Weak rate of convergence of the Euler-Maruyama scheme 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 rate of convergence of the Euler-Maruyama scheme for stochastic differential equations with non-regular drift. Journal of Computational and Applied Mathematics, Elsevier, 2017, 326C, pp.138-158. <10.1016/j.cam.2017.05.015>. <hal-00840211v5>

HAL Id: hal-00840211
https://hal.inria.fr/hal-00840211v5

Submitted on 6 Jul 2017

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† Kazuhiro Yasuda‡

Abstract

We consider an Euler-Maruyama type approximation method for a stochastic differential equation (SDE) with a non-regular 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 one is the error of regularization of the drift coefficient within a given class of parametrized functions. The second one is the error of the regularized Euler-Maruyama scheme. After an optimization procedure with respect to the parameters 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

$$X_t = x + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s + \int_0^t b(s, X_s)\,\mathrm{d}s, \tag{1}$$

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

*Department of Mathematical Sciences — Ritsumeikan University and Japan Science and Technology Agency— 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577, Japan.
†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.
‡Faculty of Science and Engineering — Hosei University — 3-7-2, Kajino-cho, Koganei-shi, Tokyo, 184-8584, Japan.

$\xi_1, \ldots, \xi_n$ with zero mean and identity variance matrix. The Euler-Maruyama scheme consists in computing iteratively for $k = 0, \ldots, n-1$ with $t_k = \frac{kT}{n}$,

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

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

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

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

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

Roughly speaking, when σ is uniformly elliptic and σ, b are at least of class $\mathcal{C}^4$ in space, $d_f(X, \overline{X})$ 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 [5, 6]. This is a non-trivial consequence of the regularity of the density of X_T. Weakening the conditions on the coefficients b or σ naturally implies 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 [37] and R. Mikulevičius and C. Zhang [36]. More precisely, if[1] $a = \sigma\sigma^*, b \in \mathrm{H}^{\alpha/2,\alpha}(\overline{H})$ and $f \in \mathrm{H}^{2+\alpha}(\mathbb{R}^d)$ for $\alpha \in (0, 1) \cup (1, 2) \cup (2, 3)$ then $d_f(X, \overline{X})$ 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 α, that is $\mathcal{C}^k$ 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 σ, 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, 3, 9, 20–22, 28, 38, 39, 43, 49] and time [42]. Other related studies can be found in [8, 10, 12, 35]. Still, optimal weak rates of convergence remain an elusive subject.

Our method of analysis is to give first a perturbation result in order to compare $\mathbb{E}[f(X_T)] - \mathbb{E}[f(X^\epsilon_T)]$ where X^ϵ has a smooth drift coefficient b_ϵ which approximates b. Then we provide some bounds on the weak rate of convergence for the Euler-Maruyama scheme for X^ϵ 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.

[1] For exact definitions of the functional spaces, we refer to Section 2.

Putting these two results together we finally get various weak rates of convergence according to the choices of $\mathfrak{M}$ and $\mathfrak{F}$. We get then results which are different from the ones in [37] 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 in 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 $\overline{X}$ and the Euler-Maruyama scheme $\overline{X}^{\epsilon}$ measured by $d_f(X, \overline{X})$. 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 [25].

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 analysis techniques to tackle these problems.

Outline. Notations, hypotheses and our main results are given in Sections 2, 3 and 4. The perturbation formulae are proved in Section 5. Sections 6 provides us with estimates on the distance $d_p(b, b_\epsilon)$, which will be defined in Section 3. Section 7 provides a weak rate of convergence when the drift coefficient is smooth. Section 8 is devoted to extending our results when the drift is not approximated. 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 $|m| = \max_{|x|=1, x\in\mathbb{R}^k} |mx|$. We define similarly the norm of a multi-linear transformation. We do not make any particular difference in notation for norms of real numbers, vectors or multi-linear transforms.

Spaces of functions. For $r > 0$, we denote by $\mathcal{C}^r_{\mathrm{b}}(U)$ 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 $f \in \mathcal{C}^r_{\mathrm{b}}(U)$ for any multi-index $\alpha = (\alpha_1, \dots, \alpha_j) \in \{1, \dots, d\}^j$ of length $j \leq \lfloor r \rfloor$ as either $\partial^j_{x_{\alpha_1},\dots,x_{\alpha_j}} \equiv \frac{\partial^j}{\partial x_{\alpha_1} \cdots \partial x_{\alpha_j}}$. The multi-linear transform is denoted by $\partial^j_x \equiv \{\partial^j_{x_{\alpha_1},\dots,x_{\alpha_j}}; \alpha \in \{1, \dots, d\}^j\}$. For $f = (f^1, \dots, f^d) \in (\mathcal{C}^k_{\mathrm{b}}(U))^d$, $k \in \mathbb{N}$, we define the following semi-norm

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

For $k \in \mathbb{N}$, the norm on $\mathcal{C}^k_{\mathrm{b}}(U)$ is $\|f\|_{b,k} = \sum_{j=0}^k \|\partial^j_x f\|_\infty$. For $r \notin \mathbb{N}$, the norm on $\mathcal{C}^r_{\mathrm{b}}(U)$ is

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

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 exist 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 $\mathcal{C}_{\mathrm{p}}(\mathbb{R}^d)$.

