ON GENERALIZED BOUNDED VARIATION AND APPROXIMATION OF SDES

Rainer Avikainen

Preprint 383 September 2009

2000 *Mathematics Subject Classification.* 60H10, 41A25, 26A45, 65C20, 65C30.

ISSN 1457-9235

ON GENERALIZED BOUNDED VARIATION AND APPROXIMATION OF SDES

Rainer Avikainen

September 2009

2000 *Mathematics Subject Classification.* 60H10, 41A25, 26A45, 65C20, 65C30.

ON GENERALIZED BOUNDED VARIATION AND APPROXIMATION OF SDES

RAINER AVIKAINEN

Abstract. We consider upper bounds 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. We extend the results of [3], where g was a function of bounded variation, to a class of generalized bounded variation containing functions of polynomial variation. This is obtained by compensating for the variation by the tail of the distribution of X and $\hat{X}$. 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 in this particular case, exponential variation of the function g is also allowed. An application to the multilevel Monte Carlo method is considered.

1. Introduction

1.1. **Background.** 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. In [3] we computed bounds for the error $\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})|$, where $K \in \mathbb{R}$, by reducing it to the error $\left\|X - \hat{X}\right\|_p$. This gave us a tool to compute convergence rates $\gamma > 0$ for irregular functionals of stochastic differential equations, i.e. in

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

where $1 \le p < \infty$, $g \in BV$, X_T is a diffusion, and X_T^π is an approximation of X_T corresponding to a partition π of the interval $[0,T]$.

Inequalities of the type (1.1) play an important role in two fields of financial mathematics. It is an integral part of the multilevel Monte Carlo method for SDEs, developed by M. B. Giles [11, 12], to approximate the expected payoff of an option with a significant improvement in the computational complexity of the problem. The inequality (1.1) is required to determine the complexity of the algorithm for options with a non-Lipschitz payoff, as the complexity is expressed in terms

2000 *Mathematics Subject Classification.* 60H10, 41A25, 26A45, 65C20, 65C30.

Key words and phrases. Generalized bounded variation, stochastic differential equations, approximation, Euler scheme, multilevel Monte Carlo method.

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.

of the convergence parameter γ. This motivates us to look for optimal values of γ in (1.1). Another application is the L_p-variation of backward stochastic differential equations with non-Lipschitz terminal condition, due to C. Geiss, S. Geiss and E. Gobet [8]. When studying fractional smoothness for BSDEs, convergence rates of certain conditional expectations are needed. In this context the inequality (1.1) again appears.

Our goal is to extend the convergence results of type (1.1) in [3] to a larger class of functions, namely functions that are unbounded or have infinite variation. The starting point is the result in [3], where we showed that if X has a bounded density f_X, then

$$\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| \leq 3\,(\sup f_X)^{\frac{p}{p+1}} \left\| X - \hat{X} \right\|_p^{\frac{p}{p+1}} \tag{1.2}$$

for all $K \in \mathbb{R}$ and all $0 < p < \infty$. We also proved optimality of the power of the L_p-norm on the right hand side of (1.2), i.e. that the power $p/(p+1)$ is the largest possible power in general. However, additional information about the distribution of X and $\hat{X}$ enables us to show estimates better than (1.2). An example is the following result in [10, Discussion after Proposition 3.5]:

Theorem 1.1. *Suppose that $X, \hat{X} \sim N(0,1)$ and $(X, \hat{X})$ is a Gaussian random vector, and let $p \geq 2$. Then*

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

for all $K \in \mathbb{R}$.

Therefore it is natural to take as an assumption the statement

$$\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| \leq C(p, X) \left\| X - \hat{X} \right\|_p^{\beta_p}, \tag{1.3}$$

where the exponent is given by an unspecified function $\beta_p : [1,\infty) \to (0,\infty)$, and to show results of the type (1.1) for as large class of functions as possible.

The problem of determining the exponent β_p is related to fractional smoothness. The statement of Theorem 1.1 is equivalent to knowing the fractional smoothness of the indicator function in terms of Malliavin Besov spaces [9]. Taking another class of test random variables $\{X, \hat{X}\}$ would result in a changed notion of fractional smoothness, and a different power of the L_p-norm on the right hand side of (1.3). The inequality (1.2) with the power $p/(p+1)$ gives the most general situation enabling us to take any random variables X and $\hat{X}$ such that X has a bounded density. The boundedness assumption of the density of X is essential, because without it the statement of Equation (1.3) would contradict itself, unless we replaced the indicator function by a Lipschitz function. Namely, with the choice $X \equiv K$ and $\hat{X} \equiv K - \varepsilon$

for $\varepsilon > 0$, the left hand side of Equation (1.3) does not converge at all as $\varepsilon \to 0$, but $\hat{X}$ converges to X in L_p.

1.2. **Results.** We develop an extension of bounded variation by compensating for the variation of functions by the tail of the distribution of X and $\hat{X}$. If the tail behavior is described by a function $\varphi : \mathbb{R} \to \mathbb{R}$, which vanishes at $\pm\infty$, then we define a space of functions of φ-bounded variation, to be called BV_φ. The space BV_φ is a Banach space, and its size depends on the decay of the function φ, i.e. faster decay of φ allows more compensation and thus more variation for functions in BV_φ. A function in BV_φ can be represented as an integral of $1/\varphi$ with respect to a signed measure, and is by definition left-continuous. The latter condition can be relaxed by adding jumps, also compensated by φ. This is analogous to the spaces BV and NBV in [17]. If $X, \hat{X} \in L_p$ and $\left\| X - \hat{X} \right\|_p \le C_p < \infty$ for all $1 \le p < \infty$, then we show in Lemma 9.1 that the function φ decays faster than any polynomial, allowing polynomial variation for the functions in BV_φ.

Given a function $g \in BV_\varphi$, and assuming the condition (1.3), we show in Theorem 6.2 that, if $1 \le q < \infty$ and $0 < \theta < 1$, then

$$\left\| g(X) - g(\hat{X}) \right\|_q^q \le C(\theta, q, g, \varphi) C(p, X)^{1-\theta} \left\| X - \hat{X} \right\|_p^{(1-\theta)\beta_p} \tag{1.4}$$

for every $1 \le p < \infty$. This is a natural extension of [3, Theorem 2.4], where we showed an analogous result with $g \in BV$ and $\theta = 0$. We also show that Equation (1.4) holds if the left-continuity assumption of functions in BV_φ is dropped.

By concavity arguments we can further extend the class of functions we can handle. Suppose that $0 < r \le 1$ and $g = |f|^r \operatorname{sgn} f$ for some $f \in BV$. In Theorem 7.1 we show that, if $0 < q < \infty$, then

$$\left\| g(X) - g(\widehat{X}) \right\|_q^q \le C(r, q, f) C(p, X) \left\| X - \widehat{X} \right\|_p^{\beta_p(rq \wedge 1)}$$

for every $1 \le p < \infty$. This is valid e.g. for functions with frequent variation on a finite interval, like the function in Example 7.3.

We have sharp convergence results on the space of Lipschitz functions and BV on a finite interval $[a, b]$, so we can apply the real interpolation method to get sharp convergence rates in the interpolation spaces $(Lip([a, b]), BV([a, b]))_{\theta,q}$ with parameters $0 < \theta < 1$ and $1 \le q \le \infty$. Theorem 8.4 shows that for g in one such space, and random variables X and $\widehat{X}$ with values in $[a, b]$,

$$\left\| g(X) - g(\hat{X}) \right\|_p \le C(p, X, \theta, q, g) \left\| X - \hat{X} \right\|_p^{1-\theta\left(1-\frac{1}{1+p}\right)}$$

for every $1 \le p < \infty$. The rate is optimal by Theorem 8.6.

All of the results above can be applied to the approximation of solutions of stochastic differential equations. Let X be a solution of an

SDE such that X_T has a bounded density. Given an approximation $(X_t^\pi)_{t\in[0,T]}$ of X with

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

then using (1.4), we show in Corollary 10.2 that for $1 \le q < \infty$ and $0 < \varepsilon < \gamma$,

$$\|g(X_T) - g(X_T^\pi)\|_q^q \le C(q, \gamma, \varepsilon, X)\, |\pi|^{\gamma-\varepsilon} \,, \tag{1.5}$$

whenever g is a function of polynomial variation. If X^π is the Euler scheme, then in Theorem 11.4 we find estimates for the decay of the function φ that are better than in the general case. This extends the result (1.5) to functions with variation higher than polynomial. If the coefficients of the SDE are bounded, then even exponential variation is possible.

We use our results on the Euler scheme to determine the variance parameter β in the multilevel Monte Carlo method of Giles [12]. By Corollary 12.1 we get $\beta = 1/2 - \varepsilon$ for any $\varepsilon > 0$, for functions g with variation related to the decay of φ given in Theorem 11.4. In the case that g has polynomial variation, Theorem 11.6 provides a logarithmic expression for ε, i.e., $\varepsilon = C/(-\log|\pi|)^{1/3}$, which converges to zero as $|\pi| \to 0$. In Corollary 12.2, we extend the variance estimate (iii') in [3, Section 6], shown for functions of bounded variation, to functions of polynomial variation. Consequently, the complexity result [3, Theorem 6.1] holds for functions of polynomial variation as well.

1.3. **Organization of the paper.** We start by recalling some preliminary definitions in Section 2. In Section 3 we define the space of functions of φ-bounded variation, BV_φ, and show that it is complete. By definition the functions in BV_φ are left-continuous, but in Section 4 we extend the class to different types of discontinuity using the idea of compensating jumps by the tail probabilities. We present an alternative characterization of the space BV_φ in terms of integrals of $1/\varphi$ with respect to signed measures in Section 5. The main convergence result of type (1.4) is presented in Section 6. In Section 7 we use a simple concavity trick to deal with certain functions for which the compensation idea of φ-bounded variation fails. Section 8 contains another extension to find sharp convergence rates for functions in the real interpolation spaces between Lipschitz and BV-functions defined on a finite interval. In Section 9 we consider a typical situation where the functions in BV_φ can have polynomial variation. The convergence result of Section 6 is applied to approximation of stochastic differential equations in Section 10. In Section 11 we recall the Euler scheme and use specific information about the scheme to get better convergence results. Finally, we apply the results concerning the Euler scheme to the multilevel Monte Carlo method in Section 12.

2. Preliminaries

Let us first recall a set of definitions, starting with $\mathbb{N} = \{1, 2, \dots\}$.

Definition 2.1. Let $(X, \mathcal{F})$ be a measurable space. A partition of $F \in \mathcal{F}$ is a countable collection $\{F_i\}_{i=1}^{\infty} \subset \mathcal{F}$ such that $F_i \cap F_j = \emptyset$ if $i \neq j$, and $\cup F_i = F$.

Definition 2.2 (Signed measure)**.** A signed measure μ on a measurable space $(X, \mathcal{F})$ is a set function $\mu : \mathcal{F} \to \mathbb{R}$ such that

$$\mu(F) = \sum_{i=1}^{\infty} \mu(F_i)$$

for all $F \in \mathcal{F}$ and all partitions $\{F_i\}_{i=1}^{\infty}$ of F.

Remark 2.3. A signed measure is always finite, i.e. $|\mu(X)| < \infty$.

Definition 2.4 (Total variation measure)**.** The total variation measure $|\mu|$ of a signed measure μ is the set function

$$|\mu| : \mathcal{F} \to [0, \infty), \ |\mu|(F) = \sup \sum_{i=1}^{\infty} |\mu(F_i)|,$$

where the supremum is taken over all partitions $\{F_i\}_{i=1}^{\infty}$ of F.

Remark 2.5. The total variation measure $|\mu|$ is always a finite positive measure, i.e. $|\mu|(X) < \infty$, see [18, Theorems 6.2 and 6.4]. Moreover, $|\mu(F)| \leq |\mu|(F)$ for all $F \in \mathcal{F}$.

Definition 2.6. A measure $\mu : X \to [0, \infty]$ on a measure space $(X, \mathcal{F})$ is σ-finite, if there exist a partition $\{F_i\}$ of X such that $\mu(F_i) < \infty$ for all $i \in \mathbb{N}$, i.e. the space X can be written as a countable union of measurable sets of finite measure.

We also recall the definition of functions of bounded variation:

Definition 2.7. Given a function $f : \mathbb{R} \to \mathbb{R}$, set

$$T_f(x) := \sup \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})|, \tag{2.1}$$

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 we call $V(f)$ the (total) variation of f.

Remark 2.8. We will occasionally use the fact that a left-continuous function $f \in BV$ has a unique representation $f(x) = c + \mu((-\infty, x))$, where $c \in \mathbb{R}$ and μ is a signed measure. Conversely, any signed measure μ defines a function $f(x) = \mu((-\infty, x)) \in BV$, which is left-continuous and vanishes as $x \to -\infty$. Moreover, $V(f) = |\mu|(\mathbb{R})$. See Theorems 8.13 and 8.14 in [17].

3. Generalization of Bounded Variation

Our aim in this section is to define functions g that may be of unbounded variation on the real line, but have a strong enough bound on the variation to enable us to show a result of the type

$$\left\| g(X) - g(\hat{X}) \right\|_q^q \leq C(p, q, g, X) \left\| X - \hat{X} \right\|_p^{\beta_p}, \; 1 \leq p, q < \infty,$$

which is proved for BV-functions in [3, Theorem 2.4] with $\beta_p = \frac{p}{p+1}$. We obtain the bound by compensating for the variation of g by a function φ vanishing at infinity, resulting in a notion of φ-bounded variation. Let us now define such functions rigorously, and show that the functions of φ-bounded variation generate a Banach space.

