infty)}$, with $K \in \mathbb{R}$. Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then we can define the fractional smoothness, with respect to $2 \leq p < \infty$ and a given class of random variables $\{X, \hat{X} : \Omega \to \mathbb{R}\}$, to be the largest possible power $\beta_p > 0$ in the inequality

$$\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| \leq C_p \left\|X - \hat{X}\right\|_p^{\beta_p}. \qquad (7.1)$$

For example, for the class of Gaussian random variables we have the following:

Theorem 2 ([13, p. 12]). *Suppose that $X, \hat{X} \sim N(0,1)$ and $(X, \hat{X})$ is a Gaussian random vector. Then, for all $K \in \mathbb{R}$ and $p \geq 2$, we have*

$$\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| \leq C_p \left\|X - \hat{X}\right\|_p.$$

The statement of Theorem 2 is equivalent to the knowledge of the fractional smoothness of the indicator function in terms of Malliavin Besov spaces by [12]. However, if we generalize this by adjusting the class $\{X, \hat{X}\}$ of random variables, we come up with a new concept of fractional smoothness that leads us outside the conventional Malliavin Besov setting.

If we were to choose the random variables X and $\hat{X}$ freely, we could take e.g. $X \equiv K$ and $\hat{X} = X - \varepsilon$ for $\varepsilon > 0$. Then, as $\varepsilon \to 0$, the right hand side of Equation (7.1) would converge, whereas the left hand side would not. Hence we need to restrict the possible choices of X and $\hat{X}$ to avoid this. A minimal assumption is to say that one of the random

variables, say X, has a bounded density with respect to the Lebesgue measure. This gives optimal results in the following sense:

Theorem 3 ([1, Lemma 3.4, Theorem 2.4 (ii) and (iii)]).

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

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

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

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

for all random variables X and $\hat{X}$ such that X has a bounded density, 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*

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

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

By Theorem 3 we have defined a new weaker notion of fractional smoothness in the sense of Equation (7.1). Choosing the class $\{X, \widehat{X} : X \text{ has a bounded density}\}$ results in the optimal power $\beta_p = p/(p+1)$, which is strictly smaller than the power $\beta_p = 1$ in the Gaussian case.

8. RESULTS

8.1. **General case.** Let us first consider the setting of Section 7. We assume throughout this section that X and $\hat{X}$ are random variables such that X has a bounded density f_X, and for some $1 \leq p < \infty$, $C(p, X) > 0$, and $\beta_p > 0$, they satisfy

$$\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| \leq C(p, X) \left\| X - \hat{X} \right\|_p^{\beta_p} \tag{8.1}$$

for all $K \in \mathbb{R}$. This is justified by Theorem 3, which gives us $\beta_p = p/(p + 1)$ if no further assumptions are made. We denote the class of functions of bounded variation on the real line by BV, defined in detail in [1, Definition 2.1], and the variation of $g \in BV$ by $V(g)$. A basic example is the payoff function of the binary option, $g = \chi_{[K,\infty)}$, which has $V(g) = 1$. The information given in Equation (8.1) implies an extension to functions of bounded variation, such that the rate β_p remains unchanged:

Theorem 4 ([1, Theorem 2.4 (i) with general β_p]). *If $g \in BV$ and $1 \le q < \infty$, then*

$$\left\| g(X) - g(\hat{X}) \right\|_q^q \le 3^q V(g)^q C(p, X) \left\| X - \hat{X} \right\|_p^{\beta_p}.$$

We can further extend this by allowing the function g to have unbounded variation. This is based on the idea of compensating for the variation of g by the tail probabilities of X and $\hat{X}$.

Let μ be a set function defined on bounded Borel sets, such that the restriction of μ to a compact set is a signed measure. We associate it with the distribution function

$$g^\mu(x) = \begin{cases} \mu([0, x)), & \text{for } x > 0, \\ -\mu([x, 0)), & \text{for } x \le 0. \end{cases}$$

Then we can show that μ possesses a unique Jordan decomposition $\mu = \mu^+ - \mu^-$ on the real line [2, Theorem 3.3], and therefore has a unique total variation measure $|\mu| = \mu^+ + \mu^-$.