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 $\mathcal{C}_{\mathrm{e}}(\mathbb{R}^d)$.

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

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

We denote by $\mathcal{C}^k_a(\mathbb{R}^d)$ with $a =$ e, p or Sl 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 $\mathcal{C}_a(\mathbb{R}^d)$.

Functions of space and time. We fix a time horizon $T > 0$. Two particular domains will appear frequently. Let $H = [0,T) \times \mathbb{R}^d$ and $\overline{H} = [0,T] \times \mathbb{R}^d$. The set $\mathcal{C}_a(\overline{H})$ for $a =$ e, p, Sl denotes the set of functions $u : H \to \mathbb{R}$ which are continuous on $\overline{H}$ and such that $u(s,\cdot) \in \mathcal{C}_a(\mathbb{R}^d)$ uniformly for $s \in [0,T)$. For example, in the case that $a =$ Sl, $\lim_{|x|\to\infty} \sup_{s\in[0,T)} |u(s,x)| e^{-c|x|^2} = 0$ for every $c > 0$. The extension of the norms from the spaces $\mathcal{C}^k_b(\mathbb{R}^d)$ to the spaces $\mathcal{C}^k_b(\overline{H})$ 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 $\mathrm{H}^{\alpha/2,\alpha}(\overline{H})$ the space of continuous functions which are α-Hölder continuous in space uniformly in time and $\alpha/2$-Hölder continuous in time uniformly in space. Similarly, $\mathrm{H}^\alpha(\mathbb{R}^d)$ is the space of α-Hölder continuous functions on $\mathbb{R}^d$ with their associated natural norm.

We denote by $\mathrm{W}^{1,2}_p(H)$ the space of functions in $\mathrm{L}^p(H)$ whose weak derivatives (first in time, first and second in space) are also in $\mathrm{L}^p(H)$. The space $\mathrm{W}^{1,2}_{p,\mathrm{loc}}(H)$ is the space of functions in $\mathrm{W}^{1,2}_p(K)$ for any compact $K \subseteq H$. Embedding theorems for these spaces may be found in [32, II.§ 3].

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

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

Constants will be denoted by the same symbol C or K. Unless stated explicitly they may change values from one line to the next. They may depend on the time constant T but this will not be directly indicated as we suppose from the start that T is fixed. In many cases, we make the effort to explicitly state the dependence of the constants on the parameters of the problem.

When using Hölder's inequality 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 an *almost conjugate pair* if $p > 1$, $q > 1$ and $p^{-1} + q^{-1} < 1$

Due to the use of these parameters, some constants will depend on the choice of (p, q). In some cases, $(p, q) = (1, \infty)$. If that case is allowed it means that the constant is finite in that case too. On the other hand, if for example, the constant C depends on (p, q) a conjugate pair with $p \in (1, \infty)$ (or equivalently $p > 1$) this means that for $p = 1$ or $p = \infty$, the constant may not be finite.

In general, we not write the domains over which space integrals are evaluated unless a confusion may arise.

3 The approximation framework to obtain the weak rate of convergence

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

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

$$\exists \Lambda \geq \lambda > 0 \text{ such that } \lambda|\xi|^2 \leq \xi^* a(t, x)\xi \leq \Lambda|\xi|^2,\ \forall (t, x) \in \overline{H},\ \forall \xi \in \mathbb{R}^d, \tag{H1}$$

$$\sigma \text{ is uniformly continuous on } \overline{H}. \tag{H2}$$

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

Sometimes, we need stronger regularity condition on σ. We define $\mathfrak{H}(\lambda, \Lambda)$ as the space of functions σ in $\mathfrak{S}(\lambda, \Lambda)$ for which

$$\sigma \in \mathrm{H}^{1/2,1}(\overline{H}). \tag{H2'}$$

Hypothesis (H2') means that σ 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 $b : \overline{H} \to \mathbb{R}^d$ such that

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

When $\sigma \in \mathfrak{S}(\lambda, \Lambda)$ and $b \in \mathfrak{B}(\Lambda)$, there exists a unique weak solution in some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ for the SDE (1) (see e.g. [46]). Here B is a d-dimensional Wiener process defined on the probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