Definition 3.1 (Bump function)**.** Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $0 < \varphi(x) \leq 1$ for all $x \in \mathbb{R}$, $\varphi(0) = 1$, and φ is increasing in $(-\infty, 0]$ and decreasing in $[0, \infty)$. Then φ is called a bump function.

Definition 3.2. Let $\mathcal{M}$ be the set of all set functions

$$\mu : \{F \in \mathcal{B}(\mathbb{R}) : F \text{ bounded }\} \to \mathbb{R}$$

that can be written as a difference $\mu = \mu^1 - \mu^2$ of two non-negative measures $\mu^1, \mu^2 : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ such that $\mu^i(K) < \infty$ for $i \in \{1, 2\}$ and all compact sets $K \subset \mathbb{R}$.

A set function $\mu \in \mathcal{M}$ restricted to $\mathcal{B}(K)$ is a signed measure for all compact sets $K \subset \mathbb{R}$. It is not necessarily a signed measure on $\mathcal{B}(\mathbb{R})$, because it can be undefined for unbounded sets. However, we now show that a set function in $\mathcal{M}$ has an optimal decomposition corresponding to the Jordan decomposition of signed measures.

Theorem 3.3. *Suppose that $\mu \in \mathcal{M}$. Then there exist a unique decomposition $\mu = \mu^+ - \mu^-$, where $\mu^+, \mu^- : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ are non-negative measures such that $\mu^+(K) < \infty$ and $\mu^-(K) < \infty$ for all compact sets $K \subset \mathbb{R}$, with the property that $\mu^+(E) \leq \mu^1(E)$ and $\mu^-(E) \leq \mu^2(E)$ for all other decompositions $\mu = \mu^1 - \mu^2$ and all $E \in \mathcal{B}(\mathbb{R})$.*

Proof. Let $N \in \mathbb{Z}$. Then μ is a signed measure on the interval $[N, N + 1)$. By the Hahn decomposition theorem [18, Theorem 6.14] there exist sets $A^N, B^N \in \mathcal{B}([N, N + 1))$ such that $A^N \cup B^N = [N, N + 1)$, $A^N \cap B^N = \emptyset$, and the Jordan decomposition $\mu_{|[N,N+1)} = \mu_N^+ - \mu_N^-$ on the interval $[N, N + 1)$ satisfies

$$\mu_N^+(E) = \mu(A^N \cap E) \;\text{ and }\; \mu_N^-(E) = -\mu(B^N \cap E)$$

for $E \in \mathcal{B}([N, N+1))$. We define the sets

$$A^\infty := \bigcup_{N \in \mathbb{Z}} A^N \quad \text{and} \quad B^\infty := \bigcup_{N \in \mathbb{Z}} B^N,$$

which have the properties $A^\infty \cup B^\infty = \mathbb{R}$ and $A^\infty \cap B^\infty = \emptyset$. Now we set

$$\mu^+(E) = \sum_{N \in \mathbb{Z}} \mu(A^N \cap E) \quad \text{and} \quad \mu^-(E) = \sum_{N \in \mathbb{Z}} -\mu(B^N \cap E)$$

for $E \in \mathcal{B}(\mathbb{R})$. Then μ^+ and μ^- are non-negative measures on $\mathcal{B}(\mathbb{R})$ that are finite on all compact sets and bounded from above by measures of any other decomposition $\mu = \mu^1 - \mu^2$, since $\mu \le \mu^1$ and

$$\mu^+(E) = \sum_{N \in \mathbb{Z}} \mu(A^N \cap E) \le \sum_{N \in \mathbb{Z}} \mu^1(A^N \cap E) \le \mu_1(E)$$

for all $E \in \mathcal{B}(\mathbb{R})$. Similarly $\mu^- \le \mu^2$, since $-\mu \le \mu^2$. The finiteness of μ^+ and μ^- on compact sets now follows from the corresponding property in the definition of $\mathcal{M}$. We conclude that the measures μ^+ and μ^- give the Jordan decomposition of μ, and the sets A^∞ and B^∞ give the Hahn decomposition of $\mathbb{R}$ induced by μ. □

Definition 3.4. For $\mu \in \mathcal{M}$, we define

$$\|\mu\|_\varphi = \int_{\mathbb{R}} \varphi(x)\, d|\mu|(x),$$

where $|\mu| := \mu^+ + \mu^-$ is the σ-finite measure given by the optimal decomposition of μ in Theorem 3.3. Moreover, we define

$$\mathcal{M}_\varphi = \{\mu \in \mathcal{M} : \|\mu\|_\varphi < \infty\}.$$

Remark 3.5. If we restrict μ to $\mathcal{B}([-N, N])$, then the measure $|\mu|$ in the Definition 3.4 is the total variation measure of μ in the classical sense.

Lemma 3.6. $(\mathcal{M}_\varphi, \|\cdot\|_\varphi)$ *is a normed space.*

Proof. Let $\mu \in \mathcal{M}_\varphi$. If $\mu(F) = 0$ for all $F \in \mathcal{F}_b := \{F \in \mathcal{B}(\mathbb{R}) : F \text{ bounded }\}$, then obviously $\|\mu\|_\varphi = 0$. To show the opposite, take $F \in \mathcal{F}_b$. Since F is bounded, by the positivity of φ the condition $\|\mu\|_\varphi = 0$ implies that $|\mu|(F) = 0$. Thus $|\mu(F)| \le |\mu|(F) = 0$, so $\mu(F) = 0$, and $\|\cdot\|_\varphi$ is positive definite. Let $a \in \mathbb{R}$. Since $|a\mu|(F) = |a||\mu|(F)$, the homogeneity property $\|a\mu\|_\varphi = |a|\, \|\mu\|_\varphi$ follows from the properties of the integral. So does the triangle inequality, since $|\mu_1 + \mu_2|(F) \le |\mu_1|(F) + |\mu_2|(F)$ for $\mu_1, \mu_2 \in \mathcal{M}_\varphi$. Thus we conclude that $\mathcal{M}_\varphi$ is a vector space and $\|\cdot\|_\varphi$ is a norm in $\mathcal{M}_\varphi$. □

Definition 3.7 (The class BV_φ)**.** For any $\mu \in \mathcal{M}_\varphi$, we define the distribution function related to μ by

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

where $[0,0) = \emptyset$. We denote the class of all such functions by BV_φ.

Remark 3.8. By definition a function $g \in BV_\varphi$ is left-continuous with at most countably many jumps, and $g(0) = 0$. The size of the class BV_φ depends on the decay of φ; we will show in Theorem 3.10 that no decay for φ, i.e., $\varphi \equiv 1$ leads back to BV. Using the uniqueness of the signed measure representation of BV functions we can show that there is one-to-one correspondence between measures in $\mathcal{M}_\varphi$ and functions in BV_φ, i.e. for $\mu_1, \mu_2 \in \mathcal{M}_\varphi$ we have $g^{\mu_1} = g^{\mu_2}$ if and only if $\mu_1 = \mu_2$.

Theorem 3.9. *BV_φ is a Banach space with respect to $\|g^\mu\|_\varphi := \|\mu\|_\varphi$.*

Proof. It is easy to see that BV_φ is a vector space, and Lemma 3.6 ensures that $\|\cdot\|_\varphi$ is a norm. Recall that X is a Banach space if and only if every absolutely convergent sum of elements of X converges [16, Theorem III.3]. Let $(g^{\mu_i})_{i=1}^\infty \subset BV_\varphi$ be a sequence that converges absolutely, i.e. $\sum_{i=1}^\infty \|g^{\mu_i}\|_\varphi < \infty$. By definition this is equivalent to

$$\sum_{i=1}^\infty \int_{\mathbb{R}} \varphi(x)\, d|\mu_i|(x) < \infty. \tag{3.1}$$

Here $(\mu_i)_{i=1}^\infty$ is a sequence in $\mathcal{M}_\varphi$, and each μ_i has the representation $\mu_i = \mu_i^+ - \mu_i^-$ according to Theorem 3.3. For any $x > 0$ we have

$$\begin{aligned}
\sum_{i=1}^\infty |g^{\mu_i}(x)| &= \sum_{i=1}^\infty |\mu_i([0,x))| \le \sum_{i=1}^\infty |\mu_i|([0,x)) \\
&= \sum_{i=1}^\infty \int_{[0,x)} d|\mu_i|(z) = \sum_{i=1}^\infty \frac{1}{\varphi(x)} \int_{[0,x)} \varphi(x)\, d|\mu_i|(z) \\
&\le \frac{1}{\varphi(x)} \sum_{i=1}^\infty \int_{[0,x)} \varphi(z)\, d|\mu_i|(z) \le \frac{1}{\varphi(x)} \sum_{i=1}^\infty \|g^{\mu_i}\|_\varphi < \infty,
\end{aligned}$$

and for $x \le 0$ we can do a similar computation. This shows that the sum $\sum_{i=1}^\infty g^{\mu_i}(x)$ exists for all $x \in \mathbb{R}$. Moreover, since $\mu_i^+ \le |\mu_i|$ and $\mu_i^- \le |\mu_i|$, we also get that

$$\sum_{i=1}^\infty \mu_i^+([0,x)) < \infty \text{ for } x > 0, \text{ and } \sum_{i=1}^\infty \mu_i^+([x,0)) < \infty \text{ for } x \le 0.$$

Similar results hold for μ_i^-. We define

$$\mu^1 := \sum_{i=1}^\infty \mu_i^+ \quad \text{and} \quad \mu^2 := \sum_{i=1}^\infty \mu_i^-.$$

Now $\mu^1, \mu^2 : \mathcal{B}(\mathbb{R}) \to [0,\infty]$ are measures as sums of measures, and finite on all compact sets. Thus they satisfy the conditions of Definition 3.2 and hence $\mu := \mu^1 - \mu^2 \in \mathcal{M}$. For $x > 0$ we have

$$\begin{aligned}\mu([0,x)) &= \sum_{i=1}^{\infty} \mu_i^+([0,x)) - \sum_{i=1}^{\infty} \mu_i^-([0,x)) \\ &= \sum_{i=1}^{\infty} \big(\mu_i^+([0,x)) - \mu_i^-([0,x))\big) \\ &= \sum_{i=1}^{\infty} \mu_i([0,x)) = \sum_{i=1}^{\infty} g^{\mu_i}(x),\end{aligned}$$

and for $x \le 0$,

$$\mu([x,0)) = -\sum_{i=1}^{\infty} g^{\mu_i}(x).$$

Moreover, the function g^μ defined by μ is in BV_φ if the condition

$$\int_{\mathbb{R}} \varphi \, d|\mu| < \infty \tag{3.2}$$

is satisfied. Because $\mu = \mu^1 - \mu^2$ is not necessarily the optimal (Jordan) decomposition of μ, we get by Theorem 3.3 that $|\mu| \le \mu^1 + \mu^2$. This implies that

$$\int_{\mathbb{R}} \varphi \, d|\mu| \le \int_{\mathbb{R}} \varphi \, d(\mu^1 + \mu^2) = \int_{\mathbb{R}} \varphi \, d\mu^1 + \int_{\mathbb{R}} \varphi \, d\mu^2.$$

Now

$$\begin{aligned}\int_{\mathbb{R}} \varphi \, d\mu^1 &= \int_{\mathbb{R}} \varphi \, d\left(\sum_{i=1}^{\infty} \mu_i^+\right) = \sum_{i=1}^{\infty} \int_{\mathbb{R}} \varphi \, d\mu_i^+ \\ &\le \sum_{i=1}^{\infty} \int_{\mathbb{R}} \varphi \, d|\mu_i| < \infty,\end{aligned}$$

where the second equality follows from [6, p. 179]. We can do a similar computation for the integral with respect to μ^2. Thus the condition (3.2) is satisfied, and we conclude that $g^\mu \in BV_\varphi$.

