



**HAL**  
open science

# Weak approximation errors for stochastic differential equations with non-regular drift

Arturo Kohatsu-Higa, Antoine Lejay, Kazuhiro Yasuda

► **To cite this version:**

Arturo Kohatsu-Higa, Antoine Lejay, Kazuhiro Yasuda. Weak approximation errors for stochastic differential equations with non-regular drift. 2015. hal-00840211v2

**HAL Id: hal-00840211**

**<https://inria.hal.science/hal-00840211v2>**

Preprint submitted on 29 May 2015 (v2), last revised 6 Jul 2017 (v5)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Weak rate of convergence of the Euler-Maruyama scheme for stochastic differential equations with non-regular drift

Arturo Kohatsu-Higa\*

Antoine Lejay<sup>†</sup>

Kazuhiro Yasuda<sup>‡</sup>

## Abstract

We consider an Euler-Maruyama type approximation method for a stochastic differential equation (SDE) with a discontinuous drift and regular diffusion coefficient. The method regularizes the drift coefficient within a certain class of functions and then the Euler-Maruyama scheme for the regularized scheme is used as an approximation. This methodology gives two errors, the first is the error of regularization of the drift coefficient within a given class of parametrized functions and the second is the error of the regularized Euler-Maruyama scheme. After an optimization procedure with respect to this parameter we obtain various rates, which improve other known results.

**Keywords.** Stochastic differential equation, Euler-Maruyama scheme, discontinuous drift, weak rate of convergence, Malliavin calculus.

**MSC (2010):** Primary 65C30, 60H10

## 1 Introduction

The Euler-Maruyama scheme is a simple and efficient numerical scheme to simulate solutions of multi-dimensional stochastic differential equations (SDE's) such as

$$X_t = x + \int_0^t \sigma(s, X_s) dB_s + \int_0^t b(s, X_s) ds, \quad (1)$$

where  $B$  is a multi-dimensional Brownian motion. In many situations, one is interested in computing quantities of the type  $\mathbb{E}[f(X_T)]$  for some  $T > 0$  and  $f \in \mathfrak{F}$  where  $\mathfrak{F}$  is a class of functions. For a fixed number  $n$  of steps, consider  $n$  independent Gaussian random vectors  $\xi_1, \dots, \xi_n$  with zero mean and identity variance matrix. The Euler-Maruyama scheme consists in computing iteratively for  $k = 0, \dots, n-1$  with  $t_k = \frac{kT}{n}$ ,

$$\bar{X}_0 = x, \quad \bar{X}_{t_{k+1}} = \bar{X}_{t_k} + \sigma(t_k, \bar{X}_{t_k}) \xi_k \sqrt{\frac{T}{n}} + b(t_k, \bar{X}_{t_k}) \frac{T}{n}. \quad (2)$$

Then  $\mathbb{E}[f(X_T)]$  is approximated by  $\mathbb{E}[f(\bar{X}_T)]$ . Practically, this latter quantity is approximated by an empirical mean over  $N$  samples of  $\bar{X}_T$ , giving rise to the unavoidable Monte Carlo error.

The *weak error* is defined for  $f \in \mathfrak{F}$  as

$$d_f(X, \bar{X}) := |\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]|.$$

This quantity depends on the regularity of the coefficients as well as the class of functions  $\mathfrak{F}$ . The smaller, the better. This convergence rate has practical consequences in the design of simulation methods.

Roughly speaking, when  $\sigma$  is uniformly elliptic and  $\sigma, b$  are at least of class  $C^2$  in space,  $d_f(X, \bar{X})$  converges to 0 at a rate  $n^{-1}$ , even if  $\mathfrak{F}$  is the class of bounded measurable functions or the class of Dirac distribution functions [6, 7].

\*Department of Mathematical Sciences — Ritsumeikan University and Japan Science and Technology Agency— 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577, Japan.

<sup>†</sup>Université de Lorraine, IECL, UMR 7502, Vandœuvre-lès-Nancy, F-54500, France & CNRS, IECL, UMR 7502, Vandœuvre-lès-Nancy, F-54500, France & Inria, Villers-lès-Nancy, F-54600, France.

<sup>‡</sup>Faculty of Science and Engineering — Hosei University — 3-7-2, Kajino-cho, Koganei-shi, Tokyo, 184-8584, Japan.

This is a non-trivial consequence of the regularity of the density of  $X_T$ . Weakening the conditions on the coefficients  $b$  or  $\sigma$  naturally imply slower convergence rates.

Using the regularity of the solution of the PDE  $\partial_t u = Lu$  with  $u(0, x) = f(x)$  where  $L$  is the infinitesimal generator of  $X$  with coefficients in some Hölder space and  $\mathfrak{F}$  a class of Hölder continuous functions, the weak rate of convergence was given in R. Mikulevičius and E. Platen [32] and R. Mikulevičius and Zhang [31]. More precisely, if<sup>1</sup>  $a, b \in H^{\alpha/2, \alpha}(\bar{H})$  and  $f \in H^{2+\alpha}(\mathbb{R}^d)$  for  $\alpha \in (0, 1) \cup (1, 2) \cup (2, 3)$  then  $d_f(X, \bar{X})$  converges to zero at a rate  $n^{-\alpha/2}$  for  $\alpha \in (0, 2) \setminus \{1\}$  and  $n^{-1}$  for  $\alpha \in (2, 3)$ .

Due to its strong connection with PDE regularity results, this setting excludes integer values of  $\alpha$ , that is  $C^k$  coefficients, and strongly links  $\mathfrak{F}$  to the Hölder regularity of the coefficients.

We propose an alternative approach in which we decouple the effect of the regularity of the drift and the one of the terminal functions.

This article aims at studying the weak rate of convergence under various conditions on the diffusion coefficient  $\sigma$ , the drift  $b$  and the class  $\mathfrak{F}$  of terminal conditions. Conditions as weak possible are sought, including situations where the drift is discontinuous.

Few articles have been devoted to the convergence of the Euler-Maruyama scheme when the drift coefficient  $b$  presents some discontinuities in space [2, 4, 10, 19–21, 33, 36, 43] and time [35]. Other related studies can be found in [9, 12, 30]. Still, optimal weak rates of convergence remain an elusive subject.

Our method of analysis is to first give a perturbation result in order to compare  $\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]$  where  $X^\epsilon$  has a smooth drift coefficient  $b_\epsilon$  which approximates  $b$ . Then we provide some bounds on the weak rate of convergence for the Euler-Maruyama scheme for  $X^\epsilon$  when one regularizes the drift coefficient within a certain class, say  $\mathfrak{M}$  (See Figure 1). We will then analyze the possible interplay of the class of test functions  $\mathfrak{F}$  and of the class  $\mathfrak{M}$  of regularized coefficients with respect to the rate of convergence.

Putting these two results together we finally get various weak rates of convergence according to the choice of  $\mathfrak{M}$  and  $\mathfrak{F}$ . We then get results which are different from the ones in [32] under different sets of hypotheses. For example, a lower regularity condition on the terminal condition could be imposed.

Still, one may be interested in the performance of the Euler-Maruyama scheme on the light of the results in Figure 1. For this reason, we also perform an analysis to study the distance between the Euler-Maruyama scheme  $\bar{X}$  and the Euler-Maruyama scheme  $\bar{X}^\epsilon$ . This gives a bound on the weak rate of convergence of the Euler-Maruyama scheme even with a discontinuous coefficient. Lastly, in Section 8.3, we show a special situation with an SDE with an alternating drift where the weak rate of convergence is equal to  $n^{-1}$ . This result is backed by numerical experiments which are presented together with an abridged, short version of the results in [24].

As an approximation of the drift is used to establish the rate of convergence, the rate we give cannot be said to be optimal. Still the results presented here give estimates of rates of convergence that may help to design simulation methods for particular situations so that convergence of the simulation method is assured.

It also points towards the need of new techniques to tackle these problems.

**Outline.** Notations, hypotheses and our main results are given in Sections 2, 4 and 3. The perturbation formulae are proved in Section 5. Sections 6 provides us with estimates on  $d_p(b, b_\epsilon)$ . Section 7 provides a weak rates of convergence when the drift coefficient is smooth. Section 8 is devoted to extending our results when the diffusivity  $\sigma$  is constant. In Section 8.3, we show that the weak rate of convergence could be  $n^{-1}$  in a special situation. Finally, comments on numerical simulations are given in Section 8.4.

## 2 Notations, spaces and norms

**Matrices.** Vectors in  $\mathbb{R}^d$  are usually considered as column vectors unless explicitly stated otherwise. For a vector or matrix  $m$ ,  $m^*$  denotes the transpose of  $m$ . For a  $d \times k$ -matrix  $m$ , we define its norm as  $|m| = \max_{|x|=1} |mx|$ . We define similarly the norm of a multi-linear transformation. Note that we do not make any particular difference in notation for norms of real numbers, vectors or multi-linear transformations.

<sup>1</sup>For exact definitions of the functional spaces, we refer to Section 2.

**Spaces of functions.** For  $r > 0$ , we denote by  $C_b^r(U)$  the space of real valued, bounded, continuous functions on an open set  $U \subseteq \mathbb{R}^d$ , with continuous derivatives up to order  $\lfloor r \rfloor$  which are also bounded and its derivatives are  $r - \lfloor r \rfloor$ -Hölder continuous. We denote the high order derivatives of a function  $f \in C_b^r(U)$  for any multi-index  $\alpha = (\alpha_1, \dots, \alpha_j) \in \{1, \dots, d\}^j$  of length  $j \leq \lfloor r \rfloor$  as either  $\partial_{x_{\alpha_1}, \dots, x_{\alpha_j}}^j \equiv \frac{\partial^j}{\partial x_{\alpha_1} \dots \partial x_{\alpha_j}}$ . The multi-linear transform is denoted by  $\partial_x^j \equiv \{\partial_{x_{\alpha_1}, \dots, x_{\alpha_j}}^j; \alpha \in \{1, \dots, d\}^j\}$ . For  $f = (f^1, \dots, f^d) \in (C_b^k(U))^d$ ,  $k \in \mathbb{N}$ , we define the following semi-norm

$$\|\partial_x^j f\|_\infty := \max_{i=1, \dots, d, \alpha \in \{1, \dots, d\}^j} \sup_{x \in U} |\partial_{x_{\alpha_1}, \dots, x_{\alpha_j}}^j f^i(x)|.$$

For  $k \in \mathbb{N}$ , the norm on  $C_b^k(U)$  is  $\|f\|_{b,k} = \sum_{j=0}^k \|\partial_x^j f\|_\infty$ . For  $r \notin \mathbb{N}$ , the norm on  $C_b^r(U)$  is

$$\|f\|_{b,r} := \|f\|_{b,k} + \sup_{\substack{x, y \in U, \\ x \neq y}} \frac{|\partial_x^{\lfloor r \rfloor} f(x) - \partial_x^{\lfloor r \rfloor} f(y)|}{|x - y|^{r - \lfloor r \rfloor}}$$

The extension of the above spaces to non-open sets  $U$  is taken as usual by extending the definition of continuity and differentiability to the boundary points through appropriate limits using elements in  $U$ .

**Growth of functions.** We say that a real-valued function  $f$  has at most *polynomial growth* in  $\mathbb{R}^d$  if there exists an integer  $k$  and a constant  $C \geq 0$  such that  $|f(x)| \leq C(1 + |x|^k)$  for any  $x \in \mathbb{R}^d$ . The space of continuous functions with at most polynomial growth is denoted by  $C_p(\mathbb{R}^d)$ .

We say that a real-valued function  $f$  has at most *exponential growth* if for some constants  $c_1$  and  $c_2$ ,  $|f(x)| \leq c_1 e^{c_2 |x|}$  for all  $x \in \mathbb{R}^d$ . The corresponding space of continuous functions with at most exponential growth is denoted by  $C_e(\mathbb{R}^d)$ .

The space of real valued continuous functions that are *slowly increasing* is denoted by  $C_{SI}(\mathbb{R}^d)$ . That is,  $f \in C_{SI}(\mathbb{R}^d)$  if and only if for every  $c > 0$ ,

$$\lim_{|x| \rightarrow \infty} |f(x)| e^{-c|x|^2} = 0.$$

We denote by  $C_a^k(\mathbb{R}^d)$  with  $a = e, p$  or  $SI$  the space of continuous functions  $f$  with continuous derivatives up to order  $k$  such that  $f$  and its derivatives of order up to  $k \in \mathbb{N}$  belong to the space  $C_a(\mathbb{R}^d)$ .

**Functions of space and time.** Two particular domains will appear frequently. Let  $H = [0, T] \times \mathbb{R}^d$  and  $\bar{H} = [0, T] \times \mathbb{R}^d$  for  $T > 0$ . The set  $C_a(\bar{H})$  for  $a = e, p, SI$  denotes the set of functions  $u : H \rightarrow \mathbb{R}$  which are continuous on  $\bar{H}$  and such that  $u(s, \cdot) \in C_a(\mathbb{R}^d)$  uniformly for  $s \in [0, T]$ . For example, in the case that  $a = SI$ ,  $\lim_{|x| \rightarrow \infty} \sup_{s \in [0, T]} |u(s, x)| e^{-c|x|^2} = 0$  for every  $c > 0$ . The extension of the norms from the spaces  $C_b^k(\mathbb{R}^d)$  to the spaces  $C_b^k(\bar{H})$  is done by taking supremums with respect to time. That is,  $\|f\|_{b,k} := \sup_{t \in [0, T]} \|f(t, \cdot)\|_{b,k}$ .

We set  $H^{\alpha/2, \alpha}(\bar{H})$  is the space of continuous functions which are  $\alpha$ -Hölder continuous in space uniformly in time and  $\alpha/2$ -Hölder continuous in time uniformly in space.

We denote by  $W_p^{1,2}(H)$  the space of functions in  $L^p(H)$  whose weak derivatives (first in time, first and second in space) are also in  $L^p(H)$ . The space  $W_{p,loc}^{1,2}(H)$  is the space of functions in  $W_p^{1,2}(K)$  for any compact  $K \subseteq H$ . Embedding theorems for these spaces may be found in [28, II. § 3].

We will also use the following norms  $\|f\|_{L^r, q(H)} = (\int_0^T (\int |f(s, x)|^q dx)^{r/q} ds)^{1/r}$  for  $0 < r, q < +\infty$  and  $\|f\|_{L^\infty, q(H)} = \sup_{t \in [0, T]} \|f(t, \cdot)\|_{L^q(\mathbb{R}^d)}$ . In particular, we simplify the notation using  $\|f\|_{L^q, q(H)} = \|f\|_{L^q(H)}$ .

**Miscellany.** The density of a  $d$ -dimensional Gaussian random vector with mean zero and covariance matrix  $tId$  is denoted by  $g_t(x) := \frac{1}{\sqrt{2\pi t^d}} \exp\left(-\frac{|x|^2}{2t}\right)$ .

Constants will be denoted by the same symbol  $C$  or  $K$  and unless stated explicitly they may change values from one line to the next. In many cases, we make the effort to explicitly state the dependence of the constants on the various parameters of the problem.

When using Hölder's inequality we will say that  $(p, q)$  is a conjugate pair if  $p > 1, q > 1$  and  $p^{-1} + q^{-1} = 1$ . Similarly, we say that  $(p, q)$  is a near conjugate pair if  $p > 1, q > 1$  and  $p^{-1} + q^{-1} < 1$ .

In general, we will not write the domains over which space integrals are evaluated unless there is the possibility of confusion.

### 3 The approximation framework to obtain the weak rate of convergence

In this section we explain in detail how the analysis of the approximation procedure described in the Introduction will be carried out. We first give our standing assumptions through the article.