Remark 2. As shown in [48], 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 $\overline{X}$ the continuous Euler-Maruyama scheme of X with a time step T/n. For this, we define for fixed $n \in \mathbb{N}$, $\phi(s) \equiv \phi_n(s) = \sup\{t \leq s \,|\, t = kT/n \text{ for } k \in \mathbb{N}\}$, then

$$\overline{X}_t = x + \int_0^t \sigma(\phi(s), \overline{X}_{\phi(s)})\,\mathrm{d}B_s + \int_0^t b(\phi(s), \overline{X}_{\phi(s)})\,\mathrm{d}s. \tag{3}$$

The vector $(\overline{X}_0, \overline{X}_{t_1}, \dots, \overline{X}_T)$ defined by the above equation at time $(t_0, t_1, \dots, t_n)$ has the same distribution as the vector $(\overline{X}_{t_0}, \overline{X}_{t_1}, \dots, \overline{X}_{t_n})$ defined by the recursive relation (2). Yet $\overline{X}$ 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 $\mathfrak{B}(\Lambda)$ for drift functions, we define a distance d_p that depends on a parameter $p > 0$ (for example, p is the exponent of the functional space $\mathrm{L}^p(\overline{H})$).

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 $\mathcal{C}^k$. Approximations b_ϵ 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 (1), 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 $\lim_{\epsilon\to 0} d_p(b_\epsilon, b) = 0$. Using b_ϵ, 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

$$X_t^\epsilon = x + \int_0^t \sigma(s, X_s^\epsilon)\,\mathrm{d}B_s + \int_0^t b_\epsilon(s, X_s^\epsilon)\,\mathrm{d}s. \tag{4}$$

The approximation quality of X^ϵ 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_p 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

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

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

The couple of parameters (p, q) measures the strength of the convergence of b_ϵ 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 $d_f(X, \overline{X})$, we use the triangle inequality

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

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

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

$$d_f(X^{\epsilon}, \overline{X}^{\epsilon}) := |\mathbb{E}[f(X_T^{\epsilon})] - \mathbb{E}[f(\overline{X}_T^{\epsilon})]| \leq \frac{C_{\epsilon}}{n^{\delta}}. \tag{6}$$

This is in general the case when one choose a regularization b_{ϵ} of b by using mollifiers. Clearly the parameters β and δ 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 $\mathrm{O}(\epsilon^{\gamma})$ then the distance between the drift coefficient and its approximation $b_{\epsilon} \in \mathfrak{M}$ is measured by the parameter γ. 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 δ in (6). 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 $\epsilon \to 0$ as $b \notin \mathfrak{M}$. The rate of explosion is characterized by the parameter β.

Therefore the parameters δ, β and γ 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 δ, β and γ 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 in Figure 1.

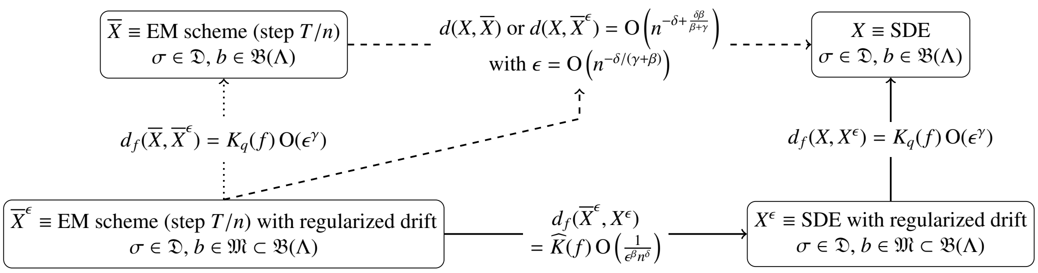

Figure 1: Computation of the weak rate of convergence for $f \in \mathfrak{F}$.

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

Theorem 1. *Let us consider suitable classes of 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', γ, β, δ 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* $d_p(b, b_\epsilon) = \mathrm{O}(\epsilon^\gamma)$ *and*

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

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

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

In the above result one clearly sees that, given (5), γ measures the irregularity of b with respect to the class $\mathfrak{M}$ and the law of X. The parameter δ measures the weak error within the class $\mathfrak{M}$, for $f \in \mathfrak{F}$. The parameter β 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 $\widehat{K}(f)$ 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.* [32, Chapter IV §5]), on which the results in [37] are based on. Alternatively, the solution are sought in Sobolev spaces, which allows to consider weak conditions on the regularity of the drift [32, Chapter IV §7]. The usual probabilistic representation for such solution also holds in this situation.