Let us call a continuous function $\varphi : \mathbb{R} \to \mathbb{R}$ a bump function if $0 < \varphi(x) \le 1$ for all $x \in \mathbb{R}$, $\varphi(0) = 1$, and φ is increasing in $(-\infty, 0]$ and decreasing in $[0, \infty)$. Given a bump function φ, we define BV_φ to be the space of all functions g^μ such that

$$\|g^\mu\|_\varphi = \int_{\mathbb{R}} \varphi(x) \, d|\mu|(x) < \infty.$$

We define the space BV_φ rigorously and show that it is a Banach space in [2, Section 3].

We can also characterize the space BV_φ in a more intuitive and, in some cases, more practical way. For a bump function φ, we consider the class of all functions

$$g(x) = \begin{cases} \int_{[0,x)} \frac{1}{\varphi} \, d\nu, & \text{for } x > 0, \\ -\int_{[x,0)} \frac{1}{\varphi} \, d\nu, & \text{for } x \le 0, \end{cases}$$

where ν is a signed measure and $[0, 0) = \emptyset$. Then this class is identical to BV_φ by [2, Theorem 5.1].

The idea is to exploit the decay of the function φ to compensate for the variation of g. Let us take a bump function $\varphi^{X,\hat{X}}$ that gives an upper bound for the tail probabilities of X and $\hat{X}$, i.e.

$$\mathbb{P}(X \ge K) \vee \mathbb{P}(\hat{X} \ge K) \le \varphi^{X,\hat{X}}(K) \text{ for } K > 0,$$

and

$$\mathbb{P}(X \le K) \vee \mathbb{P}(\hat{X} \le K) \le \varphi^{X,\hat{X}}(K) \text{ for } K \le 0.$$

Here $a \vee b = \max\{a, b\}$. Then we have the following convergence result:

Theorem 5 ([2, Theorem 6.2]). *Let $1 \leq p \leq \infty$, $0 < \theta < 1$, and consider the bump function $\varphi^{X,\hat{X}}$. If $1 \leq q < \infty$ and $g^\mu \in BV_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}}$, then*

$$\left\| g^\mu(X) - g^\mu(\hat{X}) \right\|_q^q \leq 2^\theta \|g^\mu\|_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}}^q \, C(p,X)^{1-\theta} \left\| X - \hat{X} \right\|_p^{(1-\theta)\beta_p} .$$

This is a natural extension of Theorem 4, which can intuitively be considered as the case $\theta = 0$ in Theorem 5. Namely, if we set $\varphi \equiv 1$, then the condition $g \in BV_\varphi$ implies $g \in BV$. The faster the decay of the function φ is, the larger is the space BV_φ. We show in [2, Section 9] that the boundedness of X and $\hat{X}$ in L_p for all $1 \leq p < \infty$ imply that φ decays faster than any polynomial.

8.2. **Application to SDEs and the Euler scheme.** Our results for the random variables X and $\hat{X}$ can be directly applied to the approximation of solutions of SDEs. Suppose that $X = X_T$, a solution of the SDE (2.1) at the endpoint T, and $\hat{X}$ is a numerical approximation of X_T. In particular we are interested in the Euler scheme, $\hat{X} = X_T^E$, which converges in L_p with strong order $1/2$, i.e., it satisfies

$$\left\| X_T - X_T^E \right\|_p \leq C_p |\pi|^{1/2} \tag{8.2}$$

with $C_p \leq e^{Mp}$, see [2, Lemma 11.2].

An essential assumption in our results is that X_T has a bounded density. Apparently, no equivalent condition between the boundedness of the density of X_T and the properties of the coefficients σ and b of the SDE (2.1) is known. Sufficient conditions, however, are available. One is to assume that $\sigma, b \in C_b^\infty([0,T] \times \mathbb{R})$ and the SDE is uniformly elliptic [10, p. 263], but also a weaker condition concerning regularity and integrability of the coefficients exists [7, Theorem 2]. See also [1, Remark 4.1].