Let  $\mathfrak{S}(\lambda, \Lambda)$  be the class of measurable functions  $\sigma$  on  $\bar{H}$  with values in the space of symmetric  $d \times d$ -matrices such that

$$\exists \Lambda \geq \lambda > 0 \text{ s.t. } \lambda |\xi|^2 \leq \xi^* a(t, x) \xi \leq \Lambda |\xi|^2, \quad \forall (t, x) \in \bar{H}, \quad \forall \xi \in \mathbb{R}^d, \quad (\text{H1})$$

$$\sigma \text{ is uniformly continuous on } \bar{H}. \quad (\text{H2})$$

For  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$ , we set  $a := \sigma \sigma^*$ .

*Remark 1.* As  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  is symmetric, (H1) implies that the eigenvalues of  $\sigma$  are the square root of the corresponding eigenvalues of  $a$  and then (H1) holds for  $\sigma$  with  $\lambda$  and  $\Lambda$  replaced by  $\sqrt{\lambda}$  and  $\sqrt{\Lambda}$ .

Sometimes, we need stronger regularity condition on  $\sigma$  and we define  $\mathfrak{H}(\lambda, \Lambda)$  as the functions  $\sigma$  in  $\mathfrak{S}(\lambda, \Lambda)$  for which

$$\sigma \in \mathbf{H}^{1/2,1}(\bar{H}). \quad (\text{H2}')$$

Hypothesis (H2') means that  $\sigma$  is Lipschitz continuous in space (uniformly in time) and 1/2-Hölder continuous in time (uniformly in space).

For the drift, we consider the class  $\mathfrak{B}(\Lambda)$  of measurable functions  $b : \bar{H} \rightarrow \mathbb{R}^d$  such that

$$|b(t, x)| \leq \Lambda \text{ for all } (t, x) \in \bar{H}. \quad (\text{H3})$$

When  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  and  $b \in \mathfrak{B}(\Lambda)$ , there exists a unique weak solution in some probability space  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  for the SDE (1) (see e.g. [40]). Here  $B$  is a  $d$ -dimensional Wiener process defined in the probability space  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ .

*Remark 2.* As shown in [42], if (1) has a strong solution for  $b \equiv 0$ , then (1) also admits a strong solution.

For the solution  $X$  to (1), we denote by  $\bar{X}$  the continuous Euler-Maruyama scheme of  $X$  with a time step  $T/n$ . For this, we define for fixed  $n \in \mathbb{N}$ ,  $\phi(s) \equiv \phi_n(s) = \sup\{t \leq s \mid t = k/n \text{ for } k \in \mathbb{N}\}$ , then

$$\bar{X}_t = x + \int_0^t \sigma(\phi(s), \bar{X}_{\phi(s)}) dB_s + \int_0^t b_\epsilon(\phi(s), \bar{X}_{\phi(s)}) ds. \quad (3)$$

The vector  $(\bar{X}_0, \dots, \bar{X}_T)$  has the same distribution as the vector defined by the recursive relation (2). Yet  $\bar{X}$  defined by (3) is defined for any time  $t \in [0, T]$ , not only for discrete ones, and on the same probability space as  $X$ .

On the space  $\mathfrak{B}(\Lambda)$  of drift, we define a distance  $d_p$  that depends on a parameter  $p > 0$  (for example,  $p$  is the exponent of the functional space  $L^p(\bar{H})$ ).

We denote by  $\mathfrak{M}$  a set of “regular” measurable functions in  $\mathfrak{B}(\Lambda)$ . For example,  $\mathfrak{M}$  could be composed of functions of class  $C^k$ . Approximations  $b_\epsilon$  of  $b \in \mathfrak{B}(\Lambda)$  are sought in  $\mathfrak{M}$ , so that *a priori*  $b \notin \mathfrak{M}$ .

Let  $X \equiv X(b)$  be the random process which we want to approximate, solution of (2), which depends on the irregular function  $b$ . A first approximation is obtained by replacing the irregular function  $b$  by a “regularized” function  $b_\epsilon \in \mathfrak{M}$  by assuming that  $\lim_{\epsilon \rightarrow 0} d_p(b_\epsilon, b) = 0$ . Using  $b_\epsilon$ , we construct an approximation  $X^\epsilon := X(b_\epsilon)$  as the unique weak solution on  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  to

$$X_t^\epsilon = x + \int_0^t \sigma(s, X_s^\epsilon) dB_s + \int_0^t b_\epsilon(s, X_s^\epsilon) ds. \quad (4)$$

The approximation quality of  $X^\epsilon$  to  $X$  is measured through the class of test functions  $\mathfrak{F}$ . For this, we use a “bias” function denoted by  $d_f(\cdot, \cdot)$  for fixed  $f \in \mathfrak{F}$  defined by  $d_f(X, Y) := |\mathbb{E}[f(X_T)] - \mathbb{E}[f(Y_T)]|$  for two processes  $X$  and  $Y$ .

In our framework, we choose  $d_f$  and  $d_b$  on  $\mathfrak{B}(\Lambda)^2$  such that  $d_p(b, b_\epsilon)$  does not depend on  $f$  once  $p$  is fixed (however, the choice of  $p$  may depend of the class  $\mathfrak{F}$ ). The key point of our analysis is to reduce to situations where

$$d_f(X, X^\epsilon) = |\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq C(p, q) K_q(f) d_p(b, b_\epsilon) \quad (5)$$

for an almost conjugate pair  $(p, q)$  and a constant  $C(p, q)$  which generally explodes as  $p^{-1} + q^{-1} \rightarrow 1$ .

The couple of parameters  $(p, q)$  measures the strength of the convergence of  $b_\epsilon$  to  $b$  and the regularity of the functions in  $\mathfrak{F}$ . Conditions for which the inequality (5) holds are given in Section 5.

Finally in order to obtain the final rate of convergence  $d_f(X, \bar{X})$ , we use the triangle inequality

$$d_f(X, \bar{X}) \leq d_f(X, X^\epsilon) + d_f(X^\epsilon, \bar{X}^\epsilon) + d_f(\bar{X}^\epsilon, \bar{X}).$$

where  $\bar{X}^\epsilon$  is the continuous time Euler-scheme associated to  $X^\epsilon$  (which depends on the time step  $T/n$  and the regularized drift  $b_\epsilon$ ). As the term  $d_f(\bar{X}^\epsilon, \bar{X})$  is not always simple to quantify, we could simply estimate  $d_f(X, \bar{X}^\epsilon)$ .

Under rather general hypotheses, when  $\sigma$  and  $b_\epsilon$  belong to “good” classes of functions  $\mathfrak{D} \subset \mathfrak{S}(\lambda, \Lambda)$  and  $\mathfrak{M} \subset \mathfrak{B}(\Lambda)$  (for example  $\mathfrak{D}, \mathfrak{M} = \mathbb{H}^{\alpha/2, \alpha}(\bar{H})$  for some  $\alpha > 0$  or  $\mathfrak{D}, \mathfrak{M} = C_b^{1,3}(\bar{H})$ ), and when  $f$  belongs to a proper class of functions  $\mathfrak{F}$  (for example,  $\mathfrak{F} = C^{2+\alpha}(\mathbb{R}^d)$  or  $\mathfrak{F} = C^3(\mathbb{R}^d) \cap C_{\text{SI}}(\mathbb{R}^d)$  respectively), a weak rate of convergence of the Euler-Maruyama scheme  $\bar{X}^\epsilon$  can be given. We assume that there exist  $\delta > 0$  and some constant  $C_\epsilon = O(\epsilon^{-\beta})$  such that

$$d_f(X^\epsilon, \bar{X}^\epsilon) := |\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\bar{X}_T^\epsilon)]| \leq \frac{C_\epsilon}{n^\delta}.$$

This is in general the case when one choose a regularization  $b_\epsilon$  of  $b$  by using mollifiers. Clearly the parameters  $\beta$  and  $\delta$  depend on the classes  $\mathfrak{D}$ ,  $\mathfrak{M}$  and  $\mathfrak{F}$ .

If we assume that  $d_p(b, b_\epsilon)$  decreases to 0 at the rate  $O(\epsilon^\gamma)$  then the distance between the drift coefficient and its approximation  $b_\epsilon \in \mathfrak{M}$  is measured by the parameter  $\gamma$ . When one uses the Euler-Maruyama scheme with a regularized drift in  $\mathfrak{M}$  and a time step  $T/n$ , the weak rate of convergence is expressed by the parameter  $\delta$ . However, the constant  $C_\epsilon := \widehat{K}(f)\epsilon^{-\beta}$  depends itself on  $f \in \mathfrak{F}$  and on the drift  $b_\epsilon \in \mathfrak{M}$ . It usually explodes as  $\epsilon \rightarrow 0$  as  $b \notin \mathfrak{M}$ . The rate of explosion is characterized by the parameter  $\beta$ .

Therefore the parameters  $\delta, \beta$  and  $\gamma$  express the interplay between the rate of convergence, the regularity of the coefficient and the regularity of test functions.

Later, we provide several situations where  $\delta, \beta$  and  $\gamma$  may be computed for several classes  $\mathfrak{M}$  of drift coefficients and  $\mathfrak{F}$  of test functions, therefore achieving several weak rates of convergence.

Schematically, the above two results explained above and the corresponding analysis to obtain a weak convergence result can be seen as follows:

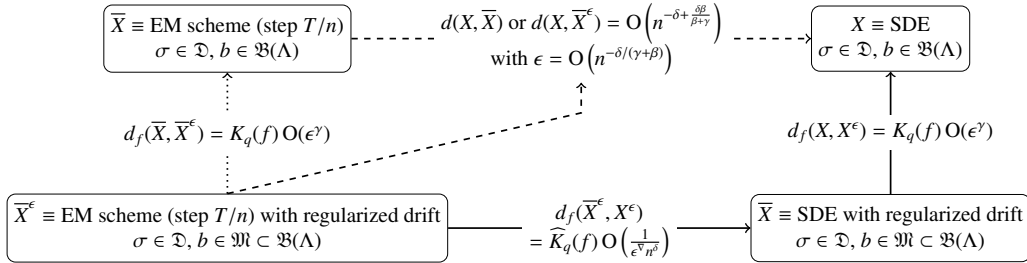


Figure 1: Computation of the weak rate of convergence for  $f \in \mathfrak{F}$ .

Optimizing over the choice of  $\epsilon$  leads to the following result, which we summarized in Figure 1.

**Theorem 1.** *Let us consider classes of suitable functions  $\mathfrak{F}$ ,  $\mathfrak{D} \subset \mathfrak{S}(\lambda, \Lambda)$  and  $\mathfrak{M} \subset \mathfrak{B}(\Lambda)$  such that for  $f \in \mathfrak{F}$ ,  $\sigma \in \mathfrak{D}$ ,  $b_\epsilon \in \mathfrak{M}$ ,  $\epsilon \in (0, 1)$ , there exist positive constants  $p', \gamma, \beta, \delta$  and an almost conjugate pair  $(p, p')$  which only depend on the definition of the classes  $\mathfrak{F}$  and  $(\mathfrak{D}, \mathfrak{M})$  such that  $d_p(b, b_\epsilon) = O(\epsilon^\gamma)$  and*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq C(p, q)K_{p'}(f)d_p(b, b_\epsilon) \text{ and } |\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\bar{X}_T^\epsilon)]| \leq \widehat{K}(f)\epsilon^{-\beta}n^{-\delta}.$$

Here  $K_{p'}(f)$  and  $\widehat{K}(f)$  are positive constants depending on  $f$ . Then for  $\epsilon = O(n^{-\delta/(\gamma+\beta)})$ , both errors above are of the same order  $O(n^{-\kappa})$  with  $\kappa = \delta - \frac{\delta\beta}{\gamma+\beta}$  and therefore gives the optimal rate

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T^\epsilon)]| = O(n^{-\kappa}).$$

In the above result one clearly sees that, given (5),  $\gamma$  measures the irregularity of  $b$  with respect to the class  $\mathfrak{M}$  and the law of  $Y$ . The parameter  $\delta$  measures the weak error within the class  $\mathfrak{M}$ , for  $f \in \mathfrak{F}$  and  $\beta$  measures how the weak error degenerates as we exit the class  $\mathfrak{M}$  when the weak rate of convergence is  $n^{-\delta}$ . The constants  $K_p(f)$  and  $\widehat{K}(f)$  will determine the method of proof to be used in order to obtain the respective rates.

## 4 SDE and PDE under weak assumptions on the coefficients

Let  $L$  be the differential operator associated to (1):

$$L := \frac{1}{2} \sum_{i,j=1}^d a_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i \frac{\partial}{\partial x_i}.$$

Solution of second-order PDE with non-divergence form operators may be sought in Hölder space when the coefficients are Hölder (See *e.g.* [28, Chapter IV §5], which was the base of the results in [32]. In an alternative approach, the solution are sought in Sobolev spaces, which allows to consider weak conditions on the regularity of the drift [28, Chapter IV §7]. The usual probabilistic representation for such solution also holds in this situation.

**Theorem 2** (Theorem 3' and Corollary p. 401, [42]). *Let  $L$  be as above with  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  and  $b \in \mathfrak{B}(\Lambda)$ . Fix  $T > 0$  and assume that  $f \in C_{\text{SI}}(\mathbb{R}^d)$ . The Cauchy problem*

$$\begin{cases} \frac{\partial u(t, x)}{\partial t} + Lu(t, x) = 0, \\ u(T, x) = f(x) \end{cases} \quad (6)$$

has a solution  $u$  in  $\cap_{p>1} W_{p,\text{loc}}^{1,2}(H) \cap C_{\text{SI}}(\overline{H})$ . This solution, satisfying  $u(s, x) = \mathbb{E}[f(X_T)|X_s = x]$ , is unique in  $W_{p,\text{loc}}^{1,2}(H) \cap C_{\text{SI}}(\overline{H})$  for  $p \geq d + 1$ . Besides,

$$f(X_T) = u(s, X_s) + \int_s^T \nabla u(r, X_r)^* \sigma(r, X_r) dB_r$$

and

$$\mathbb{E} \left[ \int_0^T |\nabla u(s, X_s) \sigma(s, X_s)|^2 ds \right] = \mathbb{E}[f(X_T)^2] - \mathbb{E}[f(X_T)]^2 = \text{Var}(f(X_T)).$$

The conditions  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  and  $b \in \mathfrak{B}(\Lambda)$  are assumed throughout the rest of the article without any further mentioning.

In the particular case where  $b \equiv 0$ , we denote by  $v$  the solution to (6) given by Theorem 2.

Let  $X$  be the solution to (1) with a general  $b$ . The Itô's formula yields

$$f(X_T) = v(0, x) + \int_0^T \nabla v(s, X_s)^* \sigma(s, X_s) dB_s + \int_0^T b^*(s, X_s) \nabla v(s, X_s) ds.$$

To avoid further confusions, we denote by  $Y$  the weak solution of the equation

$$Y_t = x + \int_0^t \sigma(s, Y_s) dB_s.$$

Therefore  $v(s, x) = \mathbb{E}[f(Y_T)|Y_s = x]$ .