It remains to show that the sum $\sum_{i=1}^n g^{\mu_i}$ converges to the element g^μ as $n \to \infty$. This follows from the fact that

$$\begin{aligned}&\left\| g^\mu - \sum_{i=1}^n g^{\mu_i} \right\|_\varphi = \left\| \mu - \sum_{i=1}^n \mu_i \right\|_\varphi = \left\| \mu^1 - \mu^2 - \sum_{i=1}^n \mu_i \right\|_\varphi \\ &= \left\| \sum_{i=1}^{\infty} \mu_i^+ - \sum_{i=1}^{\infty} \mu_i^- - \sum_{i=1}^{n} (\mu_i^+ - \mu_i^-) \right\|_\varphi = \left\| \sum_{i=n+1}^{\infty} \mu_i^+ - \sum_{i=n+1}^{\infty} \mu_i^- \right\|_\varphi\end{aligned}$$

$$= \int_{\mathbb{R}} \varphi \, d \left| \sum_{i=n+1}^{\infty} \mu_i^+ - \sum_{i=n+1}^{\infty} \mu_i^- \right| \leq \int_{\mathbb{R}} \varphi \, d \left(\sum_{i=n+1}^{\infty} \mu_i^+ + \sum_{i=n+1}^{\infty} \mu_i^- \right)$$
$$= \int_{\mathbb{R}} \varphi \, d \left(\sum_{i=n+1}^{\infty} |\mu_i| \right) = \sum_{i=n+1}^{\infty} \int_{\mathbb{R}} \varphi \, d|\mu_i| \to 0$$

as $n \to \infty$, because the sum on the right hand side converges. □

Let us show that the class BV_φ contains all correctly normalized and left-continuous functions of bounded variation. Define

$$A := BV \cap \{g \text{ left-continuous and } g(0) = 0\}.$$

Theorem 3.10. *We have $A \subset BV_\varphi$ for any bump function φ. If $\varphi \equiv 1$, then $A = BV_\varphi$.*

Proof. Suppose that $g \in BV$ is left-continuous. Then, by Theorems 8.13 and 8.14 in [17], there exist a unique signed measure μ and a constant $c \in \mathbb{R}$ such that $g(x) = c + \mu((-\infty, x))$. The assumption $g(0) = 0$ then implies that $c = -\mu((-\infty, 0))$. Thus for $x > 0$ we have

$$g(x) = c + \mu((-\infty, 0)) + \mu([0, x)) = \mu([0, x)),$$

and for $x \leq 0$ we have

$$g(x) = c + \mu((-\infty, 0)) - \mu([x, 0)) = -\mu([x, 0)),$$

which coincides with the measure representation of functions in BV_φ. Moreover,

$$\|g\|_\varphi = \int_{\mathbb{R}} \varphi \, d|\mu| \leq \int_{\mathbb{R}} d|\mu| = V(g) < \infty,$$

and $g \in BV_\varphi$.

If $\varphi \equiv 1$ and $g^\mu \in BV_\varphi$ for some $\mu \in \mathcal{M}_\varphi$, then there exists $M > 0$ such that $\|g^\mu\|_\varphi = |\mu|(\mathbb{R}) < M$. Since $\mu^+(\mathbb{R}) + \mu^-(\mathbb{R}) = |\mu|(\mathbb{R})$, we see that both μ^+ and μ^- are finite measures, μ is a signed measure on $\mathbb{R}$, and $g^\mu \in BV$ by Theorems 8.13 and 8.14 in [17].

□

The space BV_φ satisfies the following comparison properties:

Lemma 3.11. *Suppose that φ and ψ are bump functions. If $\varphi \leq \psi$, then $BV_\psi \subset BV_\varphi$. In particular, if $g \in BV_\psi$, then $\|g\|_\varphi \leq \|g\|_\psi$.*

Proof. Let $g \in BV_\psi$ and let $\mu \in \mathcal{M}_\psi$ be the measure related to g. Then

$$\|\mu\|_\varphi = \int_{\mathbb{R}} \varphi(x) \, d|\mu|(x) \leq \int_{\mathbb{R}} \psi(x) \, d|\mu|(x) = \|\mu\|_\psi < \infty,$$

which implies the statement. □

Remark 3.12. The space BV_φ is not separable, which can be seen by considering the uncountable set of functions $\chi_{(K,\infty)}/\varphi(K)$, $K \in \mathbb{R}$. Namely, we can write $\chi_{(K,\infty)}/\varphi(K) = g^{\mu_K}$ with $\mu_K = \delta_{\{K\}}/\varphi(K)$, where δ is the Dirac delta. This implies that if $K_1 \neq K_2$, then $\|g^{\mu_{K_1}} - g^{\mu_{K_2}}\|_\varphi = 2$, and it is impossible to find a countable dense subset.

4. Incorporation of jumps

By definition, functions $g^\mu \in BV_\varphi$ have only countably many jumps, are left-continuous, and vanish at the origin. We can relax the latter two restrictions by adding to the function g^μ a constant $c \in \mathbb{R}$ and a jump function Δ, which is zero outside a countable set. This extends the class BV_φ to include functions that have different types of discontinuity, i.e. points of left-continuity, right-continuity and neither left- nor right-continuity. For example, we can make $g^\mu \in BV_\varphi$ right-continuous by choosing

$$\Delta(x) = \begin{cases} g^\mu(x+) - g^\mu(x) \text{ for } x \in A, \\ 0 \text{ elsewhere,} \end{cases}$$

where A is the set of points of discontinuity of g^μ.

Definition 4.1. Define a set of jump functions

$$\Delta_\varphi = \{\Delta^\nu : \mathbb{R} \to \mathbb{R} \mid \Delta^\nu(x) = \nu(\{x\}), \nu \in \mathcal{M}_\varphi^\Delta\},$$

where

$$\mathcal{M}_\varphi^\Delta = \{\mu \in \mathcal{M}_\varphi : \mu = \sum_{i=1}^\infty \alpha_i \delta_{\{x_i\}} \text{ with } \alpha_i, x_i \in \mathbb{R}, x_i \neq x_j \text{ for } i \neq j\}$$

This gives us a set of functions that can have non-zero values only in countably many points $(x_i)_{i=1}^\infty$, and the condition $\nu \in \mathcal{M}_\varphi^\Delta$ states that

$$\|\Delta^\nu\|_\varphi := \|\nu\|_\varphi = \sum_{i=1}^\infty \varphi(x_i)|\nu(\{x_i\})| = \sum_{i=1}^\infty \varphi(x_i)|\alpha_i| < \infty.$$

We have uniqueness of the decomposition $g = c + g^\mu + \Delta^\nu$:

Theorem 4.2. *If $c_1 + g^{\mu_1} + \Delta^{\nu_1} = c_2 + g^{\mu_2} + \Delta^{\nu_2}$ with $c_1, c_2 \in \mathbb{R}$, $\mu_1, \mu_2 \in \mathcal{M}_\varphi$ and $\nu_1, \nu_2 \in \mathcal{M}_\varphi^\Delta$, then $c_1 = c_2$, $\mu_1 = \mu_2$ and $\nu_1 = \nu_2$.*

Proof. Take two functions g_1 and g_2 such that $g_i = c_i + g^{\mu_i} + \Delta^{\nu_i}$, $i \in \{1,2\}$, and suppose that $g_1 = g_2$. Define $A_i = \operatorname{supp} \Delta^{\nu_i}$, $i \in \{1,2\}$. Now $A_1 \cup A_2$ is countable and $\Delta^{\nu_i} = 0$ in $(A_1 \cup A_2)^c$. Take a sequence $(x_j) \subset (A_1 \cup A_2)^c$ such that $x_j \nearrow 0$ as $j \to \infty$. Since g^{μ_i} is left-continuous and $g^{\mu_i}(0) = 0$, it follows that $g_i(x_j) = c_i + g^{\mu_i}(x_j) \to c_i$ as $j \to \infty$, and thus $c_1 = c_2$. This implies that for $x_0 \in (A_1 \cup A_2)^c$ we have $g^{\mu_1}(x_0) = g^{\mu_2}(x_0)$. Now let $x_0 \in A_1 \cup A_2$. Again we choose a sequence $(x_j) \subset (A_1 \cup A_2)^c$ such that $x_j \nearrow x_0$ as $j \to \infty$, and by left-continuity of g^{μ_i} we get that $g^{\mu_1}(x_0) = g^{\mu_2}(x_0)$. Thus $g^{\mu_1} = g^{\mu_2}$ everywhere, and also $\Delta^{\nu_1} = \Delta^{\nu_2}$. $\square$

Theorem 4.3. *$BV \subset \{g = c + g^\mu + \Delta^\nu : c \in \mathbb{R},\ g^\mu \in BV_\varphi,\ \Delta^\nu \in \Delta_\varphi\}$.*

Proof. Theorem 3.10 shows that the component g^μ covers all $g \in BV$ that are left-continuous and satisfy $g(0) = 0$. Since the latter condition can be relaxed by adding a constant $c \in \mathbb{R}$, we only need to deal with the points of discontinuity of g. But if $\{x_i\}_{i=1}^\infty$ is the set of these points, then we can alter the left-continuous part g^μ by adding a function Δ^ν with $\nu(x) = \sum_{i=1}^\infty \lambda_i \chi_{\{x_i\}}(x)$, where the coefficients λ_i are the necessary changes at the points x_i. Then $\Delta^\nu \in \Delta_\varphi$, since

$$\|\Delta^\nu\|_\varphi = \sum_{i=1}^\infty \varphi(x_i)|\nu(\{x_i\})| = \sum_{i=1}^\infty \varphi(x_i)|\lambda_i| \le \sum_{i=1}^\infty |\lambda_i| \le V(g),$$

and any $g \in BV$ admits a representation $g = c + g^\mu + \Delta^\nu$. $\square$

5. Alternative characterization of BV_φ

In this chapter we characterize the class BV_φ in a more intuitive way. Given a bump function φ and a signed measure ν, we can generate a function in BV_φ by computing the integral of $1/\varphi$ with respect to ν.

Definition 5.1 (Class BVR_φ)**.** For a bump function φ, we denote by BVR_φ 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 $[0,0) = \emptyset$.

Theorem 5.2. *$BV_\varphi = BVR_\varphi$.*

Proof. Let $g^\mu \in BV_\varphi$ with $\mu \in \mathcal{M}_\varphi$. Then by Theorem 3.3, μ admits the Jordan decomposition $\mu = \mu^+ - \mu^-$ with

$$\int_{\mathbb{R}} \varphi\, d\mu^+ + \int_{\mathbb{R}} \varphi\, d\mu^- < \infty.$$

For $E \in \mathcal{B}(\mathbb{R})$, we define finite measures

$$\nu^+(E) = \int_E \varphi\, d\mu^+ \quad \text{and} \quad \nu^-(E) = \int_E \varphi\, d\mu^-,$$

or $d\nu^+ = \varphi\, d\mu^+$ and $d\nu^- = \varphi\, d\mu^-$. Hence, for $E \in \mathcal{B}(\mathbb{R})$, we have

$$\mu^+(E) = \int_E d\mu^+ = \int_E \frac{1}{\varphi}\varphi\, d\mu^+ = \int_E \frac{1}{\varphi}\, d\nu^+,$$

and similarly

$$\mu^-(E) = \int_E \frac{1}{\varphi}\, d\nu^-.$$

Thus

$$\mu(E) = \int_E d\mu^+ - \int_E d\mu^- = \int_E \frac{1}{\varphi}\, d\nu^+ - \int_E \frac{1}{\varphi}\, d\nu^- = \int_E \frac{1}{\varphi}\, d\nu$$

and $g^\mu \in BVR_\varphi$.

The converse follows by similar arguments. Let $g \in BVR_\varphi$ with the underlying signed measure ν with Jordan decomposition $\nu = \nu^+ - \nu^-$. We define a set function $\mu = \mu^1 - \mu^2$ such that

$$\mu^1(E) = \int_E \frac{1}{\varphi}\, d\nu^+ \quad \text{and} \quad \mu^2(E) = \int_E \frac{1}{\varphi}\, d\nu^-$$

for all $E \in \mathcal{B}(\mathbb{R})$. Then we get

$$d\nu^+ = \varphi\, d\mu^1 \text{ and } d\nu^- = \varphi\, d\mu^2,$$

and the finiteness of ν^+ and ν^- imply that

$$\int_{\mathbb{R}} \varphi\, d|\mu| \leq \int_{\mathbb{R}} \varphi\, d\mu^1 + \int_{\mathbb{R}} \varphi\, d\mu^2 < \infty.$$

Thus μ is in $\mathcal{M}_\varphi$ and $g = g^\mu \in BV_\varphi$. □

Example 5.3. Suppose that φ is a continuously differentiable bump function. Because $\varphi \in BV$, we may choose in Definition 5.1 a signed measure ν such that $\varphi(x) = \nu((-\infty, x))$, in accordance with Remark 2.8. Then by [1, Theorem 7.35] we have $d\nu = \varphi'\, dx$ and $g = \log\varphi \in BVR_\varphi$, and thus $\log\varphi \in BV_\varphi$ by Theorem 5.2.

6. Convergence results

Let X and $\hat{X}$ be random variables defined on a common probability space. We define a bump function $\varphi^{X,\hat{X}}$ that connects the random variables with their tail behavior. Then, assuming that the rate $\beta_p > 0$ in the error

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

is known, we find convergence rates γ_p for the error

$$\left\| g(X) - g(\hat{X}) \right\|_q^q \leq C(p, q, g, X, \hat{X}) \left\| X - \hat{X} \right\|_p^{\gamma_p},$$

for functions g in the class $BV_{\varphi^{X,\hat{X}}}$ associated with the function $\varphi^{X,\hat{X}}$.

We can apply the above principle together with results giving the value of β_p. We may use [3, Theorem 2.4] to show that the optimal power is $\beta_p = p/(p+1)$ for random variables X and $\hat{X}$ such that X has a bounded density, or take advantage of a setting with additional information about X and $\hat{X}$, such as Gaussianity as in Theorem 1.1, to obtain better powers β_p.