Theorem 2 (Theorem 3' and Corollary p. 401, [48])**.** *Let* L *be as above with* $\sigma \in \mathfrak{S}(\lambda, \Lambda)$ *and* $b \in \mathfrak{B}(\Lambda)$*. Fix* $T > 0$ *and assume that* $f \in \mathfrak{C}_{\mathrm{Sl}}(\mathbb{R}^d)$*. The Cauchy problem*

$$\begin{cases} \dfrac{\partial u(t,x)}{\partial t} + Lu(t,x) = 0, \\ u(T,x) = f(x) \end{cases} \tag{7}$$

has a solution u *in* $\cap_{p>1}\mathrm{W}^{1,2}_{p,\mathrm{loc}}(H) \cap \mathfrak{C}_{\mathrm{Sl}}(\overline{H})$*. This solution, satisfying* $u(s,x) = \mathbb{E}[f(X_T)|X_s = x]$*, is unique in* $\mathrm{W}^{1,2}_{p,\mathrm{loc}}(H) \cap \mathfrak{C}_{\mathrm{Sl}}(\overline{H})$ *for* $p \ge d+1$*. Besides,*

$$f(X_T) = u(s, X_s) + \int_s^T \nabla u(r, X_r)^* \sigma(r, X_r)\,\mathrm{d}B_r$$

and

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

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 (7) given by Theorem 2.

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

$$f(X_T) = v(0,x) + \int_0^T \nabla v(s, X_s)^* \sigma(s, X_s)\, \mathrm{d}B_s + \int_0^T b^*(s, X_s) \nabla v(s, X_s)\, \mathrm{d}s.$$

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

$$Y_t = x + \int_0^t \sigma(s, Y_s)\, \mathrm{d}B_s. \tag{8}$$

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 $\Gamma : \Omega \times [0,T] \to \mathbb{R}^d$, be a bounded, predictable stochastic process. To Γ, we relate the exponential martingale

$$\mathcal{Z}(\Gamma)_t := \exp\left(\int_0^t \Gamma_s^*\, \mathrm{d}B_s - \frac{1}{2} \int_0^t \Gamma_s^* \Gamma_s\, \mathrm{d}s \right). \tag{9}$$

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

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

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 $\frac{\mathrm{d}\mathbb{Q}^\Gamma}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t} = \mathcal{Z}(\Gamma)_t;\ t \geq 0$. Various moment properties of $\mathcal{Z}(\Gamma)$ are given in Appendix 10.1.

We let $\mathrm{d}W_t^\Gamma = \mathrm{d}B_t - \Gamma_t\, \mathrm{d}t$ denote the Wiener process on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q}^\Gamma)$. As it is well known, choosing the right definition for Γ 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)

$$Y_t^\Gamma = x + \int_0^t \sigma(s, Y_s^\Gamma)\, \mathrm{d}W_s^\Gamma.$$

As $Y^\Gamma \overset{\mathcal{L}}{=} Y$ where the previous equality means that the law of both processes are the same and therefore we do not make any distinction between them as we only consider 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 $g : \overline{H} \to \mathbb{R}^k$, we say that $g \in \mathrm{L}^{p,q}(X)$ if its associated norm

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

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 $\|g\|_{X,p} := \|g\|_{X,2,p}$, $\mathrm{L}^p(X) := \mathrm{L}^{2,p}(X)$, $\|g\|_{X,\infty} := \|g\|_\infty$. 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 μ, where Y is solution to (8). Then there exists a constant $C(p, p')$ that depends only on p, p' T, Λ, λ such that*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq C(p, p')\|b - b_\epsilon\|_{Y,p}\mathbb{E}\left[|f(Y_T) - \mu|^{p'}\right]^{1/p'}. \tag{10}$$

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 $\mathbb{E}[\mathcal{Z}(\Gamma)_T] = \mathbb{E}[\mathcal{Z}(\Gamma^\epsilon)_T] = 1$,

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

for a conjugate pair (p, p''). Hence, (10) follows from Lemma 11. □

Corollary 1. *When $\sigma \in \mathfrak{S}(\lambda, \Lambda)$ and $b, b_\epsilon \in \mathfrak{B}(\Lambda)$,*
i) *If $f \in \mathcal{C}_{\mathrm{Sl}}(\mathbb{R}^d)$, then for $p > 2$*

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

ii) *If $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{R}^d)$, 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 $f \in \mathcal{C}_{\mathrm{Sl}}(\mathbb{R}^d)$ is uniformly Lipschitz continuous with constant* $\mathrm{Lip}(f)$*, for any $p > 1$,*

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

Proof. The proof of this corollary is immediate by choosing respectively $\mu = \mathbb{E}[f(Y_T)]$ and $p' = 2$ for i), $\mu = 0$ for ii) and $\mu = f(x)$ together with the control $\mathbb{E}[|Y_T - x|^{p'}]^{1/p'} \leq \Lambda T^{1/2}$ for any $T > 0$ for iii). □

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 [16–18] allows one to consider several classes $\mathfrak{F}$ of functions for which $K_2(f) := \sqrt{\operatorname{Var} f(Y_T)}$ 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 $\mathbb{E}\left[|f(Y_T) - \mu|^2\right]^{1/2}$ as the process Y is independent of b or b_ϵ. 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 (7), $\|\nabla u\|_{X,p'}$ 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 $f \in \mathcal{C}_{\mathrm{SI}}(\mathbb{R}^d)$ is a function such that for some $1 < p' \le \infty$, $\|\nabla u\|_{X,p'} < +\infty$, where u is the solution to (7) with terminal condition f. Let (p, p') be an almost conjugate pair. Then there exists a constant $C_2(p, p')$ depending only on T, Λ, λ and (p, p') such that*