The above probabilistic representation of the solution is our first element of analysis. Our second one is given by the Girsanov theorem. Let  $\Gamma : \Omega \times [0, T] \rightarrow \mathbb{R}^d$ , be a bounded, predictable stochastic process. To  $\Gamma$ , we relate the exponential martingale

$$\mathcal{Z}(\Gamma)_t := \exp \left( \int_0^t \Gamma_s^* dB_s - \frac{1}{2} \int_0^t \Gamma_s^* \Gamma_s ds \right). \quad (7)$$

This martingale  $\mathcal{Z}(\Gamma)$  is the unique strong solution to the SDE

$$\mathcal{Z}(\Gamma)_t = 1 + \int_0^t \mathcal{Z}(\Gamma)_s \Gamma_s^* dB_s.$$

The martingale  $\mathcal{Z}(\Gamma)$  can be used to define a new measure  $\mathbb{Q} \equiv \mathbb{Q}^\Gamma$  which is absolutely continuous with respect to  $\mathbb{P}$ , with a Radon-Nikodym derivative process given by  $\frac{d\mathbb{Q}^\Gamma}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \mathcal{Z}(\Gamma)_t$ ;  $t \geq 0$ . Various moment properties of  $\mathcal{Z}(\Gamma)$  are given in Appendix 10.1.

We let  $dW_t^\Gamma = dB_t - \Gamma_t dt$  denote the Wiener process on the probability space  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q}^\Gamma)$ . As it is well known, choosing the right definition for  $\Gamma$  and using Girsanov theorem, the equations (4) and (1) have under the respective probability measures  $\mathbb{Q}^\Gamma$ , the same law as the solution of (further details will be given later)

$$Y_t^\Gamma = x + \int_0^t \sigma(s, Y_s^\Gamma) dW_s^\Gamma.$$

As  $Y^\Gamma \stackrel{\mathcal{L}}{=} Y$  where the previous equality means that the law of both processes are the same and therefore we will not make any distinction between them as we will only be considering their expectations.

Similarly, expectations under different probability measures are denoted by  $\mathbb{E}$  or  $\mathbb{E}^\mathbb{Q}$  as the underlying probability measure is well understood from the context. Many times we use the equality in law mentioned above and therefore we only use  $\mathbb{E}[f(Y_T)]$ .

## 5 A perturbation formula for $d_f(X, X^\epsilon)$

We now give some general conditions for which the key inequality (5) holds. For this, we provide a perturbation formula which compares the laws of two diffusions with different drifts.

For  $p, q \geq 1$  and a measurable function  $g : \bar{H} \rightarrow \mathbb{R}^k$ , we say that  $g \in L^{p,q}(X)$  if its associated norm

$$\|g\|_{X,q,p} := \mathbb{E} \left[ \left( \int_0^T |g(s, X_s)|^q ds \right)^{p/q} \right]^{1/p}$$

is finite. The above norm is always considered for  $X$  defined as the weak solution of (1) on  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ . We somewhat abuse the notation, using the convention  $\|g\|_{X,p} := \|g\|_{X,2,p}$ ,  $L^p(X) := L^{2,p}(X)$ ,  $\|g\|_{X,\infty} := \|g\|_\infty$ . In a similar fashion, we can define  $\|g\|_{Y,p}$ .

**Proposition 1.** *Let  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  and  $b, b_\epsilon \in \mathfrak{B}(\Lambda)$  (See Section 3). Assume that  $(p, p')$  is an almost conjugate pair and that  $f$  is a function such that  $\mathbb{E} \left[ |f(Y_T) - \mu|^{p'} \right]^{1/p'} < \infty$  for some constant  $\mu$ . Then there exists a constant  $C(p, p')$  that depends only on  $p, p', T, \Lambda, \lambda$  such that*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq C(p, p') \|b - b_\epsilon\|_{Y,p} \mathbb{E} \left[ |f(Y_T) - \mu|^{p'} \right]^{1/p'}. \quad (8)$$

*Proof.* Let us define  $\Gamma_s^\epsilon := \sigma^{-1}(s, Y_s) b_\epsilon(s, Y_s)$  and  $\Gamma_s := \sigma^{-1}(s, Y_s) b(s, Y_s)$ . Note that  $|\Gamma^\epsilon| \leq 2\Lambda\lambda^{-1}$  and  $|\Gamma| \leq \Lambda\lambda^{-1}$ . Using Girsanov's theorem and since  $\mathbb{E}[\mathcal{Z}(\Gamma)_T] = \mathbb{E}[\mathcal{Z}(\Gamma^\epsilon)_T] = 1$ ,

$$\begin{aligned} |\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(X_T)]| &= |\mathbb{E}[\mathcal{Z}(\Gamma^\epsilon)_T (f(Y_T) - \mu)] - \mathbb{E}[\mathcal{Z}(\Gamma)_T (f(Y_T) - \mu)]| \\ &\leq \mathbb{E}[(\mathcal{Z}(\Gamma^\epsilon) - \mathcal{Z}(\Gamma))^p]^{1/p} \mathbb{E}[(f(Y_T) - \mu)^{p'}]^{1/p'} \end{aligned}$$

for a conjugate pair  $(p, p')$ . Hence, (8) follows from Lemma 11.  $\square$

**Corollary 1.** *When  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  and  $b, b_\epsilon \in \mathfrak{B}(\Lambda)$ ,*

i) *If  $f \in C_{\text{Sl}}(\mathbb{R}^d)$ , then for  $p > 2$*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq C(p) \|b - b_\epsilon\|_{Y,p} \sqrt{\text{Var } f(Y_T)}.$$



ii) If  $f \in C_b(\mathbb{R}^d)$ , then for any  $p > 1$ ,

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq C(p)\|b - b_\epsilon\|_{Y,p}\|f\|_\infty.$$

iii) If  $f \in C_{\text{Sl}}(\mathbb{R}^d)$  is uniformly Lipschitz continuous with constant  $\text{Lip}(f)$ , for any  $p > 1$ ,

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq C(p)\Lambda T^{1/2}\|b - b_\epsilon\|_{Y,p}\text{Lip}(f).$$

*Proof.* The proof of this corollary is immediate using Gaussian upper bound estimates (see [3], or [39] or the discussion before Lemma 1) and by choosing respectively  $\mu = \mathbb{E}[f(Y_T)]$  for i),  $\mu = 0$  for ii) and  $\mu = f(x)$  together with the control  $\mathbb{E}[|Y_T - x|^{p'}]^{1/p'} \leq \Lambda T$  for any  $T > 0$  for iii).  $\square$

The above Corollary shows that choosing the norm  $K_2(f)$  is subject to the properties of interest for the test function  $f$ . We are not bound to consider the above hypotheses on  $f$ . In particular, the notion of *fractional smoothness* introduced in [15–17] allows one to consider several classes  $\mathfrak{F}$  of functions for which  $K_2(f) := \sqrt{\text{Var } f(Y_T)}$  is finite. These classes may even include discontinuous functions. This will be subject to future works.

In Proposition 1 and Corollary 1, the effect of the drift is isolated in the norm  $\|b - b_\epsilon\|_{Y,p}$  and the effect of the test function only appears in  $\mathbb{E}[|f(Y_T) - \mu|^2]^{1/2}$  as the process  $Y$  is independent of  $b$  or  $b_\epsilon$ . On the other hand, the condition  $p > 2$  may be too restrictive to determine a weak rate of convergence. In the next proposition, we give an alternative bound for  $d_f(X, X^\epsilon)$  which retains the irregular drift but involves a control on the gradient of the solution to (6),  $\|\nabla u\|_{X,p'}$  which is more difficult to obtain. The positive feature of this bound is that we can use the resulting inequality for any  $p > 1$ . This proposition is particularly useful for taking  $p'$  as big as possible (including  $p' = +\infty$ ).

**Proposition 2.** Let  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  and  $b, b_\epsilon \in \mathfrak{B}(\Lambda)$ . Assume that  $f \in C_{\text{Sl}}(\mathbb{R}^d)$  is such that for some  $1 < p' \leq \infty$ ,  $\|\nabla u\|_{X,p'} < +\infty$ , where  $u$  is the solution to (6) with terminal condition  $f$ . Then there exists a constant  $C_2(p, p')$  depending only on  $T, \Lambda, \lambda$  and  $(p, p')$  an almost conjugate pair such that

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq C_2(p, p')\|b - b_\epsilon\|_{Y,p}\|\nabla u\|_{X,p'}$$

with  $C_2(p, p') \xrightarrow[p \rightarrow p'/(p'-1)]{} +\infty$ .

*Proof.* Set  $\Gamma_s = \sigma^{-1}(s, X_s)(b_\epsilon(s, X_s) - b(s, X_s))$  and  $\hat{\Gamma} := \sup_{(\omega, s) \in \Omega \times [0, T]} |\Gamma_s| \leq \Lambda \lambda^{-1}$ . From Girsanov's theorem and the quadratic covariation between martingales,

$$\begin{aligned} \Delta &:= -(\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]) = \mathbb{E}[(\mathcal{Z}(\Gamma)_T - 1)f(X_T)] = \mathbb{E}\left[\int_0^T \mathcal{Z}(\Gamma)_s (b_\epsilon - b)^*(s, X_s) \nabla u(s, X_s) ds\right] \\ &= \mathbb{E}\left[\mathcal{Z}(\Gamma)_T \int_0^T (b_\epsilon - b)^*(s, X_s) \nabla u(s, X_s) ds\right]. \end{aligned}$$

The martingale property of  $\mathcal{Z}(\Gamma)$  has been used in the last step. As in the proof of Proposition 1, for  $\alpha_1 > 2$ ,  $\alpha_2 \in (1, p)$  and  $p'$  with  $\alpha_1^{-1} + \alpha_2^{-1} + (p')^{-1} = 1$ ,

$$|\Delta| \leq T^{1-\frac{1}{\alpha_1}} \mathbb{E}[|\mathcal{Z}(\Gamma)_T|^{\alpha_1}]^{\frac{1}{\alpha_1}} \mathbb{E}\left[\left(\int_0^T |(b_\epsilon - b)(s, X_s)|^2 ds\right)^{\alpha_2/2}\right]^{\frac{1}{\alpha_2}} \mathbb{E}\left[\left(\int_0^T |\nabla u(s, X_s)|^2 ds\right)^{p'/2}\right]^{\frac{1}{p'}}.$$

The result follows from Lemma 9(I) and Lemma 9(III) since  $\|b_\epsilon - b\|_{X, \alpha_2} \leq \kappa_1(\hat{\Gamma}, \alpha_2, p)\|b_\epsilon - b\|_{Y, p}$  with  $p > \alpha_2$ .  $\square$

## 6 Rate of convergence of smooth approximations of the irregular drift: Considerations for $d_p(b, b_\epsilon)$

In the previous section we have analyzed the distance  $d_f(X, X^\epsilon)$  and we have found in Propositions 1 and 2, that these distances depend on two factors. The first is the distance between  $b$  and  $b_\epsilon$  in the  $L^p(H)$ -norm. The second is a measure on the regularity of  $f$ .

The distance  $\|b - b_\epsilon\|_{Y,p}$  will be recurrent in the discussions to follow and various cases will be given.

Due to our results in Propositions 1 and 2 the smaller  $p$ , the better is our inequality, since the distance between  $b$  and  $b_\epsilon$  decrease in general with  $p$ . Furthermore, we show that this distance can be reduced to a problem of function approximation if one uses Gaussian upper bound estimates for the various processes that will appear.

The first needed ingredient is the Gaussian upper bound estimates also known as Aronson estimates.

We say that a process  $Y$  satisfies an *upper Gaussian estimate* if  $Y_t$  has a density  $\mathfrak{p}(t, x, y)$  for each  $t > 0$  that satisfies the inequality

$$\mathfrak{p}(t, x, y) \leq C_1 \mathfrak{g}_{C_2 t}(x - y), \quad \forall (t, x, y) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d, \quad (9)$$

for some positive constants  $C_1$  and  $C_2$ .

The estimate (9) holds for example if the diffusion coefficient  $a$  belongs to the Hölder space  $H^{\alpha/2, \alpha}(H)$  for some  $\alpha > 0$  (see e.g. [28, § IV.13, p. 377]). In fact, (9) is satisfied for time independent coefficients under hypotheses much weaker than (H1) and (H2) [3, 39].

If  $a$  satisfies (H1)-(H2) and that for some  $C, \eta > 0$ ,  $\sup_{t \in [0, T]} |a(t, x) - a(t, y)| \leq C|x - y|^\eta$  the (continuous) Euler-Maruyama scheme  $\bar{Y}_t$  has a density  $\bar{\mathfrak{p}}$  which satisfies a uniform Gaussian upper bound [29].

**Lemma 1.** *Assume that  $Y$  satisfies an Gaussian upper bound estimate (9). Then for any bounded function  $g : H \rightarrow \mathbb{R}^d$ ,  $p \geq 1$  and for any  $q, r \in (p \vee i, \infty]$  with  $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p \vee i}$ ,  $i = 1, 2$ , there exists a positive constant  $C_3$  such that for  $i = 1, 2$ ,*

$$\|g\|_{Y,i,p} \leq C_3 \left( \int_0^T \left( \int_{\mathbb{R}^d} |g(s, y)|^q dy \right)^{r/q} ds \right)^{1/r} =: C_3 \|g\|_{L^{r,q}(H)}. \quad (10)$$

*Proof.* Consider the case for  $\|g\|_{Y,1,p}$ . We can prove this lemma by using an explicit Gaussian upper bound estimation for the transition density, the Hölder inequality twice and an appropriate normalization of Gaussian density. In fact, first note that there exists a positive constant  $K(\lambda, \Lambda, T)$  such that for any  $\alpha > 1$

$$\left( \int_{\mathbb{R}^d} \mathfrak{p}(t, x, y)^\alpha dy \right)^{1/\alpha} \leq \frac{K(\lambda, \Lambda, T)}{\alpha^{d/(2\alpha)} t^{(\alpha-1)d/(2\alpha)}}.$$

Second, for  $p \in [1, q)$ ,  $\frac{p}{q} + \frac{1}{\alpha} = 1$  and a conjugate pair  $(a, a')$ ,

$$\begin{aligned} \|g\|_{Y,1,p}^p &\leq C \mathbb{E} \left[ \int_0^T |g(s, Y_s)|^p ds \right] = C \int_0^T \int_{\mathbb{R}^d} |g(s, y)|^p \mathfrak{p}(s, x, y) dy ds \\ &\leq C \int_0^T \left( \int_{\mathbb{R}^d} |g(s, y)|^q dy \right)^{p/q} \left( \int_{\mathbb{R}^d} \mathfrak{p}(s, x, y)^\alpha dy \right)^{1/\alpha} ds \\ &\leq C \left( \int_0^T \left( \int_{\mathbb{R}^d} |g(s, y)|^q dy \right)^{pa/q} ds \right)^{1/a} \left( \int_0^T \left( \int_{\mathbb{R}^d} \mathfrak{p}(s, x, y)^\alpha dy \right)^{a'/\alpha} ds \right)^{1/a'}. \end{aligned}$$

Letting  $pa = r$  and noting that the last integral is integrable under the restriction  $(1 - \frac{1}{\alpha})\frac{d}{2}a' < 1$ , (10) follows for  $i = 1$ . The cases  $q = \infty$  and/or  $r = \infty$  is dealt similarly.

Now, let us briefly consider  $\|g\|_{Y,2,p}$ . The proof is similar to the above one if  $p \geq 2$ . In the case that  $p \in (1, 2)$  we apply Hölder's inequality in order to obtain that  $\|g\|_{Y,i,p} \leq \|g\|_{Y,2}$  and apply the same chain of above inequalities in the case  $p = 2$ .  $\square$

*Remark 3.* Even in absence of Gaussian upper bound estimates, the *Krylov estimate* ([26] or [8, Theorem 7.6.2, p. 114]) could also be used with Hypothesis (H1) in order to get an estimate on  $\|g\|_{Y,p}$ . In the case of a time-homogeneous coefficient  $b$ , the Krylov estimate yields that

$$\|g\|_{Y,2} \leq C(\lambda, \Lambda) e^T \|g\|_{L^{2d}(H)}.$$

In case of a time-inhomogeneous coefficient, a similar estimate could be obtained when the domain is bounded. One should then estimate the exit time of  $Y$  from such domains.