Definition 6.1. Take two continuous and strictly positive functions $\varphi^+ : (0, \infty) \to (0, 1]$ and $\varphi^- : (-\infty, 0] \to (0, 1]$ with the properties that φ^+ is decreasing and $\varphi^+(0) = 1$, φ^- is increasing and $\varphi^-(0) = 1$,

$$\mathbb{P}(X \geq K) \vee \mathbb{P}(\hat{X} \geq K) \leq \varphi^+(K) \text{ for } K > 0$$

and

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

Then we define a bump function $\varphi^{X,\hat{X}} : \mathbb{R} \to (0, 1]$ by

$$\varphi^{X,\hat{X}}(K) := \begin{cases} \varphi^+(K) & \text{if } K > 0, \\ \varphi^-(K) & \text{if } K \leq 0. \end{cases}$$

The main result of this section is the following convergence theorem:

Theorem 6.2. *Let $1 \leq p \leq \infty$ and $\beta_p > 0$. Suppose that X and $\hat{X}$ are random variables that satisfy*

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

for all $K \in \mathbb{R}$. Suppose that $0 < \theta < 1$ and consider the bump function $\varphi^{X,\hat{X}}$. If $1 \leq q < \infty$ and $g \in BV_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}}$, then

$$\left\| g(X) - g(\hat{X}) \right\|_q^q \leq 2^\theta \, \|g\|^q_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}} C(p, X)^{1-\theta} \left\| X - \hat{X} \right\|_p^{(1-\theta)\beta_p}. \tag{6.1}$$

Remark 6.3. Theorem 6.2 is an extension of [3, Theorem 2.4], which considers the case $\theta = 0$ and $\beta_p = p/(p+1)$. The intuition given by plugging $\theta = 0$ into Equation (6.1) is correct for the class BV_1, which is a subspace of BV by Theorem 3.10. The statement is formally proved for all functions in BV in [3, Theorem 2.4].

Proof of Theorem 6.2. Let $g \in BV_{\varphi^{X,\hat{X}}}$ and let μ be the set function associated with g, i.e. $g = g^\mu$. We use the optimal decomposition $\mu = \mu^+ - \mu^-$ given by Theorem 3.3. Now

$$\begin{aligned} g^\mu(x)\chi_{(0,\infty)}(x) &= \mu([0, x))\chi_{(0,\infty)}(x) \\ &= \left(\mu^+([0, x)) - \mu^-([0, x))\right) \chi_{(0,\infty)}(x) \\ &= \int_{[0,x)} \chi_{(0,\infty)}(x) \, d\mu^+(z) - \int_{[0,x)} \chi_{(0,\infty)}(x) \, d\mu^-(z) \\ &= \int_{[0,\infty)} \chi_{(z,\infty)}(x) \, d\mu^+(z) - \int_{[0,\infty)} \chi_{(z,\infty)}(x) \, d\mu^-(z), \end{aligned}$$

and similarly

$$\begin{aligned} g^\mu(x)\chi_{(-\infty,0]}(x) &= -\mu([x, 0))\chi_{(-\infty,0]}(x) \\ = -\left(\int_{(-\infty,0)} \chi_{(-\infty,z]}(x) \, d\mu^+(z) \right. &\left. - \int_{(-\infty,0)} \chi_{(-\infty,z]}(x) \, d\mu^-(z) \right). \end{aligned}$$

Thus we can use these representations to get that

$$
\begin{aligned}
\left\| g^\mu(X) - g^\mu(\hat{X}) \right\|_q &\le \left\| g^\mu(X)\chi_{(0,\infty)}(X) - g^\mu(\hat{X})\chi_{(0,\infty)}(\hat{X}) \right\|_q \\
&\quad + \left\| g^\mu(X)\chi_{(-\infty,0]}(X) - g^\mu(\hat{X})\chi_{(-\infty,0]}(\hat{X}) \right\|_q \\
&\le \left\| \int_{[0,\infty)} |\chi_{(z,\infty)}(X) - \chi_{(z,\infty)}(\hat{X})| \, d\mu^+(z) \right\|_q \\
&\quad + \left\| \int_{[0,\infty)} |\chi_{(z,\infty)}(X) - \chi_{(z,\infty)}(\hat{X})| \, d\mu^-(z) \right\|_q \\
&\quad + \left\| \int_{(-\infty,0)} |\chi_{(-\infty,z]}(X) - \chi_{(-\infty,z]}(\hat{X})| \, d\mu^+(z) \right\|_q \\
&\quad + \left\| \int_{(-\infty,0)} |\chi_{(-\infty,z]}(X) - \chi_{(-\infty,z]}(\hat{X})| \, d\mu^-(z) \right\|_q .
\end{aligned}
$$

For the first term we have

$$
\begin{aligned}
&\left\| \int_{[0,\infty)} |\chi_{(z,\infty)}(X) - \chi_{(z,\infty)}(\hat{X})| \, d\mu^+(z) \right\|_q \\
&\le \int_{[0,\infty)} \left\| \chi_{(z,\infty)}(X) - \chi_{(z,\infty)}(\hat{X}) \right\|_q \, d\mu^+(z),
\end{aligned}
$$

and similarly for the other three terms. Let us now look for an upper bound for the L_q-norm in the integrand. Denote by $\psi(X, \hat{X})$ the error given in the assumption, i.e.

$$
\psi(X, \hat{X}) := C(p, X) \left\| X - \hat{X} \right\|_p^{\beta_p}
$$

and notice that $a \wedge b \le a^{1-\theta} b^\theta$ for any $a, b \ge 0$ and $0 < \theta < 1$. Since

$$
\begin{aligned}
\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| &= \mathbb{P}(X \ge K, \hat{X} < K) + \mathbb{P}(X < K, \hat{X} \ge K) \\
&\le 2(\mathbb{P}(X \ge K) \vee \mathbb{P}(\hat{X} \ge K)),
\end{aligned}
$$

it follows that, for $K > 0$ and $0 < \theta < 1$,

$$
\begin{aligned}
\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| &\le \psi(X, \hat{X}) \wedge 2 \left[\mathbb{P}(X \ge K) \vee \mathbb{P}(\hat{X} \ge K) \right] \\
&\le \psi(X, \hat{X})^{1-\theta} 2^\theta \left[\mathbb{P}(X \ge K) \vee \mathbb{P}(\hat{X} \ge K) \right]^\theta \\
&\le 2^\theta \psi(X, \hat{X})^{1-\theta} \varphi^+(K)^\theta.
\end{aligned}
$$

In a similar way we get for $K \le 0$ that

$$
\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| \le 2^\theta \psi(X, \hat{X})^{1-\theta} \varphi^-(K)^\theta,
$$

so we write for $K \in \mathbb{R}$,

$$
\mathbb{E}|\chi_{[K,\infty)}(X) - \chi_{[K,\infty)}(\hat{X})| \le 2^\theta \psi(X, \hat{X})^{1-\theta} \varphi^{X,\hat{X}}(K)^\theta.
$$

This gives an estimate for $\left\|\chi_{[z,\infty)}(X)-\chi_{[z,\infty)}(\hat{X})\right\|_q$, which in turn implies the needed estimate for the function $\chi_{(z,\infty)}$ by writing $\chi_{(z,\infty)} = \lim_{\varepsilon\to 0}\chi_{[z+\varepsilon,\infty)}$ and using the dominated convergence theorem. Moreover, we get the same estimate for the function $\chi_{(-\infty,z]}$ by looking at the complement of the interval (z,∞). Therefore,

$$
\begin{aligned}
&\left\|g^\mu(X)-g^\mu(\hat{X})\right\|_q \\
&\le \int_{[0,\infty)} \left\|\chi_{(z,\infty)}(X)-\chi_{(z,\infty)}(\hat{X})\right\|_q d(\mu^+ + \mu^-)(z) \\
&\quad + \int_{(-\infty,0)} \left\|\chi_{(-\infty,z]}(X)-\chi_{(-\infty,z]}(\hat{X})\right\|_q d(\mu^+ + \mu^-)(z) \\
&\le \int_{\mathbb{R}} 2^{\frac{\theta}{q}} \psi(X,\hat{X})^{\frac{1-\theta}{q}} \varphi^{X,\hat{X}}(z)^{\frac{\theta}{q}}\, d|\mu|(z) \\
&\le 2^{\frac{\theta}{q}} \int_{\mathbb{R}} C(p,X)^{\frac{1-\theta}{q}} \left\|X-\hat{X}\right\|_p^{\frac{(1-\theta)\beta_p}{q}} \varphi^{X,\hat{X}}(z)^{\frac{\theta}{q}}\, d|\mu|(z) \\
&\le 2^{\frac{\theta}{q}} C(p,X)^{\frac{1-\theta}{q}} \int_{\mathbb{R}} \varphi^{X,\hat{X}}(z)^{\frac{\theta}{q}}\, d|\mu|(z) \left\|X-\hat{X}\right\|_p^{\frac{(1-\theta)\beta_p}{q}} \\
&= 2^{\frac{\theta}{q}} C(p,X)^{\frac{1-\theta}{q}} \|g\|_{\left(\varphi^{X,\hat{X}}\right)^{\frac{\theta}{q}}} \left\|X-\hat{X}\right\|_p^{\frac{(1-\theta)\beta_p}{q}},
\end{aligned}
$$

which gives the assertion. □

Similar result holds for the jump functions defined in Section 4:

Lemma 6.4. *Let $1 \le q < \infty$, $0 < \theta < 1$, and $\Delta^\nu \in \Delta_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}}$ with $\nu \in \mathcal{M}^{\Delta}_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}}$. Under the assumptions of Theorem 6.2,*

$$
\left\|\Delta^\nu(X)-\Delta^\nu(\hat{X})\right\|_q^q \le 2^{q+\theta} \left\|\Delta^\nu\right\|^q_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}} C(p,X)^{1-\theta} \left\|X-\hat{X}\right\|_p^{(1-\theta)\beta_p}.
$$

Proof. First, note that the function Δ^ν can be written in the form

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

Then by arguments similar to those employed in the proof of Theorem 6.2,

$$
\begin{aligned}
&\left\|\Delta^\nu(X)-\Delta^\nu(\hat{X})\right\|_q \\
&\le \int_{\mathbb{R}} \left\|\chi_{(-\infty,z]}(X)-\chi_{(-\infty,z]}(\hat{X})\right\|_q d(\nu^+ + \nu^-)(z) \\
&+ \int_{\mathbb{R}} \left\|\chi_{(-\infty,z)}(X)-\chi_{(-\infty,z)}(\hat{X})\right\|_q d(\nu^+ + \nu^-)(z)
\end{aligned}
$$

$$
\begin{aligned}
&\leq 2 \cdot 2^{\frac{\theta}{q}} C(p, X)^{\frac{1-\theta}{q}} \int_{\mathbb{R}} \varphi^{X,\hat{X}}(z)^{\frac{\theta}{q}}\, d|\nu|(z) \left\| X - \hat{X} \right\|_p^{\frac{(1-\theta)\beta_p}{q}} \\
&\leq 2^{\frac{q+\theta}{q}} C(p, X)^{\frac{1-\theta}{q}} \left\| \Delta^\nu \right\|_{\left(\varphi^{X,\hat{X}}\right)^{\frac{\theta}{q}}} \left\| X - \hat{X} \right\|_p^{\frac{(1-\theta)\beta_p}{q}} .
\end{aligned}
$$

□

Now Theorem 6.2, Lemma 6.4, and Minkowski's inequality imply the following result for not necessarily left-continuous functions.

Corollary 6.5. *Let $1 \leq q < \infty$ and $0 < \theta < 1$. If $g = c + g^\mu + \Delta^\nu$, where $c \in \mathbb{R}$, $g^\mu \in BV_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}}$, and $\Delta^\nu \in \Delta_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}}$, then under the assumptions of Theorem 6.2,*

$$
\begin{aligned}
\left\| g(X) - g(\hat{X}) \right\|_q^q \;\leq\;& 2^{q+\theta} \left(\left\| g^\mu \right\|_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}} + \left\| \Delta^\nu \right\|_{(\varphi^{X,\hat{X}})^{\frac{\theta}{q}}} \right)^q \\
& \cdot C(p, X)^{1-\theta} \left\| X - \hat{X} \right\|_p^{(1-\theta)\beta_p} .
\end{aligned}
$$

7. Extension for fine variations

The functions in the space BV_φ have bounded variation on a compact set. By simple concavity arguments, we show a result similar to Theorem 6.2 for certain functions that have variation that is small in amplitude, but so frequent that the function has unbounded variation on a compact set. We give an example of such a function in Example 7.3.

Let X and $\widehat{X}$ be random variables on the same probability space. For $0 < r \leq 1$, we define the set

$$
BV^r := \{ g_{r,f} : \mathbb{R} \to \mathbb{R} \ : \ g_{r,f} = |f|^r \operatorname{sgn} f, \ f \in BV \},
$$

where sgn is the signum function.