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

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

Proof. Set $\Gamma_s = \sigma^{-1}(s, X_s)(b_\epsilon(s, X_s) - b(s, X_s))$ and $\hat{\Gamma} := \sup_{(\omega,s)\in\Omega\times[0,T]} |\Gamma_s| \le 2\Lambda\lambda^{-1}$. From Girsanov's theorem and the quadratic covariation between martingales,

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

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 $\alpha_1^{-1} + \alpha_2^{-1} + (p')^{-1} = 1$,

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

The result follows from Lemma 9(I) and Lemma 9(III) since $\|b_\epsilon - b\|_{X,\alpha_2} \le \varkappa_1(\hat{\Gamma}, \alpha_2, p)\|b_\epsilon - b\|_{Y,p}$ with $p > \alpha_2$. □

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 $d_f(X, X^\epsilon)$. We have found in Propositions 1 and 2 that this distance $d_f(X, X^\epsilon)$ depends on two factors. The first one is the distance between b

and b_ϵ in the $\mathrm{L}^p(H)$-norm. The second one is a measure on the regularity of f.

The distance $\|b - b_\epsilon\|_{Y,p}$ will appear recurrently in the discussions to follow. Various study 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_ϵ decreases 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.

We say that a process Y satisfies an *upper Gaussian estimate* if Y_t has a density $\mathsf{p}(t, x, y)$ for each $t \in [0, T]$ so that it satisfies the inequality

$$\mathsf{p}(t, x, y) \leq C_1 \mathsf{g}_{C_2 t}(x - y), \ \forall (t, x, y) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d, \tag{11}$$

for some positive constants C_1 and C_2.

The estimate (11) holds for example if the diffusion coefficient a satisfies (H1) and it belongs to the Hölder space $\mathrm{H}^{\alpha/2,\alpha}(H)$ for some $\alpha > 0$ (see e.g. [32, § IV.13, p. 377]). An alternative condition which involves the weak differentiability of a (but not necessarily its Hölder continuity) may be found in [30]. When a is only uniformly continuous, the upper Gaussian estimate does not necessarily hold, as the density itself may not exist (See [13] for a counter-example).

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 $\overline{Y}_t$ has a density $\overline{\mathsf{p}}$ which satisfies a Gaussian upper bound [33] where the constants C_1 and C_2 are independent of n. In general, in both cases above the constants C_1 and C_2 will depend on the constants λ and Λ appearing in condition (H1).

Lemma 1. *Assume that Y satisfies a Gaussian upper bound estimate* (11)*. Then for any $p \in [1, \infty)$, $q, r \in (p \vee i, \infty]$ with $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p \vee i}$, $i = 1, 2$, and every function $g : H \to \mathbb{R}^d$ with $\|g\|_{\mathrm{L}^{r,q}(H)} < \infty$, there exists a positive constant $C_3 \equiv C_3(p, q, r, i, T)$ such that for $i = 1, 2$,*

$$\|g\|_{Y,i,p} \leq C_3 \left(\int_0^T \left(\int_{\mathbb{R}^d} |g(s, y)|^q \, \mathrm{d}y \right)^{r/q} \mathrm{d}s \right)^{1/r} =: C_3 \|g\|_{\mathrm{L}^{r,q}(H)}. \tag{12}$$

Proof. Consider the case for $i = 1$, i.e., $\|g\|_{Y,1,p}$. 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 using (11), there exists a positive constant $K(\alpha) := C_1^{1/\alpha} (2\pi)^{-d/(2\alpha)}$ such that for any $\alpha > 1$,

$$\left(\int_{\mathbb{R}^d} \mathsf{p}(t, x, y)^\alpha \, \mathrm{d}y \right)^{1/\alpha} \leq \left(\int_{\mathbb{R}^d} \frac{C_1^\alpha}{(2\pi C_2 t)^{\alpha d/2}} \exp\left(-\frac{\alpha |y - x|^2}{C_2 t} \right) \mathrm{d}y \right)^{1/\alpha} \leq \frac{K(\alpha)}{\alpha^{d/(2\alpha)} t^{(\alpha-1)d/(2\alpha)}}. \tag{13}$$

Second, for $p \in [1, q)$, $\frac{p}{q} + \frac{1}{\alpha} = 1$ and a conjugate pair (a, a'), we have from (13) that

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

Letting $pa = r$ and noting that the last time integral is finite under the restriction $(1 - \frac{1}{\alpha})\frac{d}{2}a' < 1$, (12) follows for $i = 1$. In fact, the condition $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p \vee 1}$ is equivalent to $(1 - \frac{1}{\alpha})\frac{d}{2}a' < 1$. The cases $q = \infty$ and/or $r = \infty$ are proven similarly.