The proof of the following lemma is a variation of Lemma 1 for discretized processes.

**Lemma 2.** *Assume that  $Y$  satisfies the Gaussian upper bound estimate (9) uniformly in  $n \in \mathbb{N}$  and that  $g : H \rightarrow \mathbb{R}^d$  is bounded. Then for any  $p \geq 1$  and any  $q, r \in (p \vee i, \infty]$  with  $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p \vee i}$ ,  $i = 1, 2$ , there exists some positive constant  $C_1 = C(p, r, q, \lambda, \Lambda, T)$  and  $C_1 = C(T, p, i)$  such that*

$$\left( \mathbb{E} \left[ \left( \int_0^T |g(\phi(s), Y_{\phi(s)})|^i ds \right)^{p/i} \right] \right)^{1/p} \leq C_1 \|g\|_{L^{r,q}(H)} + \frac{C_2}{n^{1/i}} |g(0, x)|.$$

*Proof.* As in the proof of Lemma 9 we prove by cases. Here we only give the proof for the case  $p \geq 2$ ,  $i = 2$ .

The time integral is actually a sum, so that with the Jensen inequality,

$$\begin{aligned} \left( \int_0^T |g(\phi(s), Y_{\phi(s)})|^i ds \right)^{\frac{p}{i}} &\leq 2^{\frac{p-i}{i}} \left( \int_{t_1}^T |g(\phi(s), Y_{\phi(s)})|^i ds \right)^{\frac{p}{i}} + 2^{p/i-1} \frac{T^{\frac{p}{i}}}{n^{\frac{p}{i}}} |g(0, x)|^p \\ &\leq 2^{\frac{p-i}{i}} T^{\frac{p-i}{p}} \left( \int_{t_1}^T |g(\phi(s), Y_{\phi(s)})|^p ds \right) + 2^{\frac{p-i}{i}} \frac{T^{\frac{p}{i}}}{n^{p/i}} |g(0, x)|^p. \end{aligned}$$

For a conjugate pair  $(\alpha, \alpha')$  with  $\alpha = \frac{q}{q-p}$ , there exists  $K(\lambda, \Lambda, T) > 0$  such that

$$\left( \int_{\mathbb{R}^d} \rho(t, x, y)^\alpha dy \right)^{1/\alpha} \leq \frac{K(\lambda, \Lambda, T)}{(2\pi)^{d/(2\alpha')} T^{d/(2\alpha')}},$$

where  $\rho$  is the density of  $Y$ . Then there exist some constant  $C(\alpha) = K(2\pi)^{-\frac{d}{2\alpha'}}$  such that

$$\begin{aligned} \mathbb{E} \left[ \int_{t_1}^T |g(\phi(s), Y_{\phi(s)})|^p ds \right] &= \frac{T}{n} \left\{ \sum_{k=1}^{n-1} \int_{\mathbb{R}^d} \rho(t_k, x, y) |g(t_k, y)|^p dy \right\} \\ &\leq \frac{C(\alpha)}{T^{d/2\alpha'-1}} \cdot \frac{1}{n} \sum_{k=1}^{n-1} \frac{n^{d/2\alpha'}}{k^{d/2\alpha'}} \left( \int_{\mathbb{R}^d} |g(t_k, y)|^q dy \right)^{p/q}. \end{aligned}$$

Finally, for a conjugate pair  $(a, a')$  satisfying  $a'd/2\alpha' < 1$ , the first sum can be rewritten as

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{n-1} \frac{n^{d/2\alpha'}}{(Tk)^{d/2\alpha'}} \left( \int_{\mathbb{R}^d} |g(t_k, y)|^q dy \right)^{p/q} &= \int_{t_1}^T \frac{1}{\phi(s)^{d/2\alpha'}} \left( \int_{\mathbb{R}^d} |g(\phi(s), y)|^q dy \right)^{p/q} ds \\ &\leq C \left( \int_0^T \frac{1}{s^{a'd/2\alpha'}} ds \right)^{1/a'} \left( \int_{t_1}^T \left( \int_{\mathbb{R}^d} |g(\phi(s), y)|^q dy \right)^{pa/q} ds \right)^{1/a}. \end{aligned}$$

The result follows choosing  $a = r/p$ . □

The above inequalities will be used for  $g = b - b_\epsilon$  and  $b = \partial^k b_\epsilon$ . Instead of the processes  $Y$  and  $\bar{Y}$  — the continuous Euler-Maruyama scheme associated to  $Y$  — sometimes other weighted combinations such as  $\alpha Y + (1-\alpha)\bar{Y}$  or  $\alpha \bar{Y} + (1-\alpha)\bar{Y}_{\phi(\cdot)}$  may also appear. These processes will also satisfy an Gaussian upper bound estimate as indicated in Lemma 13 in the Appendix.

Now we concentrate on the case that  $g = b - b_\epsilon$ . In this case, in order to know the rate at which  $g$  goes to zero in the above norms one has to go over the history of function approximation.

Unfortunately, our knowledge of these problems is limited, we therefore report here our findings with the possibility that there may be better results available.

To start, the space of smooth functions is dense in the space  $L^{dp}(H)$ . Therefore, one has to choose a subclass of functions in order to make sense of the approximation problem.

We now provide a standard clarifying example of approximations using regularized coefficients.

*Example 1.* Let  $b(t, x) = \mathbf{1}_{[\zeta_1, \zeta_2]}(x)$  for  $x \in \mathbb{R}$  and  $\zeta_1 < \zeta_2$ . Define for  $\epsilon > 0$ ,

$$b_\epsilon(x) := \begin{cases} 0, & x \in (-\infty, \zeta_1 - 2\epsilon) \cup (\zeta_2 + 2\epsilon, \infty), \\ \frac{1}{2\epsilon}x - \frac{\zeta_1 - 2\epsilon}{2\epsilon}, & x \in [\zeta_1 - 2\epsilon, \zeta_1), \\ -\frac{1}{2\epsilon}x + \frac{\zeta_2 + 2\epsilon}{2\epsilon}, & x \in (\zeta_2, \zeta_2 + 2\epsilon], \\ 1, & x \in [\zeta_1, \zeta_2]. \end{cases}$$

If we use  $b_\epsilon$  as an approximation of  $b$ ,  $b_\epsilon$  is Lipschitz continuous. Furthermore, we also have the following rates of convergence of  $b_\epsilon$  to  $b$ : for  $p \geq 1$ ,

$$\left( \int_{-\infty}^{\infty} |b_\epsilon(x) - b(x)|^p dx \right)^{\frac{1}{p}} = \left( \frac{4\epsilon}{p+1} \right)^{\frac{1}{p}} = \mathcal{O}\left(\epsilon^{\frac{1}{p}}\right). \quad (11)$$

The above approximation is only once differentiable with  $\|b'_\epsilon\|_\infty = \mathcal{O}(\epsilon^{-1})$  and  $\|b'_\epsilon\|_{L^p(H)} = \mathcal{O}(\epsilon^{-1+\frac{1}{p}})$ . Here,  $\epsilon$  is the regularization parameter.

Further regularity of the approximation  $b_\epsilon$  can be obtained, if we use a mollifier with the Gaussian kernel, that is  $b_\epsilon(x) := \int_{-\infty}^{+\infty} b(u)g_\epsilon(x-u) du$ , then we have the same order of the convergence as the above (11) with  $b_\epsilon \in C_b^\infty(\mathbb{R})$  and  $\|b_\epsilon\|_{H^\alpha(\mathbb{R})} = \mathcal{O}(\epsilon^{-\alpha/2})$ ,  $\|\partial_x^k b_\epsilon\|_{L^p(\mathbb{R})} = \mathcal{O}(\epsilon^{-\frac{1}{2}(k-\frac{1}{p})_+})$ , for any  $k \in \mathbb{N}$ .

The next step is to consider any general subset  $D \subset \mathbb{R}^d$  and consider approximations of  $\mathbf{1}_D(x)$ . This problem is already extremely difficult and has generated a long list of articles where the geometry of the set  $D$  plays an important role in order to determine its possible approximations. For example, one may refer to [11], Chapter 5.9 where the discussion is concentrated on functions  $b$  which are density functions and  $pd = 1$ . In fact, the literature on kernel density estimation is extremely large and deals with the particular case of functions  $b$  which are density functions.

For example, [38], uses the kernel estimators based on shift invariant spaces to approximate functions in  $W_p^k(H)$ . They provide uniform bounds on the optimal  $L^p(H)$  distance depending on the kernel regularization parameter  $\epsilon$  and the  $W_p^k(H)$  norm of the function to approximate if the so-called Strang-Fix conditions are satisfied for the generator of the shift invariant space.

Starting from this result, many other researchers have studied other classes of objective functions to approximate and subclasses of functions (replacing the kernel estimators based on shift invariant spaces) that are used to build the approximations have been studied in the literature. A high number of these results are found in journals on approximation theory. Just as an example, we refer to [27].

In the setup of Fourier analysis, this topic is also a classical topic as one wants to know what is the error committed when approximating the full infinite order Fourier series by a finite order one. See for example, Chapter 6 in [37].

A different approach, which may seem unnatural at the beginning is the use of wavelet theory results. In wavelet theory, one actually approximates regular functions using discontinuous functions which is the reverse of the current situation here. Still, theoretical results in this area seem to be useful due to their generality. This is a subject that needs to be further deepened by experts in this area. For more references on this matter, see [1, Eq. (84)], [13, 14] or [41] for the relation of this problem with Besov spaces.

The above theoretical developments are only a part of the large literature of the theory of function approximation. From the above, one can also guess that the rate of  $d_p(b, b_\epsilon)$  will vary greatly depending on the dimension, the irregularity of the function  $b$  and the particular norm considered.

In order to be able to give a comparative set of results, in what follows, we will make the following assumption for the approximation sequence  $b_\epsilon$  in the rest of the article. For this, we define,  $b_\epsilon^1 := b_\epsilon^* a^{-1} b_\epsilon$ ,  $\mathbf{b}_\epsilon = (b_\epsilon, b_\epsilon^1)$ .

**Hypothesis.** When  $d = 1$ , the triplet  $(\mathbf{b}_\epsilon, \mathbf{b}, Y)$  satisfied for  $i \in \{1, 2\}$ : For any  $\alpha > 0$  small enough, there exists  $p \geq 1$

and  $q, r \in (p \vee i, \infty]$  with  $\frac{1}{2q} + \frac{1}{r} < \frac{1}{p \vee i}$  such that

$$\begin{aligned} \|b - b_\epsilon\|_{L^{r,q}(H)} &\leq C\epsilon^{\frac{1}{2q}-\alpha}, \\ \|\mathbf{b}_\epsilon\|_{\mathbb{H}^k(\mathbb{R})} &= O(\epsilon^{-k/2}), \quad \forall k \in \mathbb{N} \\ \text{and } \|\partial_x^k \mathbf{b}_\epsilon\|_{L^{r,q}(H)} &\leq \begin{cases} C\epsilon^{-\frac{k}{2}+\frac{1}{2}-\alpha}, & \text{if } i = 2, \\ C\epsilon^{-(\frac{k}{2}-1+\alpha)_+}, & \text{if } i = 1 \end{cases} \quad \forall k \in \mathbb{N}. \end{aligned} \tag{H4}$$

Recall that the Gaussian upper bound estimates are satisfied as stated in (9) and (24). This assumption will simplify the understanding of our results and avoid technical issues which have been discussed before Lemmas 1 and 2 as well as in Lemma 13 in the Appendix.

## 7 Weak method to obtain the rates of convergence of $d_f(X^\epsilon, \bar{X}^\epsilon)$

We now exhibit some situations where the assumptions of Corollary 1 hold. We will therefore assume hypothesis (H4) and we try to obtain an estimate of weak rate of convergence under different spaces  $\mathfrak{F}$  and  $\mathfrak{M}$ . In particular, we will always have that for  $f \in \mathfrak{F}$  and that for the function  $b$  there exists a sequence in  $b_\epsilon \in \mathfrak{M}$  so that  $K_{p'}(f) < \infty$  and  $d_p(b, b_\epsilon) \leq C\epsilon^\gamma$  for some  $\gamma > 0$ .

### 7.1 Case of smooth coefficients

The results in [32] require the coefficients to be Hölder continuous. Of course, the convergence rate is better for smooth coefficients. But in order to achieve a rate asymptotically equal to  $n^{-1}$ , we need that  $a \in \mathbb{H}^{\alpha/2, \alpha}(\bar{H})$  with  $\alpha > 2$  and a terminal condition  $f \in \mathbb{H}^{2+\alpha}(\mathbb{R}^d)$ . That is,  $f$  is at least four times differentiable.

**Lemma 3.** Assume that  $\sigma \in \mathfrak{H}(\lambda, \Lambda)$ .

Let  $f \in C_{\text{SI}}(\mathbb{R}^d)$ . Then for any integer  $k$ , there exists a function  $B_{\text{SI}}(f, x)$  such that

$$\|\mathbb{E}[f(\bar{Y}_T)^k]\|^{1/k} + \|\mathbb{E}[f(Y_T)^k]\|^{1/k} \leq B_{\text{SI}}(f, x) \text{ with } \sup_{x \in \mathbb{R}^d} B_\epsilon(f, x) e^{-C|x|^2} < +\infty \text{ for all } C > 0.$$

*Proof.* These inequalities follow immediately from (9) applied to the densities  $\bar{\mathbf{p}}(t, x, y)$  and to  $\mathbf{p}(t, x, y)$ .  $\square$

For the next proposition we define  $S(g)(s, x) := \frac{g(s, x) - g(\phi(s), x)}{s - \phi(s)}$ ,  $\mathcal{S}(g)(s, x) := \sup_{t \in (\phi(s), s)} |S(g)(t, x)|$  and the following measure of irregularity for a function  $g : H \rightarrow \mathbb{R}^d$ :

$$\|g\|_{r, q, n} := \|g\|_{L^{r,q}(H)} + \frac{1}{n} |g(0, x)|.$$

**Proposition 3.** Assume  $\sigma \in \mathfrak{D} := C_b^{1,3}(\bar{H}) \cap \mathfrak{S}(\lambda, \Lambda)$ ,  $b, b_\epsilon \in \mathfrak{B}(\Lambda)$  with  $b_\epsilon \in \mathfrak{M} := C_b^{1,3}(\bar{H}) \cap \mathfrak{B}(\Lambda)$  and  $f$  in  $C_{\text{SI}}^3(\mathbb{R}^d)$ . Then for any  $p > 1$  and any constants  $q, r \in (p \vee 2, \infty]$ ,  $q_1, r_1 \in (p, \infty]$  such that  $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p \vee 2}$  and  $\frac{d}{2q_1} + \frac{1}{r_1} < \frac{1}{p}$ , there exists a constant  $C$  which does not depend on the function  $b_\epsilon$  but depends upon  $p, q, r, q_1, r_1, d, T, f, \Lambda$  and  $\sigma$ , such that

$$\|\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\bar{X}_T^\epsilon)]\| \leq \frac{C(\|b_\epsilon\|_\infty)}{n} (1 + \|\partial_x^1 \mathbf{b}_\epsilon\|_{r_1, q_1, n}) P_3 \left( \|\partial_x^1 \mathbf{b}_\epsilon\|_{r, q, n}, \|\partial_x^2 \mathbf{b}_\epsilon\|_{r, q, n}, \|\partial_x^3 \mathbf{b}_\epsilon\|_{r, q, n}, \|\mathcal{S}(\mathbf{b}_\epsilon)\|_{r, q, n} \right).$$

Furthermore  $P_3(x_1, \dots, x_4)$  is an explicit polynomial of order 3 in the variables  $(x_1, \dots, x_4)$  which also satisfies that any product of the type  $x_{i_1} x_{i_2} x_{i_3}$  appearing in  $P_3$  has  $0 \leq i_1 + i_2 + i_3 \leq 3$ ,  $i_1, i_2, i_3 \in \{0, \dots, 3\}$  (we let  $x_0 \equiv 1$ ) and any product containing  $x_4$  can have only a term  $x_1$  or a constant multiplying it.