Theorem 7.1. *Let $1 \leq p \leq \infty$, $\beta_p > 0$, $0 < r \leq 1$, and $g_{r,f} \in BV^r$. Suppose that X and $\hat{X}$ are random variables that satisfy*

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

for all $K \in \mathbb{R}$, and suppose X has a bounded density f_X. If $0 < q < \infty$, then

$$
\left\| g_{r,f}(X) - g_{r,f}(\widehat{X}) \right\|_q^q \leq 3^{2q} V(f)^{rq \vee 1} C(p, X) \left\| X - \widehat{X} \right\|_p^{\beta_p (rq \wedge 1)} .
$$

Proof. For $0 < r \leq 1$ and $x, y \geq 0$, we have

$$
|x^r - y^r| \leq |x - y|^r,
$$

because the function x^r, $x \geq 0$, is increasing and concave. Similarly for $x, y \leq 0$ we get

$$
||x|^r - |y|^r| \leq |x - y|^r.
$$

If $xy < 0$, then

$$||x|^r \operatorname{sgn} x - |y|^r \operatorname{sgn} y| \le |x|^r + |y|^r \le 2|x-y|^r. \tag{7.1}$$

Thus we see that Equation (7.1) holds for all $x, y \in \mathbb{R}$. By the assumption and [3, Proof of Theorem 2.4], if $f \in BV$ then

$$\left\| f(X) - f(\widehat{X}) \right\|_q^q \le 3^q V(g)^q C(p, X) \left\| X - \widehat{X} \right\|_p^{\beta_p}. \tag{7.2}$$

Thus,

$$\begin{aligned}
\left\| g_{r,f}(X) - g_{r,f}(\widehat{X}) \right\|_q^q &\le \left\| 2|f(X) - f(\widehat{X})|^r \right\|_q^q \\
&= 2^q \left\| f(X) - f(\widehat{X}) \right\|_{rq}^{rq} \\
&\le 2^q \left\| f(X) - f(\widehat{X}) \right\|_{rq\vee 1}^{(rq\vee 1)(rq\wedge 1)} \\
&\le 3^{2q} V(f)^{rq\vee 1} C(p, X) \left\| X - \widehat{X} \right\|_p^{\beta_p (rq\wedge 1)},
\end{aligned}$$

as desired. □

Remark 7.2. If we only know that X and $\widehat{X}$ are random variables such that X has a bounded density, then the optimal power in Equation (7.2) is $\beta_p = p/(p+1)$ by [3, Theorem 2.4].

Example 7.3. Let $0 < r < 1$ and define a function $g : \mathbb{R} \to \mathbb{R}$ such that

$$g(1/k) = \frac{(-1)^k}{k}$$

for $k = 1, 2, \ldots$, and $g(0) = 0$. Elsewhere on the interval $[0, 1]$ we define g by linear interpolation, and outside $[0, 1]$ by continuous constant extension. Then $g \notin BV$, because

$$V(g) \ge \sum_{k=1}^{\infty} \frac{1}{k} = \infty,$$

but $f := |g|^{1/r} \operatorname{sgn} g \in BV$, since $1/r > 1$ and

$$V(f) \le \sum_{k=1}^{\infty} \frac{2}{k^{1/r}} < \infty.$$

Therefore $g_{r,f} = g \in BV^r$ and the result of Theorem 7.1 holds for $g_{r,f}$.

For $r = 1$ the spaces BV^r and BV are equal. By Example 7.3 we see that BV^r is not included in BV for $0 < r < 1$. However, the converse is true by the following theorem. Hence BV is a nontrivial subspace of BV^r for $0 < r < 1$.

Theorem 7.4. *Let* $0 < r \le 1$. *Then* $BV \subset BV^r$.

Proof. The case $r = 1$ is trivial, so let us consider $0 < r < 1$. Take $g \in BV$ such that $g \geq 0$. We wish to represent g as $g = |f|^r \operatorname{sgn} f$ for some $f \in BV$. The condition $g \geq 0$ requires that $f \geq 0$, and the representation simplifies to $g = f^r$. Therefore the function $f = g^{1/r}$ gives the correct representation. Moreover, $f \in BV$ by the mean value theorem, since

$$\sum_{i=0}^{N} |g(x_i)^{1/r} - g(x_{i-1})^{1/r}| = \sum_{i=0}^{N} \frac{1}{r}\, \xi^{\frac{1}{r}-1} |g(x_i) - g(x_{i-1})|$$

for a partition $-\infty < x_0 < \cdots < x_N < \infty$ and some value $\xi \in [\min(g(x_i), g(x_{i-1})), \max(g(x_i), g(x_{i-1}))] \subset [0, \sup_{x\in\mathbb{R}} g(x)]$. This implies that

$$V(f) \leq \frac{1}{r} \left(\sup_{x\in\mathbb{R}} g(x)\right)^{\frac{1}{r}-1} V(g) < \infty.$$

Thus $g = f^r \in BV^r$.

If g has values in the reals, then we write $g = g^+ - g^-$, where $g^+ = \max(0, g)$ and $g^- = \max(0, -g)$ are the positive and negative parts of g. It it easy to see that both parts are in BV. Now by the first part of the proof the functions g^+ and g^- are in BV^r and have the representations $g^+ = (f^+)^r = |f^+|^r \operatorname{sgn} f^+$ with $f^+ \in BV$ and $g^- = (f^-)^r = |f^-|^r \operatorname{sgn} f^-$ with $f^- \in BV$. Then also $f^+ - f^- \in BV$. For all $x \in \mathbb{R}$ at least one of g^+ and g^-, and respectively of f^+ and f^-, is always zero. Thus, due to the pointwise nature of the representation, we have $g = g^+ - g^- = |f^+|^r \operatorname{sgn} f^+ + |-f^-|^r \operatorname{sgn}(-f^-) = |f^+ - f^-|^r \operatorname{sgn}(f^+ - f^-) \in BV^r$.

□

Remark 7.5. We may also define

$$BV_\varphi^r := \{g_{r,f} : \mathbb{R} \to \mathbb{R} \ : \ g_{r,f} = |f|^r \operatorname{sgn} f,\ f \in BV_\varphi\},$$

where φ is a bump function. This could be a subject for further investigation.

8. Interpolation between Lipschitz and BV

Let $Lip([a, b])$ be the space of Lipschitz functions on the interval $[a, b]$, $a < b$, and $BV([a, b])$ the space of functions of bounded variation on $[a, b]$. It is known that

Lemma 8.1.

(i) $Lip([a, b])$ *is a Banach space with respect to*

$$\|f\|_{Lip} = \|f\|_\infty \vee \sup_{\substack{x,y\in[a,b] \\ x\neq y}} \frac{|f(x) - f(y)|}{|x - y|}.$$

(ii) $BV([a,b])$ *is a Banach space with respect to*

$$\|f\|_{BV} = |f(a)| + V(f).$$

Remark 8.2. If we continuously extend a function $g \in BV([a,b])$ to be constant outside the interval $[a,b]$, then the extension is a function in $BV(\mathbb{R})$ defined in Definition 2.7, and has the same variation as g.

We also have $Lip([a,b]) \subset BV([a,b])$, so that the spaces $Lip([a,b])$ and $BV([a,b])$ form an interpolation couple. We may use the real interpolation method described in [5] to improve the convergence results for functions in the interpolation spaces $(Lip([a,b]), BV([a,b]))_{\theta,q}$ equipped with the norm $\|\cdot\|_{\theta,q}$. Let us first recall the definition.

Definition 8.3. The K-functional related to the spaces $Lip([a,b])$ and $BV([a,b])$ is

$$\begin{aligned} &K(g,t; Lip([a,b]), BV([a,b])) \\ &= \inf\{\|g_1\|_{Lip} + t\,\|g_2\|_{BV} : g = g_1 + g_2, g_1 \in Lip([a,b]), g_2 \in BV([a,b])\}. \end{aligned}$$

Then for $0 < \theta < 1$ and $1 \le q \le \infty$, the interpolation space is

$$(Lip([a,b]), BV([a,b]))_{\theta,q} = \{g \in BV([a,b]) : \|g\|_{\theta,q} < \infty\},$$

where for $1 \le q < \infty$ the norm is defined by

$$\|g\|_{\theta,q} = \left(\int_0^\infty \left[t^{-\theta} K(g,t; Lip([a,b]), BV([a,b]))\right]^q \frac{dt}{t}\right)^{\frac{1}{q}},$$

and for $q = \infty$,

$$\|g\|_{\theta,\infty} = \sup_{t>0} t^{-\theta} K(g,t; Lip([a,b]), BV([a,b])).$$

Theorem 8.4. *Let $1 \le p < \infty$, $0 < \theta < 1$ and $1 \le q \le \infty$. Suppose that X and $\hat{X}$ are random variables with values in $[a,b]$, and that X has a bounded density f_X. Then*

$$\left\|g(X) - g(\hat{X})\right\|_p \le \left(3^{1+1/p} (\sup f_X)^{\frac{1}{1+p}}\right)^\theta \|g\|_{\theta,q} \left\|X - \hat{X}\right\|_p^{1-\theta\left(1-\frac{1}{1+p}\right)}$$

for $g \in (Lip([a,b]), BV([a,b]))_{\theta,q}$.

Proof. By the Lipschitz property, if $g \in Lip([a,b])$ then

$$\left\|g(X) - g(\hat{X})\right\|_p \le \left\|X - \hat{X}\right\|_p \|g\|_{Lip}.$$

On the other hand, by [3, Theorem 2.4 (i) with $q = p$], if $g \in BV([a,b])$ then

$$\left\|g(X) - g(\hat{X})\right\|_p \le 3^{1+1/p} (\sup f_X)^{\frac{1}{1+p}} \left\|X - \hat{X}\right\|_p^{\frac{1}{1+p}} \|g\|_{BV}.$$

We define a linear operator $T : BV([a,b]) \to L_p$ by

$$Tg = g(X) - g(\hat{X}).$$

The results above show that T is bounded and thus an admissible operator. Then by the interpolation theorem [5, Theorem V.1.12] we get for $g \in (Lip([a,b]), BV([a,b]))_{\theta,q}$ that

$$\left\| g(X) - g(\hat{X}) \right\|_p \leq \left(3^{1+1/p} (\sup f_X)^{\frac{1}{1+p}} \right)^{\theta} \left\| X - \hat{X} \right\|_p^{1-\theta+\frac{\theta}{1+p}} \|g\|_{\theta,q},$$

which gives the assertion. □

Theorem 8.5. *Let* $0 < \alpha < 1$ *and* $g : [0,1] \to \mathbb{R}$, $g(x) = x^\alpha$. *Then* $g \in (Lip([0,1]), BV([0,1]))_{1-\alpha,\infty}$.

Proof. We simplify the notation of the K-functional by omitting the spaces. By definition,

$$\|g\|_{(Lip([0,1]),BV([0,1]))_{\theta,\infty}} = \sup_{t>0} t^{-\theta} K(g,t)$$

for $0 < \theta < 1$. Then by choosing $g_2 = g \in BV([0,1])$, we have

$$K(g,t) \leq t \|g\|_{BV},$$

so that

$$\sup_{0<t\leq 1} t^{-\theta} K(g,t) \leq \|g\|_{BV} \sup_{0<t\leq 1} t^{1-\theta} < \infty$$

for all $0 < \theta < 1$. Therefore to conclude that g is in the interpolation space $(Lip([0,1]), BV([0,1]))_{1-\alpha,\infty}$, we need to show that

$$\sup_{t>1} t^{-(1-\alpha)} K(g,t) < \infty.$$

Let $x_0 \in (0,1]$. We can write $g = g_1 + g_2$, where

$$g_1(x) = \begin{cases} 0, & 0 \leq x \leq x_0 \\ g(x) - g(x_0), & x_0 < x \leq 1 \end{cases}$$

and

$$g_2(x) = \begin{cases} g(x), & 0 \leq x \leq x_0 \\ g(x_0), & x_0 < x \leq 1. \end{cases}$$

Obviously $g_1 \in Lip([0,1])$ and $g_2 \in BV([0,1])$. Now

$$\begin{aligned} \|g_1\|_{Lip} + t\|g_2\|_{BV} &= \|g_1\|_\infty \vee \sup_{\substack{x,y\in[0,1] \\ x\neq y}} \frac{|g_1(x) - g_1(y)|}{|x-y|} + t(g_2(0) + V(g_2)) \\ &= (1 - x_0^\alpha) \vee \alpha x_0^{\alpha-1} + t x_0^\alpha. \end{aligned}$$

There exists $m_\alpha \in (0,1)$ such that $\alpha x_0^{\alpha-1} > (1-x_0^\alpha)$ for all $x_0 \in (0, m_\alpha)$. In particular, for $t > 1$ we can choose $x_0 = m_\alpha/t \in (0, m_\alpha)$, so that

$$\begin{aligned} \|g_1\|_{Lip} + t\|g_2\|_{BV} &= \alpha x_0^{\alpha-1} + t x_0^\alpha = x_0^{\alpha-1}(\alpha + t x_0) \\ &= (m_\alpha/t)^{\alpha-1}(\alpha + m_\alpha). \end{aligned}$$

This implies

$$K(g,t) \leq \frac{\alpha + m_\alpha}{m_\alpha^{1-\alpha}} t^{1-\alpha}$$

for $t > 1$, and thus

$$\sup_{t>1} t^{-(1-\alpha)} K(g,t) \leq \frac{\alpha + m_\alpha}{m_\alpha^{1-\alpha}} < \infty.$$

□

Theorem 8.6. *Let $1 \leq p < \infty$. The convergence rate $1 - \theta \left(1 - \frac{1}{1+p}\right)$ in Theorem 8.4 is optimal, i.e., if*

$$\left\| g(X) - g(\hat{X}) \right\|_p \leq C(p, X, \theta, q, g) \left\| X - \hat{X} \right\|_p^r$$

for all random variables, parameter values, and functions g considered in Theorem 8.4, then $r \leq 1 - \theta \left(1 - \frac{1}{1+p}\right)$.