For X_T with a bounded density and the Euler scheme, we get the following convergence result using Theorems 3 and 4, the knowledge about the upper bound for the constant C_p, and an optimization argument over the powers p:

Theorem 6 ([1, Theorem 5.4]). *Let $1 \leq q < \infty$ and $g \in BV$. Then there exist $M > 0$ and $m \in (0,1)$ such that for $|\pi| < m$ we have*

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

Remark 7. In [1] we used an upper bound $C_p \leq e^{Mp^2}$ for the constant in Equation (8.2). The dropping of the square, which is due to the optimal constant in the Burkholder–Davis–Gundy inequality shown in [4], gives a minor improvement to the power of the logarithm in Theorem 6.

We can extend the result of Theorem 6 to a larger class of functions using the approach described in the previous section. With the additional information given by the special case $X = X_T$ and $\hat{X} = X_T^E$, we get an explicit estimate for the decay of the bump function φ^{X_T, X_T^E}, which controls the tail probabilities of X_T and X_T^E.

Theorem 8 ([2, Theorem 11.4]). *We can choose the function φ^{X_T, X_T^E} in a way that $\varphi^{X_T, X_T^E} \leq \varphi_E^{X_T}$, where $\varphi_E^{X_T}$ is a function such that*

(i) *if the functions σ and b are bounded, i.e. $|\sigma|, |b| < M$, we have for $z_0 = |x_0| + MT$ that*

$$\varphi_E^{X_T}(z) = \begin{cases} e^{-\frac{(|z|-z_0)^2}{2M^2T}} & \text{if } |z| > z_0, \\ 1 & \text{if } |z| \leq z_0. \end{cases}$$

(ii) *if the functions σ and b are Lipschitz, then there exist constants $M > 0$ and $z_0 > 0$ such that*

$$\varphi_E^{X_T}(z) = \begin{cases} |z|^{-M \log(1+|z|-z_0)} & \text{if } |z| > z_0, \\ 1 & \text{if } |z| \leq z_0. \end{cases}$$

Now we can apply Theorems 5 and 8 to get the following convergence result:

Theorem 9 ([2, Corollary 10.2 for the Euler scheme]). *Let $1 \leq q < \infty$ and $0 < \varepsilon < 1/2$. Then for $\theta = \frac{\varepsilon}{1-\varepsilon}$, the function $\varphi_E^{X_T}$, and $g^\mu \in BV_{\left(\varphi_E^{X_T}\right)^{\frac{\theta}{q}}}$ there exists $M > 0$ such that*

$$\left\| g^\mu(X_T) - g^\mu(X_T^E) \right\|_q^q \leq 3 \left(e^{\frac{M}{\theta}} \sup f_{X_T} \right)^{1-2\varepsilon} \| g^\mu \|_{\left(\varphi_E^{X_T}\right)^{\frac{\theta}{q}}}^q |\pi|^{\frac{1}{2}-\varepsilon}.$$

If the SDE (2.1) has bounded coefficients, then we can use Theorem 8 to achieve a similar rate as in Theorem 6, improved according to Remark 7, for functions of polynomial variation:

Theorem 10 ([2, Theorem 11.6]). *Suppose that the coefficients σ and b of the SDE (2.1) are bounded. Let $g : \mathbb{R} \to \mathbb{R}$ be a function with a representation $g = g^\mu$ such that there exists $s \in \{0, 1, 2, \dots\}$ with*

$$\int_{\mathbb{R}} \varphi \, d|\mu| \leq \int_{\mathbb{R}} \varphi(x) |x|^s \, dx$$

for all bump functions φ. Then for any $1 \leq q < \infty$ there exist $M > 0$ and $m \in (0, 1)$ such that

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

for $|\pi| < m$.