*Proof.* As we need to put both the solution process and the approximation process in the same probability space we re-define the processes as follows: Let  $\Gamma_s^\epsilon := \sigma^{-1} b_\epsilon(s, X_s^\epsilon)$  and  $\bar{\Gamma}_s^\epsilon := \sigma^{-1} b_\epsilon(\phi(s), \bar{X}_{\phi(s)}^\epsilon)$ . Accordingly, we define

the measures  $\mathbb{Q}^{\Gamma^\epsilon}$  and  $\mathbb{Q}^{\tilde{\Gamma}^\epsilon}$  as in (7). Then under  $\mathbb{Q}^{\Gamma^\epsilon}$ ,  $X^\epsilon \stackrel{\mathcal{L}}{=} Y$  and under  $\mathbb{Q}^{\tilde{\Gamma}^\epsilon}$ ,  $\bar{X}^\epsilon \stackrel{\mathcal{L}}{=} \bar{Y}$ . Therefore we consider the following decomposition:

$$\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\bar{X}_T^\epsilon)] = \mathbb{E}[\mathcal{Z}(\Gamma^\epsilon)_T f(Y_T)] - \mathbb{E}[\mathcal{Z}(\tilde{\Gamma}^\epsilon)_T f(\bar{Y}_T)]. \quad (12)$$

We can consider the above problem as follows. For  $y \in \mathbb{R}^d$  and  $z \in \mathbb{R}$ , define

$$g(y, z) = \exp(z)f(y)$$

and let the system to approximate be  $(Y, V^\epsilon) := (Y, \log(\mathcal{Z}(\Gamma^\epsilon)))$ . This system is approximated using  $(\bar{Y}, \log(\mathcal{Z}(\tilde{\Gamma}^\epsilon)_T))$  and becomes a diffusion because

$$V_t^\epsilon = \int_0^t (\sigma^{-1} b_\epsilon)^*(s, Y_s) dW_s + \frac{1}{2} \int_0^t b_\epsilon^* a^{-1} b_\epsilon(s, Y_s) ds.$$

The theory developed in Section 4.1 of [18] allows us to write and analyze the error (12) in general. As this methodology has been explained in detail in that reference, we only sketch briefly the proof here. In the present case, we have to perform the same analysis as in Section 4.1 of [18] yet we have to find the exact dependence on  $b_\epsilon$  and its derivatives of each term in the error expansion of the difference  $g(Y_T, \log \mathcal{Z}(\Gamma^\epsilon)_T) - g(\bar{Y}_T, \log \mathcal{Z}(\tilde{\Gamma}^\epsilon)_T)$ .

The first step in that methodology consists of proving the following error decomposition

$$\begin{aligned} g(Y_T, \log \mathcal{Z}(\Gamma^\epsilon)_T) - g(\bar{Y}_T, \log \mathcal{Z}(\tilde{\Gamma}^\epsilon)_T) &= \sum_{i,j=0}^d c_{i,j}^0(T) \int_0^T c_{i,j}^1(t) \int_{\phi(t)}^t c_{i,j}^2(s) dW_s^i dW_t^j \\ &\quad + \sum_{i,j,k=0}^d c_{i,j,k}^0(T) \int_0^T c_{i,j,k}^1(t) \left[ \int_0^t c_{i,j,k}^2(s) \left( \int_{\phi(s)}^s c_{i,j,k}^3(u) dW_u^i \right) dW_s^j \right] dW_t^k, \end{aligned} \quad (13)$$

for some explicit adapted processes  $\{(c_{i,j}^1(t), c_{i,j,k}^2(t))_{t \geq 0} : 0 \leq i, j, k \leq d, 0 \leq i_1 \leq 2, 0 \leq i_2 \leq 3\}$ , where by abuse of notation,  $dW_t^0$  denotes  $dt$ . In the above expansion, the terms  $c_{i,j}^1, c_{i,j,k}^2$  are computed in the same spirit as in the proof of Lemma 4.2 in [18].

The explicit calculations are long and tedious. Therefore we will only give here the dependence of each of the above terms on  $b_\epsilon$  and its derivatives:

- 1)  $c_{i,j}^0$  and  $c_{i,j,k}^0$  depend linearly on  $\exp(\theta(\log(\mathcal{Z}(\Gamma^\epsilon)_T) + (1 - \theta)(\log(\mathcal{Z}(\tilde{\Gamma}^\epsilon)_T)))$  with a random coefficient and no constant term. Therefore these terms depend on  $b_\epsilon^* \sigma^{-1}$  and  $b_\epsilon^1$ . Here  $\theta$  denotes a uniform  $[0, 1]$  random variable independent of all the other variables. To analyze these terms, Lemma 10 is used.
- 2)  $c_{i,j,k}^1$ , for  $k \neq 0$ , linearly depends on  $\nabla(b_\epsilon^* \sigma^{-1})$  with random coefficients, which are independent of  $b_\epsilon$ .
- 3)  $c_{i,j,0}^1$  depends linearly on  $\partial_t b_\epsilon^1$ .
- 4)  $c_{i,j}^1$  are linear functions of  $\nabla(b_\epsilon^* \sigma^{-1}), \partial_t(b_\epsilon^* \sigma^{-1}), \nabla b_\epsilon^1$  and  $\partial_t b_\epsilon^1$ .
- 5)  $c_{i,j}^2, c_{i,j,k}^2$  and  $c_{i,j,k}^3$  do not depend on  $b_\epsilon$ .

All the random coefficients mentioned above do not depend on  $b_\epsilon$  but only depend upon  $\sigma, f$  and their derivatives up to order 3.

In order to finish the argument and obtain the polynomial function  $P_3$  one follows the same steps as in the proof of Theorem 4.2 in section 4.5.1 in [18]. From there one obtains the result by appropriately using Hölder's inequality together with Lemmas 1 and 2. Some of the needed Gaussian upper bound estimates have already been mentioned previously and the rest are obtained in Lemma 13 in the Appendix. In particular, the derivatives of  $b_\epsilon$  appear due to the use of the iterative use of the duality relation of Malliavin Calculus in (13). For this reason, one needs the regularity assumptions on  $f, \sigma$  and  $b_\epsilon$ .  $\square$

*Remark 4.* We recall that in order to obtain the decomposition (13) is important the description of the so called error process  $\mathcal{E}_t := Y_t - \bar{Y}_t$  which may be written as follows:

$$\mathcal{E}_t = \sum_{i=1}^d \int_0^t (\alpha_i(s) \mathcal{E}_s)^* dW_s^i + \int_0^t G_s dW_s,$$

where  $G_s := \sigma(s, \bar{Y}_s) - \sigma(\phi(s), \bar{Y}_{\phi(s)})$ , and for  $i = 1, \dots, d$ ,  $\alpha_i(s)$  is a  $d \times d$ -matrix whose  $(j, k)$ th-component is

$$\alpha_i^{j,k}(s) = \int_0^1 \partial_k \sigma_{j,i}(s, \theta Y_s + (1 - \theta) \bar{Y}_s) d\theta.$$

Now we can reach a first global conclusion. There are many ways of combining the results of previous sections and therefore we show only one of those possible combinations. We will use the same setting in future conclusions.

*Conclusion 1.* Let  $d = 1$ ,  $\sigma \in \mathfrak{D} := C_b^{1,3}(\bar{H}) \cap \mathfrak{S}(\lambda, \Lambda)$ ,  $\mathfrak{M} = C_b^{1,3}(\bar{H})$ ,  $\mathfrak{F} = C_{\text{SI}}^3(\mathbb{R})$ .

First, with Proposition 1 and Lemma 1 with parameters  $q_0 = p_0 \vee 2$ , we obtain a rate of  $\epsilon^{1/2(p_0 \vee 2)}$  for any  $p_0 > 1$ . Therefore  $\gamma = 1/4$  in Theorem 1.

Next, we use Hypothesis (H4) with respect to all triplets corresponding to the functions in  $\mathbf{b}_\epsilon$  where the processes appearing in Proposition 3 may be  $Y, \bar{Y}, Y_t^\alpha, \bar{Y}_t^\alpha$ .

From Proposition 3 we obtain that the rate depends on the norms  $\|\partial_x^i \mathbf{b}_\epsilon\|_{r,q,n}$ . These norms are bounded according to hypothesis (H4). Note that  $\partial_t \mathbf{b}_\epsilon = 0$  and that the properties of the polynomial  $P_3$  determine the rate. Finally this gives that the rate is given by  $\epsilon^{-1-\rho} n^{-1} + \epsilon^{-2} n^{-2}$  for any  $\rho > 0$  small enough. Note that the term  $\epsilon^{-2} n^{-2}$  comes from the estimate of  $\frac{1}{n} |g(0, x)|$  in the norm  $\|g\|_{r,q,n}$  for each of the derivatives of  $\mathbf{b}_\epsilon$ .

Then we can take  $\beta = 1 + \rho, \delta = 1$  in Theorem 1. Therefore the optimal rate becomes  $\kappa = 1 - \frac{1+\rho}{1/4+1+\rho}$ .

As a conclusion, we have that the final rate achieved is as close as  $n^{-1/5}$  as desired if we let  $\rho \rightarrow 0$ . But this is achieved with a constant that will go to infinity.

Note that this result improves the one in [32] when  $\alpha < \frac{2}{3}$ .

*Remark 5.* Various other conclusions of the type above can be reached using Proposition 2 or other types of weak approximations results such as the ones in [32].

## 8 Strong method to obtain the rate of convergence for $d_f(X^\epsilon, \bar{X}^\epsilon)$

In all the previous sections, we have considered situations where a smooth approximations for the irregular drift coefficient are used. Therefore naturally, control on the higher order derivatives are needed. In this section, we try to analyze the case when for some reasons, one wants to limit the number of higher order derivatives of  $b_\epsilon$  being considered. This requires a semi-strong type method and leads to different rates.

Let  $b_\epsilon \in \mathfrak{M} \subset \mathfrak{B}(\Lambda)$  be a family of approximations of  $b$  satisfying (H4). We also need stronger regularity conditions on  $\sigma$  so that  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  through all this section.

We still use the same definitions for  $X, \bar{X}, X^\epsilon, \bar{X}^\epsilon, Y$  and  $\bar{Y}$ .

We will also use an additional intermediate approximation process

$$\tilde{X}_t^\epsilon = x + \int_0^t \sigma(s, \tilde{X}_s^\epsilon) dB_s + \int_0^t b_\epsilon(\phi(s), \tilde{X}_{\phi(s)}^\epsilon) ds.$$

Our class of test functions is  $\mathfrak{F} := C_{\text{SI}}(\mathbb{R}^d)$ . We set  $K_q(f) := \mathbb{E}[|f(Y_T) - \mathbb{E}[f(Y_T)]|^q]^{1/q}$  for  $f \in \mathfrak{F}$ . For any two processes  $X^1$  and  $X^2$  we define

$$D_{g,p}(X^1, X_\phi^2) := \mathbb{E} \left[ \left( \int_0^T |g(s, X_s^1) - g(\phi(s), X_{\phi(s)}^2)|^2 ds \right)^{p/2} \right]^{1/p}.$$

In what follows, our application of Lemmas 1 and 2 and Hypothesis (H4) is restricted to the case  $i = 2$  for the rest of the article. This is due to the strong approach to the proofs.

Proposition 1 may be applied even for a non-anticipative drift and diffusion coefficient as the lemmas used in their proof are written in general form in the Appendix. Under the above conditions, for  $(p, p')$  an almost conjugate pair,

$$|\mathbb{E}[f(Z^{(1,j)})] - \mathbb{E}[f(Z^{(2,j)})]| \leq C_1(p, p') \mathbb{E} \left[ \left( \int_0^T |\beta^{(1,j)}(s) - \beta^{(2,j)}(s)|^2 ds \right)^{p/2} \right]^{1/p} K_{p'}(f) \quad (14)$$

where  $(Z^{(i,j)}, \beta^{(i,j)})$  for  $i, j = 1, 2$  are defined as:

- (i)  $Z^{(1,1)} = X_T^\epsilon$  and  $\beta^{(2,1)}(s) = b_\epsilon(s, Y_s)$ .
- (ii)  $Z^{(2,1)} = \widetilde{X}_T^\epsilon$  and  $\beta^{(1,1)}(s) = b(\phi(s), Y_{\phi(s)})$ .
- (iii)  $Z^{(1,2)} = \overline{X}_T$  and  $\beta^{(1,2)}(s) = b(\phi(s), \overline{Y}_{\phi(s)})$ .
- (iv)  $Z^{(2,2)} = \widetilde{\overline{X}}_T^\epsilon$  and  $\beta^{(2,2)}(s) = b_\epsilon(\phi(s), \overline{Y}_{\phi(s)})$ .

This result is used in Lemma 4, Proposition 5 and Lemma 6 below to yield the following control.

**Proposition 4.** *Assume  $\sigma \in \mathfrak{H}(\lambda, \Lambda)$ ,  $b, b_\epsilon \in \mathfrak{B}(\Lambda)$ , and Hypothesis (H4) is satisfied for  $i = 2$ . For any  $p > 1$ ,  $r, q \in (p \vee 2, \infty]$  such that  $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p \vee 2}$  and  $(p, p')$  an almost conjugate pair, there exists a positive constant  $C$  depending only on  $p, r, q, \Lambda, \lambda$  and  $T$  such that for  $f \in C_{\text{Sl}}(\mathbb{R}^d)$*

$$\|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T^\epsilon)]\| \leq C \left( K_{p'}(f) \left( D_{b_\epsilon, p}(Y, Y_\phi) + \|b - b_\epsilon\|_{L^{r,q}(H)} \right) + \sqrt{\text{Var}(f(Y_T) - f(\overline{Y}_T))} \right).$$

*Remark 6.* The rate of the last term  $\sqrt{\text{Var}(f(Y_T) - f(\overline{Y}_T))}$  can be determined in particular cases. For example, if  $f$  is Lipschitz then a direct application of the strong rate of convergence shows that the rate of this term is  $n^{-1/2}$  when  $\sigma \in \mathfrak{H}(\lambda, \Lambda)$ . These rates may vary as it has been shown in [5] if  $f$  is a non-Lipschitz function.

The proof of this Proposition is divided in two parts. The first follows from Lemma 4 and the second from Proposition 5 which are proven below.

**Lemma 4.** *Under the conditions of Proposition 4, for  $p > 1$ , there exists a constant  $C$  which depends only on  $p, T, q, r \in (p \vee 2, \infty]$  and the constants  $C_1$  and  $C_2$  appearing in (9) such that for  $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p \vee 2}$  and  $(p, p')$  an almost conjugate pair,*

$$\|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]\| \leq CK_{p'}(f)\|b - b_\epsilon\|_{L^{r,q}(H)}.$$

*Proof.* This follows from Proposition 1 and Lemma 1. □

## 8.1 Rate of convergence of the Euler-Maruyama scheme with smooth approximations of the drift

Using Proposition 1 between  $X^\epsilon$  and  $\overline{X}^\epsilon$  we may also give a result with a lower rate of convergence but which is valid under a broader class of functions  $f$ .