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

Remark 3. Even in the absence of Gaussian upper bound estimates, the Krylov estimate ([29] or [7, Theorem 5.6.2, p. 114]) could also be used with Hypothesis (H1) in order to get an estimate on $\|g\|_{Y,p}$.

The proof of the following lemma is a variation of Lemma 1 for discretized processes. For this, recall that the definition of the function ϕ has been given before equation (3).

Lemma 2. *Assume that $Y \equiv Y^n$ satisfies the Gaussian upper bound estimate* (11) *uniformly in $n \in \mathbb{N}$. That is, the constants C_1, C_2 appearing in* (11) *do not depend on n . Then for any $p \geq 1$, any $q, r \in (p \vee i, \infty]$ with $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p \vee i}$, $i = 1, 2$, and $g : H \to \mathbb{R}^d$ with $\|g\|_{\mathrm{L}^{r,q}(H)} < \infty$, there exist some positive constants $C_1' = C(p, r, q)$ and $C_2' = C(p, i)$ such that*

$$\left(\mathbb{E}\left[\left(\int_0^T |g(\phi(s), Y_{\phi(s)})|^i \, \mathrm{d}s\right)^{p/i}\right]\right)^{1/p} \leq C_1' \|g\|_{\mathrm{L}^{r,q}(H)} + \frac{C_2'}{n^{1/i}} |g(0, x)|.$$

Define the norm

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

Proof. As in the proof of Lemma 1 we prove the above statement 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 using Jensen inequality,

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

For a conjugate pair (α, α') with $\alpha = \frac{q}{q-p}$, there exists $K(\alpha) > 0$ such that (13) is satisfied. Then there exists some constant $C(\alpha) := K(2\pi)^{-\frac{d}{2\alpha'}}$ such that

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

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

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

The result follows choosing $a = r/p$. As before, the condition $a'd/2\alpha' < 1$ is equivalent to $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p}$. □

Remark 4. The above inequalities will be used for $g = b - b_\epsilon$ and $b = \partial^k b_\epsilon$. Instead of the processes Y and $\overline{Y}$ — the continuous Euler-Maruyama scheme associated to Y —, other weighted combinations such as $\alpha Y + (1-\alpha)\overline{Y}$ or $\alpha\overline{Y} + (1-\alpha)\overline{Y}_{\phi(\cdot)}$ may also appear. These processes satisfy also a 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 $\mathrm{L}^{dp}(H)$. 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 $\epsilon > 0$,

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

Note that b_ϵ is Lipschitz continuous and is an approximation of b. Furthermore, we also have the following rates of convergence of b_ϵ to b: for $p \geq 1$,

$$\left(\int_{-\infty}^{\infty}|b_\epsilon(x) - b(x)|^p \,\mathrm{d}x\right)^{\frac{1}{p}} = \left(\frac{4\epsilon}{p+1}\right)^{\frac{1}{p}} = O\left(\epsilon^{\frac{1}{p}}\right). \tag{14}$$

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

Further regularity of the approximation b_ϵ can be obtained, if we use a mollifier with the Gaussian kernel, that is $b_\epsilon(x) := \int_{-\infty}^{+\infty} b(u)\mathrm{g}_\epsilon(x-u)\,\mathrm{d}u$, then we have the same order of the convergence as the above (14) with $b_\epsilon \in \mathcal{C}_b^\infty(\mathbb{R})$ and $\|b_\epsilon\|_{\mathrm{H}^\alpha(\mathbb{R})} = \mathrm{O}(\epsilon^{-\alpha/2})$, $\|\partial_x^k b_\epsilon\|_{\mathrm{L}^p(\mathbb{R})} = \mathrm{O}(\epsilon^{-\frac{1}{2}(k-\frac{1}{p})})$, for any $k \in \mathbb{N}$, $\alpha \in (0,1)$ and $p \geq 1$ (these estimates follow by using Hölder inequalities and standard renormalization arguments for the Gaussian kernel function).

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, [45], uses the kernel estimators based on shift invariant spaces to approximate functions in $\mathrm{W}_p^k(\mathbb{R}^d)$. They provide uniform bounds on the optimal $\mathrm{L}^p(\mathbb{R}^d)$ distance depending on the kernel regularization parameter ϵ and the $\mathrm{W}_p^k(\mathbb{R}^d)$ 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 [31].

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 [44].

A different approach, which may seem unnatural at the beginning, relies on wavelet theory. In wavelet theory, one actually approximates regular functions using discontinuous functions. This idea 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)], [14, 15] or [47] 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_ϵ in the rest of the article. For this, we define, $b_\epsilon^1 := b_\epsilon^* a^{-1} b_\epsilon$, $\mathbf{b}_\epsilon = (b_\epsilon, b_\epsilon^1)$ and similarly for $\mathbf{b} = (b, b^1)$.