In Theorems 6 and 10 we achieve a convergence rate that is asymptotically $1/2$ as the mesh size decreases. Similarly in Theorem 9 we get a rate arbitrarily close to $1/2$. While it is not clear whether the rate could be equal to $1/2$, we can provide an example showing that rates better than $1/2$ are impossible:

Theorem 11 ([1, Theorem 7.2]). *Let S be the geometric Brownian motion and $T = 1$. Then there exists $K_0 > 0$ such that*

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

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

8.3. **Application to the multilevel Monte Carlo method.** The results of Section 8.2 provide answers to the computation of the variance parameter β in the assumption (iii) of Theorem 1. For functions g of bounded variation, the parameter β is asymptotically $1/2$ by Theorem 6, and this implies the following result for the multilevel Monte Carlo method and the Euler scheme.

Theorem 12 ([1, Theorem 6.1]). *Let $g \in BV$ and $\alpha \geq 1/2$. Suppose that the assumptions (i),(ii), and (iv) of Theorem 1 hold, and $0 < \varepsilon < \min\{\sqrt{2}c_1, 1/e\} =: \varepsilon_0$, with $c_1 > 0$ taken from assumption (i). Then there exists $\tilde{c}_4 > 0$ such that the computational complexity of the multilevel estimator $\widehat{Y}$ is given by*

$$C(\widehat{Y}) \leq \tilde{c}_4 \varepsilon^{-2 - \frac{1}{2\alpha}},$$

and the mean square error of $\widehat{Y}$ has a bound

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

where the function $\Phi : (0, \varepsilon_0) \to (0, \infty)$ satisfies, for all $\delta > 0$,

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

The statement of Theorem 12 has a straightforward extension to the function classes introduced in Section 8.1. This is formulated in [2, Section 12].

9. Conclusion

We found optimal powers $\beta_p = p/(p+1)$ in the estimate (8.1) for the indicator functions, with general random variables X and $\hat{X}$ such that X has a bounded density. Then we generalized the estimate in two steps: first for functions of bounded variation, and then for a new space of functions with unbounded variation compensated by bump functions.

We applied the results to the solution of the one-dimensional SDE (2.1) and its Euler approximation. We found that the Euler approximation converges to the solution in L_p with a strong order arbitrarily close to $1/(2p)$.

We used the results for SDEs to show complexity estimates for the multilevel Monte Carlo method using the Euler scheme. We showed that for functions of bounded variation, the complexity of the multilevel estimator is of order $\varepsilon^{-2-\frac{1}{2\alpha}}$, whereas the mean square error satisfies $MSE \leq \varepsilon^{2-\delta(\varepsilon)}$ with $\delta(\varepsilon) \searrow 0$ as $\varepsilon \to 0$ and $0 < \varepsilon < \varepsilon_1$, i.e. the MSE is asymptotically ε^2. In particular this holds for the payoff function of the binary option.

References

[1] Avikainen, R. On irregular functionals of SDEs and the Euler scheme. *Finance Stoch. 13*, 3 (2009), 381–401.

[2] Avikainen, R. On generalized bounded variation and approximation of SDEs. *Preprint 383, Department of Mathematics and Statistics, University of Jyväskylä* (2009).

[3] Bally, V., and Talay, D. The law of the Euler scheme for stochastic differential equations: I. convergence rate of the distribution function. *Probab. Theory Related Fields 104*, 1 (1996), 43–60.

[4] Barlow, M. T., and Yor, M. Semi–martingale inequalities via the Garsia–Rodemich–Rumsey lemma, and applications to local times. *J. Funct. Anal. 49*, 2 (1982), 198–229.

[5] Bismut, J.-M. Conjugate convex functions in optimal stochastic control. *J. Math. Anal. Appl. 44* (1973).

[6] Bouchard, B., and Touzi, N. Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stoch. Proc. Appl. 111* (2004), 175–206.

[7] Caballero, M. E., Fernández, B., and Nualart, D. Estimation of densities and applications. *J. Theoret. Probab. 11*, 3 (1998), 831–851.