**Proposition 5.** *Assume that  $\sigma \in \mathfrak{H}(\lambda, \Lambda)$ ,  $b, b_\epsilon \in \mathfrak{B}(\Lambda)$ . Fix an almost conjugate pair  $(p, p')$ . Then for  $f \in C_{\text{Sl}}(\mathbb{R}^d)$ , there exists a constant  $C = C(p, p', T)$  such that*

$$\|\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\overline{X}_T^\epsilon)]\| \leq CD_{b_\epsilon, p}(Y, Y_\phi)K_{p'}(f), \quad (15)$$

$$\|\mathbb{E}[f(\widetilde{X}_T^\epsilon)] - \mathbb{E}[f(\overline{X}_T^\epsilon)]\| \leq C \left( \left( D_{b_\epsilon, p}(Y_\phi, \overline{Y}_\phi) + D_{\sigma^{-1}, p}(Y_\phi, \overline{Y}_\phi) \right) K_{p'}(f) + \sqrt{\text{Var}(f(Y_T) - f(\overline{Y}_T))} \right). \quad (16)$$

*Proof.* Eq. (14) also holds for  $Z^{(1,1)} = X_T^\epsilon$  and  $Z^{(2,1)} = \widetilde{X}_T^\epsilon$  with an almost conjugate pair  $(p, p')$ , then (15) holds from Lemma 5(I).

To obtain (16), consider  $\Gamma_1(s) := \sigma^{-1}(s, Y_s)b_\epsilon(\phi(s), Y_{\phi(s)})$  and  $\Gamma_2(s) = \sigma^{-1}b_\epsilon(\phi(s), \overline{Y}_{\phi(s)})$ . Therefore,

$$\begin{aligned} \|\mathbb{E}[f(\widetilde{X}_T^\epsilon)] - \mathbb{E}[f(\overline{X}_T^\epsilon)]\| &= \|\mathbb{E}[\mathcal{Z}(\Gamma_1)_T f(Y_T)] - \mathbb{E}[\mathcal{Z}(\Gamma_2)_T f(\overline{Y}_T)]\| \\ &\leq \|\mathbb{E}[(\mathcal{Z}(\Gamma_1)_T - \mathcal{Z}(\Gamma_2)_T)f(Y_T)]\| + \|\mathbb{E}[\mathcal{Z}(\Gamma_2)_T(f(Y_T) - f(\overline{Y}_T))]\| =: A_1 + A_2. \end{aligned}$$

For an almost conjugate pair  $(p, p')$ , with Lemma 11 and combining (17) and (18) applied on  $\sigma^{-1}$  in  $H^{1/2,1}(\overline{H})$  thanks to Lemma 12,

$$\begin{aligned} A_1 &\leq C \mathbb{E} \left[ \left( \int_0^T |\sigma^{-1}(s, Y_s)b_\epsilon(\phi(s), Y_{\phi(s)}) - \sigma^{-1}b_\epsilon(\phi(s), \overline{Y}_{\phi(s)})|^2 ds \right)^{p/2} \right]^{1/p} K_{p'}(f) \\ &\leq C(p, \lambda, \Lambda) \left( D_{b_\epsilon, p}(Y_\phi, \overline{Y}_\phi) + D_{\sigma^{-1}, p}(Y_\phi, \overline{Y}_\phi) \right) K_{p'}(f). \end{aligned}$$



The control of  $A_2$  follows from the Cauchy-Schwarz inequality and Lemma 9(I), so that

$$A_2 \leq C \sqrt{\text{Var}(f(Y_T) - f(\bar{Y}_T))}.$$

From here the result follows.  $\square$

We now estimate the norms in Proposition 4 using similar techniques as in Lemmas 1 and 2. For this, recall the notation  $S(g)(s, x) = \frac{g(s, x) - g(\phi(s), x)}{s - \phi(s)}$ .

**Lemma 5.** *Assume  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$ . For any  $p > 1$ ,  $r, q \in (p \vee 2, \infty]$  such that  $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p\sqrt{2}}$ , then there exists some positive constant  $C = C(p, \Lambda, T, \rho)$  such that for any  $\rho \in (0, 1/2)$ ,*

(I) *It holds that*

$$D_{g,p}(Y, Y_\phi) \vee D_{g,p}(\bar{Y}, \bar{Y}_\phi) \leq \frac{C}{n^{1/2-\rho}} (\|S(g)\|_{L^{r,q}(H)} + \|\partial_x g(\phi(\cdot), \cdot)\|_{L^{r,q}(H)}). \quad (17)$$

(II) *Suppose that for any  $\alpha \in [0, 1]$ ,  $Y_t^\alpha = \alpha Y_t + (1 - \alpha)\bar{Y}_t$  satisfies the Gaussian upper bound (9). Then*

$$D_{g,p}(Y, \bar{Y}) \leq \frac{C\|\partial_x g\|_{L^{r,q}(H)}}{\sqrt{n}}. \quad (18)$$

*Proof.* Without loss of generality, we may assume that  $p > 2$ .

(I) For example, using Minkowski's and Hölder's inequalities with a conjugate pair  $(a, a')$  together with Lemma 1,

$$\begin{aligned} & \mathbb{E} \left[ \left( \int_0^T |g(s, Z_s) - g(\phi(s), Z_{\phi(s)})|^2 ds \right)^{p/2} \right]^{1/p} \leq Cn^{-1/2} \mathbb{E} \left[ \left( \int_0^T \frac{|g(s, Z_s) - g(\phi(s), Z_{\phi(s)})|^2}{s - \phi(s)} ds \right)^{p/2} \right]^{1/p} \\ & + \mathbb{E} \left[ \left( \int_0^T \left( \frac{|g(\phi(s), Z_s) - g(\phi(s), Z_{\phi(s)})|}{|Z_s - Z_{\phi(s)}|} \right)^2 ds \right)^{ap/2} \right]^{1/ap} \mathbb{E} \left[ \sup_{s \in [0, T]} |Z_s - Z_{\phi(s)}|^{a'p} \right]^{1/a'p} \\ & \leq Cn^{-1/2} \left( \|S(g)\|_{L^{r_1, q_1}(H)} + \|\partial_x g(\phi(\cdot), \cdot)\|_{L^{r_2, q_2}(H)} \mathbb{E} \left[ \sup_{s \in [0, T]} |Z_s - Z_{\phi(s)}|^{a'p} \right]^{1/a'p} \right). \end{aligned} \quad (19)$$

Here  $\frac{d}{2q_1} + \frac{1}{r_1} < \frac{1}{p\sqrt{2}}$  and  $\frac{d}{2q_2} + \frac{1}{r_2} < \frac{1}{(ap)\sqrt{2}}$ .

In order to compute the expectation  $\mathbb{E} \left[ \sup_{s \in [0, T]} |Z_s - Z_{\phi(s)}|^b \right]$  for  $b > 0$ , let us introduce the function  $\bar{F}(x) := \mathbb{P} \left[ \sup_{s \in [0, T]} |Z_s - Z_{\phi(s)}|^b > x \right]$ . Then for any  $\rho \in (0, \frac{1}{2})$ ,

$$\mathbb{E} \left[ \sup_{s \in [0, T]} |Z_s - Z_{\phi(s)}|^b \right] = \int_0^{n^{-b(\frac{1}{2}-\rho)}} \bar{F}(x) dx + \int_{n^{-b(\frac{1}{2}-\rho)}}^\infty \bar{F}(x) dx \leq n^{-b(\frac{1}{2}-\rho)} + \int_{n^{-b(\frac{1}{2}-\rho)}}^\infty \bar{F}(x) dx.$$

For  $x \in [n^{-b(\frac{1}{2}-\rho)}, \infty)$ , we use the Dubins-Schwarz theorem to reduce the consideration of  $\bar{F}$  to the explicit law of the maximum of increments of Brownian motion  $\bar{G}(x) = \mathbb{P}(|W_1| > x)$ . That is,

$$\bar{F}(x) \leq n \mathbb{P} \left[ \sup_{s \in [0, \frac{\Lambda^2 T}{n}] } |W_s|^b > x \right] \leq \bar{G} \left( \frac{x^{1/b} \sqrt{n}}{\Lambda \sqrt{T}} \right).$$

Using the classical estimate  $\bar{G}(x) \leq e^{-c|x|^2}$  for  $x > 0$  for some positive constant  $c$ ,

$$\int_a^\infty \bar{F}(x) dx \leq C \int_{n^{-b(\frac{1}{2}-\rho)}}^\infty n e^{-c x^{2/b} n} dx \leq C n^{1-\frac{b}{2}} \int_{n^{b\rho}}^\infty e^{-cu^{2/b}} du.$$

The conclusion follows as the above integral converges to zero faster than  $n^{-b(\frac{1}{2}-\rho)}$  as  $n \rightarrow \infty$ . Hence the result follows by taking  $a$  sufficiently close to 1 in (19).

(II) The proof is similar as the above proof except that one uses the strong rate of convergence (see, e.g., [23]) which yields that  $\sup_{t \in [0, T]} \mathbb{E}[|Y_t - \bar{Y}_t|^2]^{1/2} \leq C/\sqrt{n}$ . Hence the result.  $\square$

Now using Lemma 5 and Lemma 3 on Proposition 4, we can give the convergence rate result.

**Theorem 3.** *Under the same hypothesis of Proposition 4 and  $f \in C_{\text{SI}}^1(\mathbb{R}^d)$ , there exists a positive constant  $C$  which depends only on  $q, r \in (p \vee 2, \infty]$ ,  $T, \lambda, \Lambda$  and the constants  $C_1$  and  $C_2$  appearing in (9) such that for some  $p > 1$ ,  $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p\sqrt{2}}$ , for any  $\rho \in (0, 1/2)$ ,*

$$\begin{aligned} |\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T^\epsilon)]| &\leq CK_p'(f) \left( \frac{1}{n^{1/2-\rho}} (\|S(b_\epsilon)\|_{L^{r,q}(H)} + \|\partial_x b_\epsilon(\phi(\cdot), \cdot)\|_{L^{r,q}(H)}) + \|b - b_\epsilon\|_{L^{r,q}(H)} \right) \\ &\quad + C \sqrt{\text{Var}(f(Y_T) - f(\bar{Y}_T))}. \end{aligned}$$

*Conclusion 2.* Let  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$ ,  $\mathfrak{W} = H^{1/2,1}(\bar{H})$ ,  $\mathfrak{F} = C_{\text{Lip}}(\mathbb{R}^d)$ , the space of Lipschitz functions. Assume that  $b$  is time-homogeneous, so that  $S(b_\epsilon) = 0$ . According to (H4) and Remark 6, the above rate in Theorem 3 is  $\epsilon^{-\left(\frac{1}{2}(1-\frac{1}{q}\right)} n^{-\left(\frac{1}{2}-\rho\right)} + \epsilon^{1/q}$ .

Applying Theorem 1, we achieve a rate close to  $n^{-1/3}$  by taking  $q$  close to 2. Note that for  $d \geq 2$ ,  $q$  can only be taken close to  $d$  which gives a dimension related rate of  $n^{-1/(d+1)}$ .

## 8.2 The weak error of the Euler-Maruyama scheme for $X$

We now give estimates on the distance  $d_f(\bar{X}^\epsilon, \bar{X})$ .

**Lemma 6.** *Assume  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  and  $b, b_\epsilon \in \mathfrak{B}(\Lambda)$ . Fix  $p \geq 1$ . Then there exists a constant  $C$  which depends only on  $q, r \in (p \vee 2, \infty]$ ,  $\Lambda, \lambda, T$  and the constants  $C_1$  and  $C_2$  appearing in Lemma 1 such that for some almost conjugate pair  $(p, p')$ ,  $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p\sqrt{2}}$ , we have*

$$|\mathbb{E}[f(\bar{X}_T)] - \mathbb{E}[f(\bar{X}_T^\epsilon)]| \leq CK_{p'}(f) \left( \|b - b_\epsilon\|_{L^{r,q}(H)} + \frac{1}{n^{1/2}} \right).$$

*Proof.* We apply (14) with  $Z^{(1,2)} = \bar{X}_T$  and  $Z^{(2,2)} = \bar{X}_T^\epsilon$  so that

$$|\mathbb{E}[f(\bar{X}_T)] - \mathbb{E}[f(\bar{X}_T^\epsilon)]| \leq C(p, p') K_{p'}(f) \mathbb{E} \left[ \left( \int_0^T |b(\phi(s), \bar{Y}_{\phi(s)}) - b_\epsilon(\phi(s), \bar{Y}_{\phi(s)})|^2 ds \right)^{p/2} \right]^{1/p}.$$

Therefore the result is a direct consequence of Lemma 2.  $\square$

We now deal with the case where the set of irregularity of  $b$  can be characterized. In order to introduce our main condition, we start with the following definition. For a set  $G \subset \mathbb{R}^d$ , we define  $G(\epsilon) = \{x \in \mathbb{R}^d | d(x, G) \leq \epsilon\}$ , where  $d(x, G) = \inf_{y \in G} |x - y|$  is the distance between  $x$  and  $G$ .

**Theorem 4.** *Suppose that  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$  and  $b \in \mathfrak{B}(\Lambda)$ . We assume moreover that for each  $\epsilon > 0$ , we can define an approximation  $b_\epsilon := b \mathbf{1}_{G(\epsilon)^c} + b_\epsilon \mathbf{1}_{G(\epsilon)}$  so that  $b_\epsilon$  is a Lipschitz function on  $G(\epsilon)^c$  with Lipschitz constant independent of  $\epsilon$ , the Lipschitz constant of  $b_\epsilon$  on  $G(\epsilon)$  is  $\epsilon^{-1}$  and  $\text{meas}(G(\epsilon)) = O(\epsilon^\kappa)$  for some  $\kappa > 0$ . Then for any  $f \in C_{\text{SI}}(\mathbb{R}^d)$ ,  $p > 1$ ,  $q > p \vee 2$  such that  $\frac{d}{2q} < \frac{1}{p\sqrt{2}}$ , there exists a positive constant  $C$  such that*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]| \leq C \left( n^{-\frac{1}{2} \frac{(\kappa, p)p}{qp + (\kappa, p)(p-q)}} + \sqrt{\text{Var}(f(Y_T) - f(\bar{Y}_T))} \right).$$

*Proof.* Combining Proposition 4 with Lemmas 4 and 6, we obtain that for any choice of an almost conjugate pair  $(p, p')$ ,  $q, r \in (p \vee 2, \infty]$  such that  $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p\sqrt{2}}$ , there exists constant  $C = C(p, p', q, r, \Lambda, T)$  such that

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]| \leq C \left( K_{p'}(f) (D_{b_\epsilon, p}(Y, Y_\phi) + \|b - b_\epsilon\|_{L^{r,q}(H)}) + \sqrt{\text{Var}(f(Y_T) - f(\bar{Y}_T))} + \frac{K_{p'}(f)}{n^{1/2}} \right).$$

From our hypotheses on  $b_\epsilon$ ,  $\|b - b_\epsilon\|_{L^{r,q}(H)} = O(\epsilon^{\kappa/q})$ .