Proof. Suppose that $\Omega = [0,1]$, $\mathbb{P}$ is the Lebesgue measure on $[0,1]$, and $0 < \gamma < 1$. We choose the random variables $X(x) = x$ and

$$\widehat{X}(x) = \begin{cases} 0, & x \leq \gamma, \\ x, & x > \gamma \end{cases}$$

for $x \in \Omega$. Note that X has a bounded density. We know that $g(x) = x^\alpha$ is in the interpolation space $(Lip([0,1]), BV([0,1]))_{1-\alpha,\infty}$ by Theorem 8.5. Now we have

$$\left\| g(X) - g(\widehat{X}) \right\|_p^p = \int_0^\gamma x^{\alpha p}\, dx = \frac{\gamma^{\alpha p+1}}{\alpha p + 1}$$

and

$$\left\| X - \widehat{X} \right\|_p^p = \int_0^\gamma x^p\, dx = \frac{\gamma^{p+1}}{p+1}.$$

This shows that the exponent of γ on the left-hand side of Theorem 8.4 is $(\alpha p + 1)/p$, whereas on the right-hand side it is, since $\theta = 1 - \alpha$,

$$\frac{p+1}{p}\left(1 - (1-\alpha)\left(1 - \frac{1}{1+p}\right)\right) = \frac{1+\alpha p}{p}.$$

Since the exponents coincide, the rate is optimal. □

Remark 8.7. Theorems 8.4 and 8.6 also imply that the parameter $1-\alpha$ in Theorem 8.5 is optimal, i.e. it cannot be decreased.

9. Polynomial variation and BV_φ

We show that if the random variables X and $\hat{X}$ are in L_p for all $1 \leq p < \infty$, then the function $\varphi^{X,\hat{X}}$ can be chosen in such a way that it does not depend on $\hat{X}$, and it decays faster than any polynomial, i.e. $\varphi^{X,\hat{X}}(x) = o(x^{-q})$ for all $1 \leq q < \infty$. If φ satisfies this property, then we show that BV_φ contains all functions with polynomial variation.

Lemma 9.1. *Suppose that X and $\hat{X}$ are random variables such that $X \in \bigcap_{p\in[1,\infty)} L_p$, and suppose there exists $C = (C_p)_{p\in[1,\infty)} \subset (0,\infty)$ such that $\left\|X - \hat{X}\right\|_p \leq C_p$ for all $p \in [1,\infty)$. Then we can choose the function $\varphi^{X,\hat{X}}$ such that $\varphi^{X,\hat{X}} = \varphi_C^X$, where the function φ_C^X is a bump function that decays faster than any polynomial.*

Proof. The triangle inequality gives that $\hat{X} \in L_p$ and

$$\left\|\hat{X}\right\|_p \leq \left\|X - \hat{X}\right\|_p + \|X\|_p \leq C_p + \|X\|_p .$$

Thus by Chebychev's inequality we have for all $\lambda > 0$ that

$$\mathbb{P}(|X| \geq \lambda) \leq \frac{\mathbb{E}|X|^p}{\lambda^p}$$

and

$$\mathbb{P}(|\hat{X}| \geq \lambda) \leq \frac{\mathbb{E}|\hat{X}|^p}{\lambda^p} \leq \frac{(C_p + \|X\|_p)^p}{\lambda^p}.$$

So we have a polynomial tail estimate for X and $\hat{X}$ that depends only on the constants C_p of the L_p-estimates, not directly on $\hat{X}$. This implies that

$$\mathbb{P}(|X| \geq \lambda) \vee \mathbb{P}(|\hat{X}| \geq \lambda) \leq \inf_{p\in\mathbb{N}} \frac{(C_p \vee p + \|X\|_p)^p}{\lambda^p} \wedge 1 =: \varphi_C^X(\lambda)$$

for $\lambda > 0$. For $\lambda < 0$, we define $\varphi_C^X(\lambda) := \varphi_C^X(|\lambda|)$ and $\varphi_C^X(0) := 1$. The function φ_C^X is continuous, because $C_p \vee p + \|X\|_p \to \infty$ as $p \to \infty$. Indeed, let $x_0 \in \mathbb{R}$. Then on the interval $[-|x_0| - 1, |x_0| + 1]$, only finitely many functions in the infimum contribute, i.e. are less than one. They are all continuous in λ, and the infimum over a finite number of continuous functions is continuous, in particular at x_0. By similar reasoning we see that φ_C^X is strictly positive. It also satisfies the monotonicity properties of a bump function, and by definition decays faster than any polynomial. □

Definition 9.2. For $\mu \in \mathcal{M}$, define

$$\mathcal{J}(\mu) = \{x \in \mathbb{R} \mid \mu(\{x\}) \neq 0\}.$$

Moreover, we denote the continuous part of μ by $\mu_c = \mu_{|(\mathbb{R}\setminus\mathcal{J}(\mu))}$ and the jump part of μ by $\mu_J = \mu_{|\mathcal{J}(\mu)}$.

Remark 9.3. If φ is a bump function and $\mu \in \mathcal{M}_\varphi$, Definition 9.2 gives a decomposition of any $g^\mu \in BV_\varphi$ into a continuous part g^{μ_c} and a jump part g^{μ_J}, and the set $\mathcal{J}(\mu)$ is countable.

Theorem 9.4. *Suppose that $\mu \in \mathcal{M}$, and φ is a bump function that decays faster than any polynomial. Then $g^\mu \in BV_\varphi$ if there exist constants $s, C > 0$ such that $d|\mu_c| \leq C(1 + |x|^s)\,dx$ and $\sum_{x\in J(\mu)} \varphi(x)|\mu(\{x\})| < \infty$.*

Proof. Since φ decays faster than any polynomial, we have

$$\varphi(x) \le \tilde{C}|x|^{-(s+2)} \wedge 1,$$

and

$$\begin{aligned}
\|g\|_{\varphi} &= \int_{\mathbb{R}} \varphi \, d|\mu| = \int_{\mathbb{R}} \varphi \, d|\mu_c + \mu_J| \\
&\le \int_{\mathbb{R}} \varphi(z) C(1 + |z|^s) \, dz + \int_{\mathbb{R}} \varphi(z) |\mu(z)| \, d\delta_{\mathcal{J}(\mu)}(z) \\
&\le C\tilde{C} \int_{\mathbb{R}} (|z|^{-(s+2)} \wedge 1)(1 + |z|^s) \, dz + \sum_{x \in J(\mu)} \varphi(x) |\mu(\{x\})| < \infty.
\end{aligned}$$

This implies that $g \in BV_{\varphi}$. □

Example 9.5. Suppose $g \in C^1$, $g(0) = 0$, and the derivative satisfies $|g'(x)| \le C(1 + |x|^s)$ for some $s > 0$. By the fundamental theorem of calculus we can write $g = g^{\mu_c}$, where $d\mu_c(x) = g'(x) \, dx$. If φ is a bump function that decays faster than any polynomial, then Theorem 9.4 implies that $g \in BV_{\varphi}$. Moreover, we can add jumps to function g by defining another signed measure $\mu = \mu_c + \mu_J$, where $\mu_J = \sum_{i=1}^{\infty} \alpha_i \delta_{\{x_i\}}$, with $\alpha_i, x_i \in \mathbb{R}$ and $x_i \ne x_j$ for $i \ne j$, satisfies the assumption of Theorem 9.4.

Example 9.6. For any $s \ge 1$, the function

$$g(x) = \sum_{k=0}^{\infty} k^s \chi_{]k,k+1]}(x)$$

is in BV_{φ}, where φ is a bump function that decays faster than any polynomial. Namely, if we define

$$\mu = \sum_{k=1}^{\infty} (k^s - (k-1)^s) \delta_{\{k\}},$$

then we see that $g = g^{\mu}$, $\mathcal{J}(\mu) = \mathbb{N}$, and

$$\sum_{x \in J(\mu)} \varphi(x) |\mu(\{x\})| \le \sum_{k=1}^{\infty} \tilde{C} k^{-(s+2)} (k^s - (k-1)^s) < \infty.$$

10. Stochastic Differential Equations

We recall the setting of [3], i.e. 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

$$\begin{cases} dX_t = \sigma(t, X_t) \, dW_t + b(t, X_t) \, dt, \\ X_0 = x_0 \end{cases} \tag{10.1}$$

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 $\alpha \geq \frac{1}{2}$ such that

(i) $|f(t,x) - f(t,y)| \leq C_T |x-y|$,

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

Assumptions (i) and (ii) imply the existence of a unique adapted strong solution X of the SDE (10.1), see e.g. [14]. Moreover, we assume that

(iii) X_T has a bounded density.

Remark 10.1. There are known sufficient conditions for the assumption (iii), e.g. uniform ellipticity of the SDE. See [3, Remark 4.1] for details.

Now we can formulate a Corollary corresponding to [3, Corollary 4.2] for the function class defined in Section 3.

Corollary 10.2. *Suppose that X is the solution of* (10.1)*, and X_T has a bounded density. Let $\gamma > 0$, and let X_T^π be an approximation of X_T such that, for all $1 \leq p < \infty$, there exist constants $C_p > 0$ with*

$$\|X_T - X_T^\pi\|_p \leq C_p \, |\pi|^\gamma .$$

Then for any $1 \leq q < \infty$ and $0 < \varepsilon < \gamma$, we have for $\theta = \frac{\varepsilon}{2\gamma - \varepsilon}$, $\varphi_C^{X_T}$ according to Lemma 9.1, and $g \in BV_{\left(\varphi_C^{X_T}\right)^{\theta/q}}$, that

$$\|g(X_T) - g(X_T^\pi)\|_q^q \leq 3 \left(C_{1/\theta} \sup f_{X_T}\right)^{1-\frac{\varepsilon}{\gamma}} \|g\|^q_{\left(\varphi_C^{X_T}\right)^{\frac{\theta}{q}}} |\pi|^{\gamma - \varepsilon} .$$

Proof. By [2, Lemma A.2] we have that $X_T \in \bigcap_{p \in [1,\infty)} L_p$, so by Lemma 9.1 we can choose $\varphi^{X_T, X_T^\pi} = \varphi_C^{X_T}$, where $\varphi_C^{X_T}$ is a bump function with decay faster than any polynomial. By [3, Lemma 3.4] we have

$$\left\|\chi_{[K,\infty)}(X_T) - \chi_{[K,\infty)}(X_T^\pi)\right\|_1 \leq 3 (\sup f_{X_T})^{\frac{p}{p+1}} \|X_T - X_T^\pi\|_p^{\frac{p}{p+1}} ,$$

so by Theorem 6.2, for any $p \in [1,\infty)$ and $\theta \in (0,1)$,

$$\begin{aligned}
&\|g(X_T) - g(X_T^\pi)\|_q^q \\
&\leq 2^\theta 3^{1-\theta} (\sup f_{X_T})^{\frac{p(1-\theta)}{p+1}} \|g\|^q_{(\varphi^{X_T, X_T^\pi})^{\frac{\theta}{q}}} \|X_T - X_T^\pi\|_p^{\frac{p(1-\theta)}{p+1}} \\
&\leq 3 (\sup f_{X_T})^{\frac{p(1-\theta)}{p+1}} \|g\|^q_{(\varphi_C^{X_T})^{\frac{\theta}{q}}} C_p^{\frac{p(1-\theta)}{p+1}} |\pi|^{\frac{\gamma p(1-\theta)}{p+1}} .
\end{aligned}$$

Let $0 < \varepsilon < \gamma$. Choose $p = \frac{2\gamma}{\varepsilon} - 1$ and let $\theta = 1/p$. Note that $p > 1$ since $\varepsilon < \gamma$. Then

$$\frac{p(1-\theta)}{p+1} = \frac{p-1}{p+1} = 1 - \frac{\varepsilon}{\gamma}$$

and thus for all $g \in BV_{\left(\varphi_C^{X_T}\right)^{\theta/q}}$,

$$\|g(X_T) - g(X_T^\pi)\|_q^q \leq 3 \left(C_{1/\theta} \sup f_{X_T}\right)^{1-\frac{\varepsilon}{\gamma}} \|g\|^q_{(\varphi_C^{X_T})^{\frac{\theta}{q}}} |\pi|^{\gamma - \varepsilon} .$$

□

Remark 10.3. As the bump function $\varphi_C^{X_T}$ decays faster than any polynomial, Theorem 9.4 implies that Corollary 10.2 is valid for functions with polynomial variation.

Remark 10.4. In Corollary 10.2 the function φ^{X_T,X_T^π} depends on the distribution of X_T^π and is replaced by the uniform bound $\varphi_C^{X_T}$. However, when considering the convergence rate we are looking at partitions with small mesh size. If approximating random variables X_T^π corresponding to partitions with large mesh size had heavy tailed distributions, the use of the uniform bound could unnecessarily narrow down the class of functions. Therefore in such a case it would be better to take a more delicate approach and study the result

$$\|g(X_T) - g(X_T^\pi)\|_q^q \leq 3\left(C_{1/\theta} \sup f_{X_T}\right)^{1-\frac{\varepsilon}{\gamma}} \|g\|^q_{\left(\varphi^{X_T,X_T^\pi}\right)^{\frac{\theta}{q}}} |\pi|^{\gamma-\varepsilon}.$$

11. Euler Scheme

In the case of the Euler scheme we use specific moment estimates to improve the result of Lemma 9.1 for the decay of the function φ^{X_T,X_T^E}. We now bound it from above by explicit bump functions that do not depend on X_T^E. Before showing this in the main result of this section, Theorem 11.4, let us recall the definition and a classical moment inequality.