Hypothesis. *When $d = 1$, the triplet $(\mathbf{b}_\epsilon, \mathbf{b}, Y)$ is time-homogeneous and satisfies: For any*

$q,\ r \geq 1,$

$$\|b - b_\epsilon\|_{\mathrm{L}^{r,q}(H)} \leq C\epsilon^{\frac{1}{q}},\ \|\mathbf{b}_\epsilon\|_{\mathrm{H}^k(\mathbb{R})} \leq C\epsilon^{-\frac{k}{2}},\ \forall k \in \mathbb{N}_0$$
$$\text{and } \|\partial_x^k \mathbf{b}_\epsilon\|_{\mathrm{L}^{r,q}(H)} \leq C\epsilon^{-(\frac{k}{2}-\frac{1}{2q})},\ \forall k \in \mathbb{N}. \tag{H4}$$

The above positive constant C is independent of ϵ and k.

The above simplifying assumption will ease 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. The above assumption corresponds to an idealized asymptotic case taken from Example 1.

Still, we remark that many other cases besides the one described above can be entertained. Besides, there is no reason to restrict to the case $d = 1$.

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

We now exhibit some situations where the assumptions of Theorem 1 hold. We therefore assume Hypothesis (H4). We obtain estimates of weak rates of convergence for different spaces $\mathfrak{F}$ and $\mathfrak{M}$. In particular, we will always obtain that the subclasses of functions $\mathfrak{F}$, $\mathfrak{M}$ and $\mathfrak{B}$ are such that for $f \in \mathfrak{F}$ and $b \in \mathfrak{B}(\Lambda)$ there exists a sequence $b_\epsilon \in \mathfrak{M}$ with $K_{p'}(f) < \infty$ and $d_p(b, b_\epsilon) \leq C\epsilon^\gamma$ for some $\gamma > 0$.

We start considering an auxiliary result which shows the boundedness of the second term appearing in Proposition 1.

Lemma 3. *Assume that $\sigma \in \mathfrak{H}(\lambda, \Lambda)$.*
Let $f \in \mathcal{C}_{\mathrm{Sl}}(\mathbb{R}^d)$. Then for any integer k, there exists a function $B_{\mathrm{Sl}}(f, x)$ such that

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

Proof. These inequalities follow immediately from (11) applied to the densities $\overline{\mathsf{p}}(t, x, y)$ and to $\mathsf{p}(t, x, y)$. □

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 $g : H \to \mathbb{R}^d$:

$$\|g\|_{r,q,n} \equiv \|g\|_{r,q,n,2} = \|g\|_{\mathrm{L}^{r,q}(H)} + \frac{1}{\sqrt{n}}|g(0, x)|.$$

In the particular case that the function g does not depend on the time variable we let $\mathcal{S}(g)(s, x) := 0$.

Proposition 3. *Assume $\sigma \in \mathfrak{D} := \mathcal{C}_{\mathrm{b}}^{1,3}(\overline{H}) \cap \mathfrak{S}(\lambda, \Lambda)$, $b_\epsilon \in \mathfrak{M} := \mathcal{C}_{\mathrm{b}}^{1,3}(\overline{H}) \cap \mathfrak{B}(\Lambda)$ and f in $\mathcal{C}_{\mathrm{Sl}}^3(\mathbb{R}^d)$. Then there exist index sets $\mathcal{A}(\ell) \subset \prod_{i=1}^5 (1, \infty]^2$, $\ell \in \{0, \ldots, 3\}^4$, $\ell_1 + \cdots + \ell_4 \leq 4$, such that the* *following bound is satisfied:*

$$
\begin{aligned}
&|\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\overline{X}_T^\epsilon)]| \\
&\quad \le \frac{C(\|b_\epsilon\|_\infty)}{n}\Bigg(\sum_{\substack{\ell\in\{0,\ldots,3\}^4\\ \ell_1+\cdots+\ell_4\le 4}} \sum_{\prod_{i=1}^5 (r_i,q_i)\in\mathcal{A}(\ell)} \left\{ \prod_{i=1}^4 \|\partial_x^{\ell_i}\mathbf{b}_\epsilon\|_{r_i,q_i,n} + \|\mathbb{S}(\mathbf{b}_\epsilon)\|_{r_5,q_5,n} \times (1+\|\partial_x \mathbf{b}_\epsilon\|_{r_5,q_5,n}) \right\} \Bigg).
\end{aligned}
$$

The index sets $\mathcal{A}(\ell)$ are finite non-empty sets and can be chosen as follows.