[8] Duffie, D., and Glynn, P. Efficient Monte Carlo simulation of security prices. *Ann. Appl. Probab. 5*, 4 (1995), 897–905.

[9] El Karoui, N., Peng, S., and Quenez, M. C. Backward stochastic differential equations in finance. *Math. Finance 7*, 1 (1997), 1–71.

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

[11] Geiss, C., Geiss, S., and Gobet, E. Generalized fractional smoothness and L_p-variation of BSDEs with non-Lipschitz terminal condition. *In preparation* (2009).

[12] Geiss, S., and Hujo, M. Interpolation and approximation in $L_2(\gamma)$. *J. Appr. Theory 144* (2007), 213–232.

[13] Geiss, S., and Toivola, A. Weak convergence of error processes in discretizations of stochastic integrals and Besov spaces. *To appear in Bernoulli*.

[14] Giles, M. B. Multilevel Monte Carlo path simulation. *Oper. Res. 56*, 3 (2008), 607–617.

[15] Glasserman, P. *Monte Carlo Methods in Financial Engineering*. Springer, 2004.

[16] Gobet, E., and Makhlouf, A. L_2-time regularity of BSDEs with irregular terminal functions. *Preprint* (2008).

18

[17] GUYON, J. Euler scheme and tempered distributions. *Stochastic Process. Appl. 116*, 6 (2006), 877–904.

[18] HEINRICH, S. Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing, Third International Conference. Lecture Notes in Computer Science, vol. 2179*, S. Margenov, J. Wasniewski, and P. Y. Yalamov, Eds. Springer, Berlin, 2001, pp. 58–67.

[19] KABANOV, Y. On the Pontryagin maximum principle for linear stochastic differential equations. In *Probabilistic Models and Control of Economical Processes, CEMI*. 1978.

[20] KARATZAS, I., AND SHREVE, S. E. *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, Berlin, 1991.

[21] KLOEDEN, P. E. The systematic derivation of higher order numerical schemes for stochastic differential equations. *Milan J. Math. 70* (2002), 187–207.

[22] KLOEDEN, P. E., AND PLATEN, E. *Numerical Solutions of Stochastic Differential Equations*. Springer, Berlin, 1992.

[23] MARUYAMA, G. Markov processes and stochastic equations. *Nat. Sci. Rep. Ochanomizu Univ. 4* (1953), 40–43.

[24] MILSTEIN, G. N. Approximate integration of stochastic differential equations. *Teor. Verojatnost. i Primenen 19* (1974).

[25] MÜLLER-GRONBACH, T., AND RITTER, K. Variable subspace sampling and multi-level algorithms. In *Monte Carlo and Quasi-Monte Carlo Methods 2008*, P. L'Ecuyer and A. B. Owen, Eds. Springer-Verlag, Berlin, to appear.

[26] PARDOUX, É., AND PENG, S. Adapted solution of a backward stochastic differential equation. *Systems Control Lett. 14*, 1 (1990), 55–61.

[27] PLATEN, E. An introduction to numerical methods for stochastic differential equations. *Acta Numer. 8* (1999), 197–246.

[28] PROTTER, P. E. *Stochastic Integration and Differential Equations*. Springer-Verlag, 2005.

[29] REVUZ, D., AND YOR, M. *Continuous Martingales and Brownian Motion*. Springer, 1999.

[30] TALAY, D., AND TUBARO, L. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl. 8*, 4 (1990), 483–509.

[31] ZHANG, J. *Some fine properties of backward stochastic differential equations*. PhD thesis, Purdue University, 2001.