Definition 11.1 (Euler scheme)**.** Let X^E be the Euler scheme relative to π, i.e. $X_0^E = x_0$, and for $i = 0, \ldots, n-1$,

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

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

Lemma 11.2. *If the assumptions (i) and (ii) in Section 10 hold, and* $1 \leq p < \infty$*, then there exists* $M(x_0, T, C_T, \alpha) > 0$ *such that*

$$\left\| \sup_{t \leq T} |X_t^E| \right\|_p \leq e^{Mp}$$

and

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

Proof. The result is proved in [7, pp. 275-276] without writing explicitly the dependence of the upper bound on p. We get the explicit constant using the proof of [2, Theorem A.1] and [2, Lemma A.2], and the optimal constant in the Burkholder–Davis–Gundy inequality given in [4]. □

Remark 11.3. In [3] we used a moment estimate similar to Lemma 11.2, but with the constant e^{Mp^2}. As kindly remarked by Andreas Neuenkirch, we can drop the square in the constant by using the results in [4]. Consequently, the power of the logarithm in the convergence rate given in [3, Theorem 5.4] can be slightly improved.

Theorem 11.4. *We may 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. $|\sigma|, |b| < M_B$, we have for $z_0 = |x_0| + M_B T$ that*

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

(ii) *if the functions σ and b are Lipschitz, then for $z_0 = e^{3M}$ we have*

$$\varphi_E^{X_T}(z) = \begin{cases} |z|^{-\frac{2}{9M}\log(1+|z|-z_0)} & \text{if } |z| > z_0, \\ 1 & \text{if } |z| \leq z_0, \end{cases}$$

where $M = M(x_0, T, C_T, \alpha) > 0$.

Proof. (i) We consider the Euler approximation with n time nodes in the integral form

$$\begin{aligned} X_t^E &= x_0 + \int_0^t \sum_{k=0}^{n-1} \sigma(t_k, X_{t_k}^E)\chi_{(t_k,t_{k+1}]}(s)\, dW_s \\ &\quad + \int_0^t \sum_{k=0}^{n-1} b(t_k, X_{t_k}^E)\chi_{(t_k,t_{k+1}]}(s)\, ds,\ t \in [0,T] \text{ a.s.} \end{aligned}$$

Following the techniques used in [13] and [15], let us denote

$$L_u := \sum_{k=0}^{n-1} \sigma(t_k, X_{t_k}^E)\chi_{(t_k,t_{k+1}]}(u).$$

Then by the boundedness of σ and the Novikov condition,

$$M_t := e^{\alpha \int_0^t L_u\, dW_u - \frac{\alpha^2}{2}\int_0^t L_u^2\, du}$$

is a martingale for any $\alpha > 0$, and $\mathbb{E}M_t = 1$. Thus by Chebychev's inequality we have for $\lambda > 1$ that

$$\mathbb{P}\left(e^{\alpha \int_0^T L_u\, dW_u - \frac{\alpha^2}{2}\int_0^T L_u^2\, du} \geq \lambda\right) \leq \frac{1}{\lambda}.$$

Taking logarithms shows that

$$\mathbb{P}\left(\alpha \int_0^T L_u\,dW_u - \frac{\alpha^2}{2}\int_0^T L_u^2\,du \geq \lambda\right) \leq e^{-\lambda}$$

for $\lambda > 0$. Since

$$\int_0^T L_u^2\,du \leq M_B^2 T,$$

we get

$$\mathbb{P}\left(\int_0^T L_u\,dW_u \geq \frac{\lambda}{\alpha} + \frac{\alpha M_B^2 T}{2}\right) \leq e^{-\lambda},$$

which we can reparametrize as

$$\mathbb{P}\left(\int_0^T L_u\,dW_u \geq \lambda\right) \leq e^{\frac{\alpha^2 M_B^2 T}{2} - \lambda\alpha},$$

when $\lambda > \alpha M_B^2 T/2$. Now we may choose $\alpha = \lambda/(M_B^2 T)$ to get

$$\mathbb{P}\left(\int_0^T L_u\,dW_u \geq \lambda\right) \leq e^{-\frac{\lambda^2}{2M_B^2 T}}$$

for $\lambda > 0$. A similar proof with $\widetilde{L}_u = -L_u$ shows that

$$\mathbb{P}\left(\int_0^T L_u\,dW_u \leq \lambda\right) \leq e^{-\frac{\lambda^2}{2M_B^2 T}}$$

for $\lambda < 0$. Therefore, for $\lambda > x_0 + M_B T$,

$$\begin{aligned}\mathbb{P}\left(X_T^E \geq \lambda\right) &\leq \mathbb{P}\left(x_0 + \int_0^T L_u\,dW_u + M_B T \geq \lambda\right)\\ &\leq e^{-\frac{(\lambda-(x_0+M_B T))^2}{2M_B^2 T}},\end{aligned}$$

and for $\lambda < x_0 - M_B T$,

$$\begin{aligned}\mathbb{P}\left(X_T^E \leq \lambda\right) &\leq \mathbb{P}\left(x_0 + \int_0^T L_u\,dW_u - M_B T \leq \lambda\right)\\ &\leq e^{-\frac{(\lambda-(x_0-M_B T))^2}{2M_B^2 T}}.\end{aligned}$$

Obviously a similar proof works for the random variable X_T instead of X_T^E, so we get an upper bound for φ^{X_T, X_T^E}. Moreover, we choose the upper bound to be one on the interval $[x_0 - M_B T, x_0 + M_B T]$ to get that

$$\varphi_E^{X_T}(z) = \begin{cases} e^{-\frac{(z-(x_0+M_B T))^2}{2M_B^2 T}} & \text{if } z > x_0 + M_B T,\\ e^{-\frac{(z-(x_0-M_B T))^2}{2M_B^2 T}} & \text{if } z < x_0 - M_B T,\\ 1 & \text{elsewhere.}\end{cases}$$

By extending the set where $\varphi_E^{X_T}(z) = 1$ to $|z| < |x_0| + M_B T =: z_0$ and making the corresponding shift in the function gives the assertion.

(ii) If σ and b are Lipschitz, then we know from Lemma 11.2 that

$$\left\|X_T^E\right\|_p \leq e^{Mp},$$

where the constant $M > 0$ depends on x_0, T, C_T, and α. Now by Chebychev's inequality we have for $\lambda > 0$ that

$$\mathbb{P}(|X_T^E| \geq \lambda) \leq \frac{\mathbb{E}|X_T^E|^p}{\lambda^p} \leq \frac{e^{Mp^2}}{\lambda^p}.$$

Choose $3Mp = \log \lambda$ for $\lambda \geq \lambda_0 = e^{3M}$. This gives

$$p = \frac{\log \lambda}{3M},$$

and thus for $\lambda \geq \lambda_0$ we get

$$\mathbb{P}(|X_T^E| \geq \lambda) = \frac{e^{\frac{1}{3}p \log \lambda}}{\lambda^p} = \lambda^{-\frac{2}{3}p} = \lambda^{-\frac{2}{9M} \log \lambda}.$$

Again the same proof works for the term $\mathbb{P}(|X_T| \geq \lambda)$ because of [2, Lemma A.2]. Thus we get an upper bound

$$\varphi_E^{X_T}(z) = \begin{cases} |z|^{-\frac{2}{9M} \log |z|} & \text{if } |z| > z_0, \\ 1 & \text{if } |z| \leq z_0, \end{cases}$$

where for $z_0 = e^{3M}$. To get a bump function, we again adjust the function to be continuous by making a shift in the exponent. □

Example 11.5. Let $c > 0$, $0 < \theta < 1$, $0 < q < \infty$, and suppose that the functions σ and b are bounded. Then Theorem 11.4 implies that the functions

$$g_1(x) = \sum_{k=0}^{\infty} e^{ck^\gamma} \chi_{]k,k+1]}(x)$$

and

$$g_2(x) = e^{c|x|^\gamma} - 1$$

are in $BV_{\left(\varphi_E^{X_T}\right)^{\frac{\theta}{q}}}$ for any $0 < \gamma < 2$. Indeed, g_1 can be represented using the measure

$$\mu_1 = \delta_{\{0\}} + \sum_{k=1}^{\infty} (e^{ck^\gamma} - e^{c(k-1)^\gamma}) \delta_{\{k\}},$$

which satisfies $g = g^{\mu_1}$ and

$$\|g^{\mu_1}\|_{\left(\varphi_E^{X_T}\right)^{\frac{\theta}{q}}} \leq \sum_{k=0}^{\infty} e^{ck^\gamma} \left(e^{-\frac{(k-z_0)^2 \theta}{2M_B^2 T q}} \wedge 1 \right) < \infty.$$

Since g_2 is not differentiable at zero, define $\tilde{g}(x) := g_2'(x)$ if $x \neq 0$ and $\tilde{g}(0) := 0$. By choosing a signed measure $d\mu(z) = \tilde{g}(z)\, dz$, we see that the representation $g_2 = g^\mu$ holds, and

$$\|\mu\|_{\left(\varphi_E^{X_T}\right)^{\frac{\theta}{q}}} = \int_{\mathbb{R}} \left(\varphi_E^{X_T}(z)\right)^{\frac{\theta}{q}} |\tilde{g}(z)|\, dz$$

$$= \int_{\mathbb{R}} \left(e^{-\frac{(|z|-z_0)^2 \theta}{2M_B^2 Tq}} \wedge 1 \right) e^{c|z|^\gamma} c\gamma |z|^{\gamma-1} \, dz < \infty,$$

because the singularity at zero for $0 < \gamma < 1$ is not too strong.

Theorem 11.6. *Suppose that the coefficients σ and b of the SDE* (10.1) *are bounded, and X_T has a bounded density. Let $g : \mathbb{R} \to \mathbb{R}$ be a function with a representation $g = g^\mu$, where $\mu \in \mathcal{M}$ such that there exists $s \in \{0, 1, 2, \dots\}$ with*

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

for all bump functions φ. Then for any $1 \le q < \infty$ there exists $m \in (0,1)$ such that

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

for $|\pi| < m$ and $M > 0$ taken from Lemma 11.2.

Proof. By Corollary 10.2, Lemma 11.2, and Lemma 3.11, for $0 < \varepsilon < 1/2 = \gamma$, $C_p = e^{Mp^2}$, and $g \in BV_{\left(\varphi_E^{X_T}\right)^{\varepsilon/q}}$, it holds that

$$\begin{aligned} \left\| g(X_T) - g(X_T^E) \right\|_q^q &\le 3 \left(e^{M\left(\frac{1-\varepsilon}{\varepsilon}\right)^2} \sup f_{X_T} \right)^{1-2\varepsilon} \|g\|^q_{\left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q(1-\varepsilon)}}} |\pi|^{\frac{1}{2}-\varepsilon} \\ &\le 3 \left(e^{\frac{M}{\varepsilon^2}} \sup f_{X_T} \vee 1 \right) \|g\|^q_{\left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}}} |\pi|^{\frac{1}{2}-\varepsilon}, \end{aligned} \tag{11.1}$$

since $\frac{\varepsilon}{q(1-\varepsilon)} > \frac{\varepsilon}{q}$ implies $\left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q(1-\varepsilon)}} \le \left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}}$. Now we choose

$$\varepsilon = (-\log|\pi|)^{-1/3}$$

for $|\pi| < e^{-8}$. Then

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

and obviously

$$e^{\frac{M}{\varepsilon^2}} \sup f_{X_T} \vee 1 \le e^{\frac{M}{\varepsilon^2}} (\sup f_{X_T} \vee 1).$$

Hence

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

Let us write $\varphi^{|\pi|} := \left(\varphi_E^{X_T}\right)^{\frac{1}{q(-\log|\pi|)^{1/3}}} = \left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}}$ for convenience. By Lemma 11.7 we have

$$\|g\|_{\varphi^{|\pi|}} \le C\varepsilon^{-r} = C(-\log|\pi|)^{r/3}$$

for $r = (1+s)/2$ and $C = C(x_0, s, M_B, T, q)$, with M_B from Theorem 11.4 (i), and thus $g \in BV_{\varphi^{|\pi|}}$ for any mesh size $|\pi| < e^{-8}$. Moreover, for any $1 \le q < \infty$ there exists $m = m(x_0, s, M_B, T, q) > 0$ such that

$$\begin{aligned} \|g\|^q_{\varphi^{|\pi|}} |\pi|^{\frac{1}{(-\log|\pi|)^{1/3}}} &\le C^q e^{\log(-\log|\pi|)^{\frac{qr}{3}} + \frac{\log|\pi|}{(-\log|\pi|)^{1/3}}} \\ &\le C^q e^{\frac{qr}{3}\log(-\log|\pi|) - (-\log|\pi|)^{2/3}} \\ &\le 1 \end{aligned}$$

for $|\pi| < m$, and we get

$$\begin{aligned} \|g\|^q_{\varphi^{|\pi|}} |\pi|^{\frac{1}{2} - \frac{1+M}{(-\log|\pi|)^{1/3}}} &\le \|g\|^q_{\varphi^{|\pi|}} |\pi|^{\frac{1}{(-\log|\pi|)^{1/3}}} |\pi|^{\frac{1}{2} - \frac{2+M}{(-\log|\pi|)^{1/3}}} \\ &\le |\pi|^{\frac{1}{2} - \frac{2+M}{(-\log|\pi|)^{1/3}}} \end{aligned}$$

for $|\pi| < m$, which proves the statement. □

Lemma 11.7. *Suppose that g is a function satisfying the assumption of Theorem 11.6. Then $g \in BV_{\left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}}}$ for any $0 < \varepsilon < 1$ and $q > 0$, and*

$$\|g\|_{\left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}}} \le C\varepsilon^{-r}$$

for $r = (1+s)/2$ and a constant $C = C(x_0, s, M_B, T, q) > 0$.