In order to study  $D_{b_{\epsilon,p}}(Y, Y_\phi)$ , we remark that on the set  $G(\epsilon)^c$ ,  $b_\epsilon$  is Lipschitz with a bounded constant. Then, for  $\kappa \leq p$

$$\begin{aligned} \mathbb{E} \left[ \int_0^T |b_\epsilon(Y_s) - b_\epsilon(Y_{\phi(s)})|^p ds \right] &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \int_0^T |b_\epsilon(z) - b_\epsilon(y)|^p \mathbf{p}(s - \phi(s), y, z) \mathbf{p}(\phi(s), x, y) dy dz \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \int_0^T \mathbf{1}(y, y+z \in G(\epsilon)^c) |z-y|^p \mathbf{p}(s - \phi(s), y, z) \mathbf{p}(\phi(s), x, y) dy dz \\ &\quad + \epsilon^{-p} \int_{\mathbb{R}^d \times \mathbb{R}^d} \int_0^T \mathbf{1}(y \text{ or } y+z \in G(\epsilon)) |z-y|^p \mathbf{p}(s - \phi(s), y, z) \mathbf{p}(\phi(s), x, y) dy dz \\ &\leq Cn^{-p/2} (1 + \epsilon^{-p+\kappa}). \end{aligned}$$

The last step uses the Gaussian upper bound estimate (9) together with the estimate

$$|z-y|^p \mathbf{g}_{C(s-\phi(s))}(z-y) \leq C_1(s-\phi(s))^{p/2} \mathbf{g}_{C_1(s-\phi(s))}(z-y).$$

Therefore, we obtain that  $D_{b_{\epsilon,p}}(Y, Y_\phi) \leq Cn^{-1/2} \epsilon^{-1+\frac{\kappa}{p}}$ .

Hence, the weak error of the Euler-Maruyama scheme is bounded by  $\frac{C}{n^{1/2} \epsilon^{1-\frac{\kappa}{p}}} + C' \epsilon^{\frac{\kappa}{q}}$  for some constants  $C$  and  $C'$  that depend only on  $\Lambda, T, p, q, r$  and  $f$ . Optimizing over the choice of  $\epsilon$  and taking  $r \rightarrow \infty$  this leads to the result using Theorem 1. If  $\kappa > p$  then the rate is  $\frac{C}{n^{1/2}} + C' \epsilon^{\frac{\kappa}{q}}$ . Then the rate is  $n^{-1/2}$ .  $\square$

*Remark 7.* One may think of other examples of discontinuity sets that do not satisfy the assumptions of Theorem 4 but which may be treated with the same ideas as in the proof of the above Theorem. These are the cases of unbounded sets  $G$ .

*Conclusion 3.* Considering the proof carefully, one may say that the approximations of the drift are taken on the space  $\mathfrak{M} = \mathbf{H}^1(\mathbb{R}^d)$ . If  $\sigma \in \mathfrak{S} = \mathfrak{S}(\lambda, \Lambda)$  and  $\mathfrak{F} = C_{\text{Sl}}^1(\mathbb{R}^d)$  then for  $\kappa > 1$ , the convergence rate is  $n^{-1/2}$ . This is obtained taking  $p = \kappa$  and  $\frac{d}{2q} < \frac{1}{\kappa\sqrt{2}}$ . In the case that  $\kappa \leq 1$ , the rate is as close as desired to  $n^{-\frac{\kappa}{2(d(1-\kappa)+\kappa)}}$  which is obtained by taking  $p$  close to 1 and  $q$  close to  $d$ .

### 8.3 A particular example with a weak approximation rate of $n^{-1}$

We have seen through the previous Sections that in our proofs the rate of convergence of the Euler-Maruyama scheme is worse than the classical  $n^{-1}$  due to the irregularity of the drift coefficient  $b$ .

On the other hand, in this section, we show that a rate of convergence of order 1 could be achieved in some cases. This example which is of theoretical interest rather than a practical one, shows that a remaining problem is to determine the class of drift coefficients for which the rate  $n^{-1}$  is preserved. As this is an example, we do not give all details of the proof.

For this, we consider the following family of 1-dimensional sde

$$dX_t(x) = b(X_t(x)) dt + dW_t, \quad X_0(x) = x, \quad \text{with } b(x) := \begin{cases} -\theta, & x > 0, \\ 0, & x = 0, \\ \theta, & x < 0. \end{cases}$$

for  $\theta > 0$ . As before  $\bar{X}(x)$  denotes the Euler-Maruyama scheme associated to  $X(x)$ .

This process is called a Brownian motion with two-valued, state-dependent drift, which is related to a stochastic control problem. For more details on it, see Karatzas and Shreve [22, Section 6.5]. From Karatzas and Shreve [22, Section 6.5, (5.14), p. 441], the transition density function of a Brownian motion of  $X_t(x)$ ,  $x \geq 0$  is

$$\mathbf{p}_t(x, z) = \begin{cases} h_-(x-z, t) + \frac{\theta g_-(z)}{\sqrt{2\pi t}} \int_{x+z}^{\infty} \exp\left(-\frac{(v-\theta t)^2}{2t}\right) dv & \text{for } z > 0, \\ g_+(x) h_+(x-z, t) + \frac{\theta g_+(z)}{\sqrt{2\pi t}} \int_{x-z}^{\infty} \exp\left(-\frac{(v-\theta t)^2}{2t}\right) dv & \text{for } z \leq 0. \end{cases} \quad (20)$$

Here, we have used the functions  $h_\pm(x, t) = \exp\left(-\frac{(x \pm \theta t)^2}{2t}\right) / \sqrt{2\pi t}$  and  $g_\pm(x) = \exp(\pm 2\theta x)$ .

**Theorem 5.** For  $f \in C_p^3(\mathbb{R})$  an even function, the weak error satisfies

$$|\mathbb{E}[f(X_T(0))] - \mathbb{E}[f(\bar{X}_T(0))]| \leq \frac{C}{n},$$

where  $C$  is a positive constant.

We remark that from the analytic point of view the fact that the starting point for the diffusion is zero is strongly used through the symmetry of the density of the associated Euler-Maruyama scheme.

In order to prove this result we need the following lemma.

**Lemma 7.** (i). Assume that  $f \in C_p^2(\mathbb{R}; \mathbb{R})$  is an even function. Define  $u(t, x) = \mathbb{E}[f(X_{T-t}(x))]$ ,  $0 \leq t \leq T$  and  $x \in \mathbb{R}$ . Then for  $k = 0, 1, 2$ ,

$$\sup_{t \in [0, T]} \sup_{x \in \mathbb{R}} |\partial_x^k u(t, x)| \leq C, \quad (21)$$

where  $C$  is a positive constant. Furthermore,  $u$  satisfies the following PDE

$$\frac{\partial u}{\partial t}(t, x) + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(t, x) + b(x) \frac{\partial u}{\partial x}(t, x) = 0, \quad u(T, x) = f(x). \quad (22)$$

(ii). Assume that  $f \in C_p^3(\mathbb{R}; \mathbb{R})$  is an even function. Define  $u(t, x) = \mathbb{E}[f(X_{T-t}(x))]$ ,  $0 \leq t \leq T$  and  $x \in \mathbb{R}$ . Then

$$\sup_{t \in [0, T]} \sup_{x \in \mathbb{R} \setminus \{0\}} |\partial_x^3 u(t, x)| \leq C, \quad \sup_{t \in [0, T]} |\partial_{x+} \partial_x^2 u(t, 0)| \leq C \quad \text{and} \quad \sup_{t \in [0, T]} |\partial_{x-} \partial_x^2 u(t, 0)| \leq C,$$

where  $C$  is a positive constant and  $\partial_{x+}$ ,  $\partial_{x-}$  denote the right and left derivative in  $x$  respectively.

*Proof.* The proof is just the application of the integration by parts formula as the density of the process  $X$  is known explicitly. In fact,

$$u(t, x) = \int f(z) \mathfrak{p}_{T-t}(x, z) dz.$$

Here  $\mathfrak{p}_t(x, z)$  denotes the density associated with the process  $X_t(x)$  at the point  $z$  whose formula was given in (20).

From this formula one can verify that the density  $\mathfrak{p}_t(x, z)$  is continuously differentiable in  $x$  (even for  $x = 0$ ) and as the above formula states, one can interchange derivatives with respect to  $x$  for derivatives with respect to  $z$ . Therefore for a function  $f$  which is three times differentiable with polynomial growth and even function,  $u$  is three times differentiable for all  $x \in \mathbb{R} \setminus \{0\}$  due to the successive application of the integration by parts formula. In particular,  $u$  satisfies the PDE (22).  $\square$

Let us denote by  $\bar{\mathfrak{p}}_t(x, y)$  the density transition function of the Euler-Maruyama scheme of step  $T/n$ . The following is a crucial estimate used in the proof of Theorem 5.

**Lemma 8.** Denote  $\bar{\mathfrak{p}}_{\phi(t)}(x) := \bar{\mathfrak{p}}_{\phi(t)}(0, x)$ . Then  $\bar{\mathfrak{p}}_{\phi(t)}(x) = \bar{\mathfrak{p}}_{\phi(t)}(-x)$ . Furthermore for any  $n \in \mathbb{N}$  and any  $t \geq \frac{T}{n}$ , there exists a positive constant independent of  $n$  and  $x$  such that

$$\bar{\mathfrak{p}}_{\phi(t)}(x) \leq \frac{C}{\sqrt{\phi(t)}}.$$

The proof of Theorem 5 is achieved by using a Taylor expansion theorem on the solution of the partial differential equation (22) taking into account the fact that the space variable may be on the positive or negative real axis. Then in order to show the integrability of each term Lemma 8 is used.

## 8.4 Numerical Results

Some simulations have been done in various situations. To save space, we will not give any graphs of actual simulations. Simulations were carried out for the SDE

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

where

$$b(x) = \begin{cases} \theta_1, & x \leq 0, \\ \theta_0, & x > 0. \end{cases}$$

The expectation  $\mathbb{E}[f(X_1)]$  is computed for various definitions of  $\theta_0, \theta_1$  and  $x = 0, 1, \sigma(x) = 1, x, \sin(x)$  and  $f(x) = x^k$ ,  $k = 1, 2, 3$  and  $f(x) = \cos(x)$ . Simulations for various values of  $n$  ranging from 10 to 3.000 were carried out. The number of Monte Carlo simulations is  $n^2$ . Then  $n = 3.000$  was used as the “correct” answer and then regression lines were computed. In almost all cases the confidence interval corresponding to the simulations experiments included the rate  $n^{-1}$ . Therefore these simulation experiments support the impression that the result proved in Section 8.3 is much more general.

More details on the simulation experiments can be found in [24].

## 9 Conclusions

We have given several weak orders with respect to  $|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T^\epsilon)]|$  with discontinuous drift coefficients by analyzing the errors of  $|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]|$  (Propositions 1 and 2 as well as Corollary 1), and  $|\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\bar{X}_T^\epsilon)]|$  (Proposition 3, and Proposition 5) under various assumptions to  $b, b_\epsilon, \sigma$  and  $f$ .

Two types of proof have been given. One based on the proof of weak approximation and another one based on the proof of strong approximation. From the *Conclusions* we see that in the one dimensional case, the rate using weak type proof is  $n^{-1/5}$  and using the strong rate argument is close to  $n^{-1/3}$ . Still, one may see that the weak and strong argument are in nature very different and therefore both proofs may serve different goals. For example, the weak argument can be used to broaden the class of functions  $\mathfrak{F}$ .

Secondly, considering a particular case in Theorem 4, we see that this rate may improve or worsen depending of the size of the set of irregularity.

On the other hand, we finish by giving a simple example where the drift coefficient is discontinuous at a point and the weak rate of convergence is  $n^{-1}$ . This points out that the present article is just a glimpse to a plethora of behavior possible depending on the set of irregularity of the irregular drift  $b$ .

We also studied other cases which for reasons of space are not quoted here which reinforce the previous assertion.

This article should be considered as a first attempt at understanding the approximation quality of the Euler-Maruyama scheme for stochastic differential equations with irregular coefficients. We have (most probably) not obtained the optimal rates in each problem set-up and there remain various questions that should be treated in the near future.

## 10 Appendix

### 10.1 Estimates on stochastic exponentials

In order to give the proof of Proposition 2, we need to give some preliminary results on the Girsanov change of measure process. Let us establish basic results about the moments of exponential martingales.

**Lemma 9.** *Assume that there exists some positive constant  $\hat{\Gamma}$  such that  $|\Gamma_s| \leq \hat{\Gamma}$  for all  $(\omega, s) \in \Omega \times [0, T]$ . (I) For any  $\alpha \in \mathbb{R}$ , and  $t \in [0, T]$*

$$\mathbb{E}[\mathcal{Z}(\Gamma)_t^\alpha] \leq \exp\left(\left|\alpha\left(\alpha - \frac{1}{2}\right)\right|t\hat{\Gamma}^2\right). \quad (23)$$

(II) For any  $\alpha \geq 2$  and  $t \in [0, T]$

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \mathcal{Z}(\Gamma)_s^\alpha \right]^{1/\alpha} \leq C_\alpha \left( 1 + \hat{\Gamma} t^{1/2} \exp \left( \left( \alpha - \frac{1}{2} \right) t \hat{\Gamma}^2 \right) \right)$$

(III) Let  $g \in L^p(Y)$  for some  $p < +\infty$ . For any  $1 \leq q < p$ ,

$$\|g\|_{X, q} \leq \varkappa_1(\hat{\Gamma}, q, p) \|g\|_{Y, p} \text{ with } \varkappa_1(\hat{\Gamma}, q, p) = \exp \left( T \hat{\Gamma}^2 \frac{p+q}{2(p-q)} \right) \xrightarrow{q \uparrow p} +\infty,$$

where  $|\Gamma_s| := |\sigma^{-1}b(s, X_s)| \leq \hat{\Gamma}$ .

*Proof.* (I) Set  $M_t = \int_0^t \Gamma_s dB_s$ . Note that  $\langle M \rangle_t \leq t \hat{\Gamma}^2$ . Using the Cauchy-Schwarz inequality,

$$\mathbb{E}[\mathcal{Z}(\Gamma)_t^\alpha] \leq \mathbb{E} \left[ \exp \left( 2\alpha M_t - \frac{4\alpha^2}{2} \langle M \rangle_t \right) \right]^{1/2} \mathbb{E}[\exp((2\alpha^2 - \alpha)\langle M \rangle_t)]^{1/2}.$$

Since  $\Gamma$  is bounded, then Novikov's condition is satisfied and therefore  $(\exp(2\alpha M_t - 2\alpha^2 \langle M \rangle_t))_{t \in [0, T]}$  is an exponential martingale with mean 1. This leads to (23).

(II) We use the Burkholder-Davis-Gundy (BDG) inequality, Hölder's inequality and (I) to obtain

$$\begin{aligned} \mathbb{E} \left[ \sup_{s \in [0, t]} (\mathcal{Z}(\Gamma)_s - 1)^\alpha \right] &\leq C_\alpha \mathbb{E} \left[ \left( \int_0^t \langle \mathcal{Z}(\Gamma) \rangle_s ds \right)^{\alpha/2} \right] \leq C_\alpha \mathbb{E} \left[ \left( \int_0^t \mathcal{Z}(\Gamma)_s^2 |\Gamma_s|^2 ds \right)^{\alpha/2} \right] \\ &\leq C_\alpha \hat{\Gamma}^\alpha t^{\frac{\alpha}{2}-1} \mathbb{E} \left[ \int_0^t \mathcal{Z}(\Gamma)_s^\alpha ds \right] \leq C_\alpha \hat{\Gamma}^\alpha t^{\frac{\alpha}{2}} \exp \left( \alpha \left( \alpha - \frac{1}{2} \right) t \hat{\Gamma}^2 \right) \end{aligned}$$

(III) For  $\alpha = p/q > 1$  and its conjugate  $\alpha' = \alpha/(\alpha - 1)$ , Hölder's inequality and (23) yields

$$\begin{aligned} \mathbb{E} \left[ \left( \int_0^T |g(s, X_s)|^2 ds \right)^{q/2} \right]^{1/q} &= \mathbb{E} \left[ \mathcal{Z}(\Gamma)_T \left( \int_0^T |g(s, Y_s)|^2 ds \right)^{q/2} \right]^{1/q} \\ &\leq \exp \left( \left( \alpha' - \frac{1}{2} \right) T \hat{\Gamma}^2 \right) \mathbb{E} \left[ \left( \int_0^T |g(s, Y_s)|^2 ds \right)^{p/2} \right]^{1/p}. \end{aligned}$$

□

The proof of the next lemma is similar to the one of Lemma 9(I).