84. ONNINEN, JANI, Mappings of finite distortion: Continuity. (24 pp.) 2002
85. OLLILA, ESA, Sign and rank covariance matrices with applications to multivariate analysis. (42 pp.) 2002
86. KAUKO, VIRPI, Visible and nonexistent trees of Mandelbrot sets. (26 pp.) 2003
87. LLORENTE, JOSÉ G., Discrete martingales and applications to analysis. (40 pp.) 2002
88. MITSIS, THEMIS, Topics in harmonic analysis. (52 pp.) 2003
89. KÄRKKÄINEN, SALME, Orientation analysis of stochastic fibre systems with an application to paper research. (53 pp.) 2003
90. HEINONEN, JUHA, Geometric embeddings of metric spaces. (44 pp.) 2003
91. RAJALA, KAI, Mappings of finite distortion: Removable singularities. (23 pp.) 2003
92. FUTURE TRENDS IN GEOMETRIC FUNCTION THEORY. RNC WORKSHOP JYVÄSKYLÄ 2003. Edited by D. Herron. (262 pp.) 2003
93. KÄENMÄKI, ANTTI, Iterated function systems: Natural measure and local structure. (14 pp.) 2003
94. TASKINEN, SARA, On nonparametric tests of independence and robust canonical correlation analysis. (44 pp.) 2003
95. KOKKI, ESA, Spatial small area analyses of disease risk around sources of environmental pollution: Modelling tools for a system using high resolution register data. (72 pp.) 2004
96. HITCZENKO, PAWEŁ, Probabillistic analysis of sorting algorithms. (71 pp.) 2004
97. NIEMINEN, TOMI, Growth of the quasihyperbolic metric and size of the boundary. (16 pp.) 2005
98. HAHLOMAA, IMMO, Menger curvature and Lipschitz parametrizations in metric spaces. (8 pp.) 2005
99. MOLTCHANOVA, ELENA, Application of Bayesian spatial methods in health and population studies using registry data. (55 pp.) 2005
100. HEINONEN, JUHA, Lectures on Lipschitz analysis. (77 pp.) 2005
101. HUJO, MIKA, On the approximation of stochastic integrals. (19 pp.) 2005
102. LINDQVIST, PETER, Notes on the p-Laplace equation. (80 pp.) 2006
103. HUKKANEN, TONI, Renormalized solutions on quasi open sets with nonhomogeneous boundary values. (41 pp.) 2006
104. HÄHKIÖNIEMI, MARKUS, The role of representations in learning the derivative. (101 pp.) 2006
105. HEIKKINEN, TONI, Self-improving properties of generalized Orlicz–Poincaré inequalities. (15 pp.) 2006
106. TOLONEN, TAPANI, On different ways of constructing relevant invariant measures. (13 pp.) 2007
107. HORPPU, ISMO, Analysis and evaluation of cell imputation. (248 pp.) 2008
108. SIRKIÄ, SEIJA, Spatial sign and rank based scatter matrices with applications. (29 pp.) 2007
109. LEIKAS, MIKA, Projected measures on manifolds and incomplete projection families. (16 pp.) 2007
110. TAKKINEN, JUHANI, Mappings of finite distortion: Formation of cusps. (10 pp.) 2007
111. TOLVANEN, ASKO, Latent growth mixture modeling: A simulation study. (201 pp.) 2007
112. VARPANEN, HARRI, Gradient estimates and a failure of the mean value principle for p-harmonic functions. (66 pp.) 2008
113. MÄKÄLÄINEN, TERO, Nonlinear potential theory on metric spaces. (16 pp.) 2008
114. LUIRO, HANNES, Regularity properties of maximal operators. (11 pp.) 2008
115. VIHOLAINEN, ANTTI, Prospective mathematics teachers' informal and formal reasoning about the concepts of derivative and differentiability. (86 pp.) 2008
116. LEHRBÄCK, JUHA, Weighted Hardy inequalities and the boundary size. (21 pp.) 2008
117. NISSINEN, KARI, Small area estimation with linear mixed models from unit-level panel and rotating panel data. (230 pp.) 2009
118. BOJARSKI, B.V., Generalized solutions of a system of differential equations of the first order and elliptic type with discontinuous coefficients. (64 pp.) 2009
119. RAJALA, TAPIO, Porosity and dimension of sets and measures. (22 pp.) 2009
120. MYLLYMÄKI, MARI, Statistical models and inference for spatial point patterns with intensity-dependent marks. (115 pp.) 2009