Proof. By assumption the function g can be written using a measure $\mu \in \mathcal{M}$ such that

$$\|g\|_{\left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}}} = \int_{\mathbb{R}} \left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}} d|\mu| \le \int_{\mathbb{R}} \left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}} (x)|x|^s \, dx,$$

where $s \in \{0, 1, 2, \dots\}$, and by symmetry it is sufficient to integrate over the positive reals. Now by Theorem 11.4 (i) we get

$$\int_0^\infty \left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}} (x) x^s \, dx = \int_0^{z_0} x^s \, dx + \int_{z_0}^\infty e^{-\frac{(x-z_0)^2\varepsilon}{2M_B^2 Tq}} x^s \, dx,$$

where the integral from 0 to z_0 is finite. If $s \ge 1$, then for the other integral we get

$$\begin{aligned} &\int_{z_0}^\infty e^{-\frac{(x-z_0)^2\varepsilon}{2M_B^2 Tq}} x^s \, dx = \int_0^\infty e^{-\frac{x^2\varepsilon}{2M_B^2 Tq}} (x+z_0)^s \, dx \\ &\le 2^{s-1} \int_0^\infty e^{-\frac{x^2\varepsilon}{2M_B^2 Tq}} (x^s + z_0^s) \, dx \\ &= 2^{s-1} \left(\frac{1}{\varepsilon^{\frac{1+s}{2}}} \int_0^\infty e^{-\frac{x^2}{2M_B^2 Tq}} x^s \, dx + \frac{z_0^s}{\varepsilon^{\frac{1}{2}}} \int_0^\infty e^{-\frac{x^2}{2M_B^2 Tq}} \, dx \right) \\ &\le \widetilde{C}(z_0, s, M_B, T, q)\varepsilon^{-\frac{1+s}{2}}. \end{aligned}$$

The case $s = 0$ is similar. Thus we have

$$\|g\|_{\left(\varphi_E^{X_T}\right)^{\frac{\varepsilon}{q}}} \le C\varepsilon^{-r}$$

for $r = (1+s)/2$ and $C = C(x_0, s, M_B, T, q)$, as $z_0 = z_0(x_0, M_B, T)$. $\square$

12. Application to the Multilevel Monte Carlo Method

We can directly apply the results of this paper to extend the results of [3, Section 6] concerning the multilevel Monte Carlo method of Giles [12], with the Euler scheme as the underlying discretization. Take $M \geq 2$ and $L \geq 0$, and consider timesteps $h_l = T/M^l$ with $0 \leq l \leq L$. 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. Then 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

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

and each of the summands $\mathbb{E}[g(X_T^{E,h_l}) - g(X_T^{E,h_{l-1}})]$ with

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

where for each i 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. By construction, the estimators $\widehat{Y}_l$ are independent. Then we approximate $\mathbb{E}g(X_T)$ by the combined estimator

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

The complexity of the multilevel method is given in [12, Theorem 3.1] in terms of two parameters, a weak convergence parameter α, and a variance parameter β. The latter can be deduced from our strong convergence results. Let us choose $T = 1$ for simplicity. As an immediate consequence of Corollary 10.2 for $q = 2$ and Theorem 11.4 we get the following:

Corollary 12.1. *Let $0 < \varepsilon < 1/2$. Then for $\theta = \varepsilon/(1-\varepsilon)$ and $g \in BV_{\left(\varphi_E^{X_T}\right)^{\theta/2}}$, the variance parameter β in [12, Theorem 3.1] satisfies $\beta = 1/2 - \varepsilon$.*

In the setting of Theorem 11.6, we show that the variance property (iii') in [3, Section 6] is again satisfied, and thus the result [3, Theorem 6.1] extends from functions of bounded variation to functions of polynomial variation.

Corollary 12.2. *Suppose g is a function satisfying the assumption of Theorem 11.6. Then the variance of the multilevel estimator $\widehat{Y}_l$ satisfies*

$$Var(\widehat{Y}_l) \le cN_l^{-1} M^{-\frac{l}{2}+\frac{A\,l}{((l\log M)\vee B)^{1/3}}}$$

for $l = 0, 1, 2, \dots,$ where $c, A, B > 0$ are constants independent of l.

For the convenience of the reader, we recall the proof of [3, Theorem 6.1] with minor modifications caused by the extension.

Proof of Corollary 12.2. Let $1 \le p < \infty$ and $T = 1$. By Theorem 11.6, there exists a constant $m = m(x_0, g, M_B, T, p) \in (0,1)$, where the parameter M_B is from Theorem 11.4 (i), such that

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

for $|\pi| < m$. On the other hand, Corollary 10.2 applied for the Euler scheme implies that, for any $0 < \delta < 1/2$,

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

for all mesh sizes $|\pi| > 0$. Note that the assumption $g \in BV_{\left(\varphi_E^{X_T}\right)^{\theta/q}}$ in Corollary 10.2, with $\theta = \delta/(1-\delta)$, is satisfied by Lemma 11.7. We choose

$$\delta = \frac{C_2(x_0,T,C_T,\alpha)}{(-\log m)^{1/3}}.$$

As $|\pi| \le m$ implies $-\log|\pi| \ge -\log m =: m_0$, this implies that for all mesh sizes $|\pi| > 0$,

$$\left\| g(X_T) - g(X_T^E) \right\|_p^p \le C_5(p,T,X,g,x_0,C_T,\alpha,M_B)\,|\pi|^{\frac{1}{2}-\frac{C_2(x_0,T,C_T,\alpha)}{(-\log|\pi|\vee m_0)^{1/3}}}.$$

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

$$\left\| g(X_T) - g(X_T^E) \right\|_p^p \le C_5(M^{-l})^{\frac{1}{2}-\frac{C_2(x_0,T,C_T,\alpha)}{(l\log M\vee m_0)^{1/3}}} =: \psi(l). \tag{12.1}$$

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

$$V_l \;=\; V(\widehat{P}_l - \widehat{P}_{l-1}) \le \left(\sqrt{V(\widehat{P}_l - P)} + \sqrt{V(\widehat{P}_{l-1} - P)}\right)^2,$$

where both of the variance terms on the right hand side can be bounded from above by $\psi(l)$. First,

$$V(\widehat{P}_l - P) \le \mathbb{E}(\widehat{P}_l - P)^2 \le \psi(l),$$

where we apply the result (12.1) for $p = 2$. Similarly, $V(\widehat{P}_{l-1} - P) \le \psi(l-1)$, but here we would like to have $\psi(l)$ instead of $\psi(l-1)$. Now

$$\psi(l-1) \;=\; C_5(M^{-l+1})^{\frac{1}{2}-\frac{C_2}{((l-1)\log M\vee m_0)^{1/3}}}$$

$$\begin{aligned}
&= C_5(M^{-l})^{\frac{1}{2}-\frac{C_2}{((l-1)\log M\vee m_0)^{1/3}}} M^{\frac{1}{2}-\frac{C_2}{((l-1)\log M\vee m_0)^{1/3}}} \\
&\leq C_5(M^{-l})^{\frac{1}{2}-\frac{C_2}{((l\log M-\log M)\vee m_0)^{1/3}}} \cdot M \\
&\leq C_6(p,T,X,g,x_0,C_T,\alpha,M_B,M)(M^{-l})^{\frac{1}{2}-\frac{C_2}{\left(l\frac{\log M}{2}\vee m_0\right)^{1/3}}},
\end{aligned}$$

where the last inequality follows from the fact that for $l \geq 2$,

$$l\log M - \log M \geq l\frac{\log M}{2},$$

and for $l = 1$ we can increase the constant C_6 if $(\log M)/2 \geq m_0$, and otherwise we could use m_0 in the estimate. Collecting the above results, we get that

$$V_l \leq C_7(p,T,X,g,x_0,C_T,\alpha,M_B,M)(M^{-l})^{\frac{1}{2}-\frac{C_2}{\left(l\frac{\log M}{2}\vee m_0\right)^{1/3}}}.$$

Note that by adjusting the constant C_7, the term $V_0 := V(\widehat{P}_0)$ also satisfies the above estimate. Indeed, we have

$$V(\widehat{P}_0) \leq \mathbb{E}\widehat{P}_0^2 = \mathbb{E}g(X_T^{E,h_0})^2 = \mathbb{E}g(x_0+\sigma(0,x_0)W_T+b(0,x_0)T)^2.$$

Thus it suffices to show that $\mathbb{E}g(c_1+c_2W_1)^2 < \infty$ for $c_1, c_2 \in \mathbb{R}$. Now,

$$\begin{aligned}
\mathbb{E}g(c_1+c_2W_1)^2 &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} g(c_1+c_2x)^2 e^{-\frac{x^2}{2}}\,dx \\
&= \frac{1}{c_2\sqrt{2\pi}}\int_{\mathbb{R}} g(x)^2 e^{-\frac{(x-c_1)^2}{2c_2^2}}\,dx \\
&= \frac{1}{c_2\sqrt{2\pi}}\int_{\mathbb{R}} \left(g(x)e^{-\frac{(x-c_1)^2}{6c_2^2}}\right)^2 e^{-\frac{(x-c_1)^2}{6c_2^2}}\,dx.
\end{aligned}$$

For $x > 0$ we get

$$\begin{aligned}
\left|g(x)e^{-\frac{(x-c_1)^2}{6c_2^2}}\right| &= \left|\int_{[0,x)} d\mu(z)\; e^{-\frac{(x-c_1)^2}{6c_2^2}}\right| \\
&\leq \int_{[0,x)} e^{-\frac{(x-c_1)^2}{6c_2^2}}\, d|\mu|(z)
\end{aligned}$$

The function $e^{-\frac{(x-c_1)^2}{6c_2^2}}$ can be bounded from above by a bump function $\tilde{\Phi}(x)$ that has exponential decay, and

$$\begin{aligned}
\int_{[0,x)} e^{-\frac{(x-c_1)^2}{6c_2^2}}\, d|\mu|(z) &\leq \int_{[0,x)} \tilde{\Phi}(x)\, d|\mu|(z) \\
&\leq \int_{[0,x)} \tilde{\Phi}(z)\, d|\mu|(z) \\
&\leq \int_{[0,\infty)} \tilde{\Phi}(z)\, d|\mu|(z).
\end{aligned}$$

The proof for $x \le 0$ is similar. Therefore, by assumption we have for some $s \in \{0, 1, 2, \dots\}$ that

$$\left| g(x) e^{-\frac{(x-c_1)^2}{6c_2^2}} \right| \le \int_{\mathbb{R}} \tilde{\Phi}(z)\, d|\mu|(z) \le \int_{\mathbb{R}} \tilde{\Phi}(z) |z|^s \, dz < \infty,$$

which immediately implies that $\mathbb{E} g(c_1 + c_2 W_1)^2 < \infty$.

Returning to our variance estimate, we have

$$V(\widehat{Y}_l) = N_l^{-1} V_l \le C_7 N_l^{-1} \left(M^{-l} \right)^{\frac{1}{2} - \frac{C_2}{\left(l \frac{\log M}{2} \vee m_0 \right)^{1/3}}}.$$

We adjust the constants to get the statement. □

References

[1] Apostol, T. M. *Mathematical Analysis*, 2nd ed. Addison-Wesley, 1974.

[2] Avikainen, R. Convergence rates for approximations of functionals of SDEs. *arXiv:0712.3635v1 [math.PR]* (2007).

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

[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] Bennett, C., and Sharpley, R. *Interpolation of Operators.* Academic Press, New York, 1988.

[6] Billingsley, P. *Probability and Measure.* Wiley, 1970.

[7] Bouleau, N., and Lépingle, D. *Numerical Methods for Stochastic Processes.* Wiley, New York, 1994.

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

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

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

[11] Giles, M. B. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, A. Keller, S. Heinrich, and H. Niederreiter, Eds. Springer, Berlin, 2008, pp. 343 – 358.

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

[13] Hitczenko, P. Upper bounds for the L_p-norms of martingales. *Probab. Theory Related Fields 86*, 2 (1990), 225.

[14] Karatzas, I., and Shreve, S. E. *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, Berlin, 1991.

[15] Ledoux, M., and Talagrand, M. *Probability in Banach Spaces.* Springer, 1991.

[16] Reed, M., and Simon, B. *Functional Analysis.* Academic Press, 1980.

[17] Rudin, W. *Real and Complex Analysis*, 2nd ed. McGraw-Hill, New York, 1966, 1974.

[18] Rudin, W. *Real and Complex Analysis*, 3rd ed. McGraw-Hill, New York, 1987.

Department of Mathematics and Statistics, P.O. Box 35 (MaD), FIN-40014 University of Jyväskylä, Finland
E-mail address: rainer.avikainen@jyu.fi