**Lemma 10.** Let  $\theta$  be a random variable with the uniform distribution on  $(0, 1)$ , which is independent of the other variables and processes. Then for any  $\alpha > 1$ ,

$$\mathbb{E} \left[ \exp \left( \theta \int_0^T (\sigma^{-1}b)^*(s, Y_s) dW_s - \frac{\theta}{2} \int_0^T b^* a^{-1} b(s, Y_s) ds \right) \right]^{\alpha-1/\alpha} \leq \exp \left( \alpha T \frac{\|b\|_\infty^2}{\lambda^2} \right).$$

And also the inequality holds for  $b_\epsilon$ .

**Lemma 11.** Let  $\Gamma$  and  $\Gamma'$  be two non-anticipative functions on  $\Omega \times [0, T]$  such that  $|\Gamma|$  and  $|\Gamma'|$  are respectively bounded by the non-negative values  $\hat{\Gamma}$  and  $\hat{\Gamma}'$ . Then for  $\alpha > 1$  there exists a constant  $\varkappa_2 = \varkappa_2(\alpha, p, \hat{\Gamma}, \hat{\Gamma}')$  such that for any  $p > \alpha$ ,

$$\mathbb{E} [ |\mathcal{Z}(\Gamma)_T - \mathcal{Z}(\Gamma')_T|^\alpha ]^{1/\alpha} \leq \varkappa_2(\alpha, p, \hat{\Gamma}, \hat{\Gamma}') \mathbb{E} \left[ \left( \int_0^T |\Gamma_s - \Gamma'_s|^2 ds \right)^{p/2} \right]^{1/p}.$$

*Proof.* As the process  $\Delta_t := \mathcal{Z}(\Gamma)_t - \mathcal{Z}(\Gamma')_t$  satisfies a one dimensional linear SDE which can be solved explicitly

$$\begin{aligned}\Delta_t &= \int_0^t \Delta_s \Gamma_s^* dB_s + \int_0^t \mathcal{Z}(\Gamma')_s (\Gamma_s - \Gamma'_s)^* dB_s \\ &= \mathcal{Z}(\Gamma)_t^{-1} \left\{ \int_0^t \mathcal{Z}(\Gamma)_s \mathcal{Z}(\Gamma')_s (\Gamma_s - \Gamma'_s)^* dB_s + \int_0^t \mathcal{Z}(\Gamma)_s \mathcal{Z}(\Gamma')_s \Gamma_s^* (\Gamma_s - \Gamma'_s) ds \right\}.\end{aligned}$$

Using Minkowski, Hölder's and BDG inequality, maximum values in time of the processes  $\mathcal{Z}$  and Lemma 9(I), (II), for  $p^{-1} + q^{-1} = 1$ ,  $a^{-1} + b^{-1} = 1$  and  $t \in [0, T]$

$$\begin{aligned}\mathbb{E} [|\Delta_t^\alpha|]^{1/\alpha} &\leq C \mathbb{E} \left[ \mathcal{Z}(\Gamma)_t^{-p\alpha} \right]^{1/p\alpha} (1 + \hat{\Gamma}) \mathbb{E} \left[ \left( \int_0^t (\mathcal{Z}(\Gamma)_s \mathcal{Z}(\Gamma')_s (\Gamma_s - \Gamma'_s))^2 ds \right)^{q\alpha/2} \right]^{1/q\alpha} \\ &\leq C(1 + \hat{\Gamma}) \exp \left( \left| p\alpha - \frac{1}{2} \right| \hat{\Gamma}^2 \right) \mathbb{E} \left[ \sup_{s \in [0, T]} (\mathcal{Z}(\Gamma)_s \mathcal{Z}(\Gamma')_s)^{q\alpha} \left( \int_0^t |\Gamma_s - \Gamma'_s|^2 ds \right)^{q\alpha/2} \right]^{1/q\alpha} \\ &\leq C(\hat{\Gamma}, \hat{\Gamma}', p, q, \alpha, a, b) \mathbb{E} \left[ \left( \int_0^t |\Gamma_s - \Gamma'_s|^2 ds \right)^{aq\alpha/2} \right]^{1/aq\alpha}.\end{aligned}$$

Choosing appropriately the values of  $q$  and  $a$  one obtains the result.  $\square$

## 10.2 A lemma on Hölder functions

**Lemma 12.** *The product of two bounded functions in  $H^{1/2,1}(\bar{H})$  remains in  $H^{1/2,1}(\bar{H})$ . Besides, if  $\sigma \in \mathfrak{S}(\lambda, \Lambda)$ , then  $a = \sigma\sigma^*$ ,  $\sigma^{-1}$  and  $a^{-1}$  belongs to  $H^{1/2,1}(\bar{H})$ .*

*Proof.* The first part of the lemma is immediate since for two bounded functions  $f$  and  $g$ ,  $|f(t, x)g(t, x) - f(s, y)g(s, y)| \leq |f(t, x) - f(s, y)| \cdot \|g\|_\infty + |g(t, x) - g(s, y)| \cdot \|f\|_\infty$ .

For the second part, note that  $a^{-1}(t, x)$  is written by the normally convergent series (the Neumann series)

$$a^{-1}(t, x) = \Lambda^{-1} \sum_{k \geq 0} (\text{Id} - \Lambda^{-1}a(t, x))^k.$$

The required continuity follows easily since  $\|\text{Id} - \Lambda^{-1}a(t, x)\|^k \leq (1 - \Lambda^{-1}\lambda)^k$  for any  $(t, x) \in \bar{H}$  and  $k \geq 0$ . The same decomposition hold for  $\sigma^{-1}$ .  $\square$

## 10.3 Some Gaussian upper bound estimates

Our lemma establishes the upper bound for the random variables  $Y_t^\alpha = \alpha Y_t + (1 - \alpha)\bar{Y}_t$  and  $\bar{Y}_t^\alpha = \alpha \bar{Y}_t + (1 - \alpha)Y_{\phi(t)}$ .

**Lemma 13.** *Under the assumptions of Proposition 3, the density of  $Y_t^\alpha$  exists for all  $\alpha \in [0, 1]$  and satisfies the Gaussian upper bound estimate (9) uniformly in  $\alpha$ . Similarly, for the variable  $\bar{Y}_t^\alpha$ , for  $t \geq t_1$  its density  $\bar{p}(t, x, y)$  exists and*

$$\bar{p}(t, x, y) \leq C_1 \mathfrak{g}_{C_2 \phi(t)}(x - y), \quad \forall (t, x, y) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d, \quad (24)$$

for some positive constants  $C_1$  and  $C_2$  independent of  $n$ .

*Proof.* The proof is somewhat standard in Malliavin Calculus theory. One first uses the Lemma 7.1 in [25] to estimate the Malliavin covariance matrix of  $Y_t^\alpha$ , together with classical martingale exponential inequalities, such as Lemma A.5 in [34].

For the second inequality one needs to use the results of [29] for  $\bar{Y}_{\phi(t)}$  and explicitly convolute this upper bound with the explicit transition density of  $\alpha(\bar{Y}_t - \bar{Y}_{\phi(t)})$ . Finally the uniform ellipticity condition together with explicit bounds on the ratio  $\frac{\phi(t)}{\phi(t) + \alpha(t - \phi(t))}$  give the result.  $\square$

## References

- [1] F. Arandiga, A. Cohen, R. Donat, and N. Dyn, *Interpolation and approximation of piecewise smooth functions*, SIAM J. Numer. Anal. **43** (2005), no. 1, 41–57, DOI 10.1137/S0036142903426245.
- [2] S. Arnold, *Approximation schemes for sdes with discontinuous coefficients*, Ph.D. thesis, ETH, Zürich, 2006.
- [3] D. G. Aronson, *Non-negative solutions of linear parabolic equations*, Ann. Scuola Norm. Sup. Pisa (3) **22** (1968), 607–694.
- [4] S. Attanasio, *Stochastic flows of diffeomorphisms for one-dimensional SDE with discontinuous drift*, Electron. Commun. Probab. **15** (2010), 213–226, DOI 10.1214/ECP.v15-1545.
- [5] R. Avikainen, *On irregular functionals of SDEs and the Euler scheme*, Finance Stoch. **13** (2009), no. 1, 381–401.
- [6] V. Bally and D. Talay, *The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function*, Probab. Theory Related Fields **104** (1996), no. 1, 43–60, DOI 10.1007/BF01303802.
- [7] ———, *The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density*, Monte Carlo Methods Appl. **2** (1996), no. 2, 93–128, DOI 10.1515/mcma.1996.2.2.93.
- [8] R.F. Bass, *Diffusions and Elliptic Operators*, Springer, 1998.
- [9] F. Bernardin, M. Bossy, M. Martinez, and D. Talay, *On mean numbers of passage times in small balls of discretized Itô processes*, Electron. Commun. Probab. **14** (2009), 302–316, DOI 10.1214/ECP.v14-1479.
- [10] K. S. Chan and O. Stramer, *Weak consistency of the Euler method for numerically solving stochastic differential equations with discontinuous coefficients*, Stochastic Process. Appl. **76** (1998), no. 1, 33–44.
- [11] L. Devroye and L. Györfi, *Nonparametric Density Estimation. The  $L_1$  view*, John Wiley and Sons, 1985.
- [12] P. Étoré and M. Martinez, *Exact simulation for SDE with discontinuous drift* (2013), available at arXiv: 1301.3019v1.
- [13] M. Frazier and B. Jawerth, *Decomposition of Besov spaces*, Indiana Univ. Math. Journ. **34** (1985), 777–799.
- [14] M. Frazier, B. Jawerth, and G. Weiss, *Littlewood-Paley theory and the study of function spaces*, CBMS-AMS Regional Conf. Ser. 79, Providence, Amer. Math., 1991.
- [15] C. Geiss, S. Geiss, and E. Gobet, *Generalized fractional smoothness and  $L_p$ -variation of BsdEs with non-Lipschitz terminal condition*, Stochastic Process. Appl. **122** (2012), 2078–2116.
- [16] S. Geiss and E. Gobet, *Fractional smoothness and applications in finance*, Advanced mathematical methods for finance, Springer, Heidelberg, 2011, pp. 313–331.
- [17] S. Geiss and A. Toivola, *Weak convergence of error processes in discretizations of stochastic integrals and Besov spaces*, Bernoulli **15** (2009), no. 4, 925–954, DOI 10.3150/09-BEJ197.
- [18] E. Gobet and R. Munos, *Sensitivity analysis using Itô-Malliavin calculus and Martingales, and application to stochastic optimal control*, SIAM J. Control Optim. **43** (2005), no. 5, 1676–1713, DOI 10.1137/S0363012902419059.
- [19] N. Halidias and P.E. Kloeden, *A note on strong solution of stochastic differential equations with discontinuous a drift coefficient*, J. Appl. Math. Stoch. Anal. (2006), Art. ID 73257, 6.
- [20] ———, *A note on the Euler-Maruyama scheme for stochastic differential equations with a discontinuous monotone drift coefficient*, BIT **48** (2008), no. 1, 51–59.
- [21] R. Janssen, *Difference methods for stochastic differential equations with discontinuous coefficients*, Stochastics **13** (1984), no. 3, 199–212.
- [22] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus (2nd)*, Springer, 1998.
- [23] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Applications of Mathematics (New York), vol. 23, Springer-Verlag, Berlin, 1992.
- [24] A. Kohatsu-Higa, A. Lejay, and K. Yasuda, *On Weak Approximation of Stochastic Differential Equations with Discontinuous Drift Coefficient*, RIMS Kôkyûroku **1788** (2012), 94–106.
- [25] A. Kohatsu-Higa, *Weak Approximations: A Malliavin Calculus approach*, Mathematics of Computation **70** (2001), 135–172.
- [26] N. Krylov, *An inequality in the theory of stochastic integrals*, Th. Probab. Applic. **16** (1971), 438–448.
- [27] G.C. Kyriazis, *Approximations from Shift-Invariant Spaces*, Constr. Approx. **11** (1995), 141–164.
- [28] O.A. Ladyženskaja, V.A. Solonnikov, and N.N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type*, American Mathematical Society, 1967.
- [29] V. Lemaire and S. Menozzi, *On some non asymptotic bounds for the Euler scheme*, Electron. J. Probab. **15** (2010), no. 53, 1645–1681.
- [30] M. Martinez and D. Talay, *Time discretization of one dimensional Markov processes with generators under divergence form with discontinuous coefficients*, Electronic Journal Probab. **17** (2012), no. 27, DOI 10.1214/EJP.v17-1905.
- [31] R. Mikulevičius and C. Zhang, *Weak Euler Approximation for Itô Diffusion and Jump Processes* **33** (April 2015), no. 3, 549–571, DOI 10.1080/07362994.2015.1014102.
- [32] R. Mikulevičius and E. Platen, *Rate of convergence of the Euler approximation for diffusion processes*, Math. Nachr. **151** (1991), 233–239, DOI 10.1002/mana.19911510114.



- [33] H.-L. Ngo and D. Taguchi, *Approximation for non-smooth functionals of stochastic differential equations with irregular drift* (May 2015), 1–20 pp., available at [arxiv:1505.03600](https://arxiv.org/abs/1505.03600).
- [34] D. Nualart, *The Malliavin calculus and related topics*, 2nd ed., Probability and its Applications (New York), Springer-Verlag, 2006.
- [35] Paweł Przybyłowicz, *Optimality of Euler-type algorithms for approximation of stochastic differential equations with discontinuous coefficients*, *International Journal of Computer Mathematics* **91** (2014), no. 7, 1461–1479, DOI 10.1080/00207160.2013.844336.
- [36] A. Semrau, *Discrete approximations of strong solutions of reflecting SDE's with discontinuous coefficients*, *Bull. Pol. Acad. Sci. Math.* **57** (2009), no. 2, 169–180, DOI 10.4064/ba57-2-10.
- [37] A. I. Stepanets, *Methods of Approximation Theory*, VSP, 2005.
- [38] G. Strang and G. Fix, *A Fourier analysis of the finite-element variational method*, *Constructive Aspects of Functional Analysis* (G. Geymonant, ed.), C.I.M.E., 1971, pp. 793–840.
- [39] D. W. Stroock, *Diffusion semigroups corresponding to uniformly elliptic divergence form operators*, *Séminaire de Probabilités, XXII*, *Lecture Notes in Math.*, vol. 1321, Springer, Berlin, 1988, pp. 316–347.
- [40] D.W. Stroock and S.R.S. Varadhan, *Multidimensional Diffusion Processes*, Springer, 1979.
- [41] H. Triebel, *Theory of function spaces II*, Birkhauser, Basel, 1992.
- [42] A. J. Veretennikov, *On strong solutions and explicit formulas for solutions of stochastic integral equations*, *Math. USSR Sb.* **39** (1981), no. 3, 387–403, DOI 10.1070/SM1981v039n03ABEH001522.
- [43] L. Yan, *The Euler scheme with irregular coefficients*, *Ann. Probab.* **30** (2002), no. 3, 1172–1194.