Fix any $\ell \in \{0,\ldots,3\}^4$ such that $\ell_1+\cdots+\ell_4 \le 4$ and any combination of indices $(p_1,\ldots,p_5) \in (1,\infty)^5$ such that $(p_1,\ldots,p_4)$ are almost conjugate. Then each element $((r_1,q_1),\ldots,(r_5,q_5))$ of the finite set $\mathcal{A}(\ell)$ can be chosen freely as long as they satisfy the following rules

1. *For $i = 1,\ldots,4$, $(r_i,q_i) \in (p_i \vee 2,\infty]$ such that $\frac{d}{2q_i} + \frac{1}{r_i} < \frac{1}{p_i\vee 2}$.*
2. *$(r_5,q_5) \in (p_5,\infty]$ such that $\frac{d}{2q_5} + \frac{1}{r_5} < \frac{1}{p_5}$.*
3. *In the case that $\ell_1+\cdots+\ell_4 = 4$ then there exists $i \in \{1,\ldots,4\}$ such that $(r_i,q_i) \in (p_i,\infty]$ can be chosen as long as $\frac{d}{2q_i} + \frac{1}{r_i} < \frac{1}{p_i}$. That is, the stronger requirement $\frac{d}{2q_i} + \frac{1}{r_i} < \frac{1}{p_i\vee 2}$ can be exchanged by the one proposed here.*

The finite constant $C(\|b_\epsilon\|_\infty)$ depends on the choice of the sets $\mathcal{A}(\ell)$ and in particular on the indices $(p_i,r_i,q_i)_{i=1,\ldots,5}$ as well as λ, Λ and T.

Proof. In order to explicitly deal with the drift function, b_ϵ, and express processes on one single probability space, we perform Girsanov's transformations on both the solution process and the approximation process. Let $\Gamma_s^\epsilon := \sigma^{-1}b_\epsilon(s,X_s^\epsilon)$ and $\widetilde{\Gamma}_s^\epsilon := \sigma^{-1}b_\epsilon(\phi(s),\overline{X}_{\phi(s)})$. Accordingly, we define the measures $\mathbb{Q}^{\Gamma^\epsilon}$ and $\mathbb{Q}^{\widetilde{\Gamma}^\epsilon}$ as in (9). Then under $\mathbb{Q}^{\Gamma^\epsilon}$, $X^\epsilon \overset{\mathcal{L}}{=} Y$ and under $\mathbb{Q}^{\widetilde{\Gamma}^\epsilon}$, $\overline{X}^\epsilon \overset{\mathcal{L}}{=} \overline{Y}$.

Therefore, in order to analyze the weak error, we consider the following decomposition:

$$\mathbb{E}\left[f(X_T^\epsilon)\right] - \mathbb{E}\left[f(\overline{X}_T^\epsilon)\right] = \mathbb{E}\left[\mathcal{Z}(\Gamma^\epsilon)_T f(Y_T)\right] - \mathbb{E}\left[\mathcal{Z}(\widetilde{\Gamma}^\epsilon)_T f\left(\overline{Y}_T\right)\right]. \tag{15}$$

We can find the rate of convergence to zero of the above expression using an extended system of stochastic equation 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 a non-homogeneous diffusion because

$$V_t^\epsilon = \int_0^t (\sigma^{-1}b_\epsilon)^*(s,Y_s)\,\mathrm{d}W_s - \frac{1}{2}\int_0^t b_\epsilon^1(s,Y_s)\,\mathrm{d}s.$$

This system is approximated using $(\bar{Y},\bar{V}^\epsilon) := (\overline{Y},\log(\mathcal{Z}(\widetilde{\Gamma}^\epsilon)_T))$ with

$$\bar{V}_t^\epsilon = \int_0^t (\sigma^{-1}b_\epsilon)^*(\phi(s),\bar{Y}_{\phi(s)})\,\mathrm{d}W_s - \frac{1}{2}\int_0^t b_\epsilon^1(\phi(s),\bar{Y}_{\phi(s)})\,\mathrm{d}s.$$

The theory developed in Section 4.1 of [19] allows us to write and analyze the error (15) 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 [19]. Yet, we have to find the exact dependence on b_ϵ and its derivatives of each term in the error expansion of the difference $g\left(Y_T, \log \mathcal{Z}(\Gamma^\epsilon)_T\right) - g\left(\overline{Y}_T, \log \mathcal{Z}(\widetilde{\Gamma}^\epsilon)_T\right)$.

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

$$g\left(Y_T, \log \mathcal{Z}(\Gamma^\epsilon)_T\right) - g\left(\overline{Y}_T, \log \mathcal{Z}(\widetilde{\Gamma}^\epsilon)_T\right) = \sum_{i,j=0}^{d} c^0_{i,j}(T) \int_0^T c^1_{i,j}(t) \int_{\phi(t)}^t c^2_{i,j}(s)\, \mathrm{d}W^i_s\, \mathrm{d}W^j_t$$
$$+ \sum_{i,j,k=0}^{d} c^0_{i,j,k}(T) \int_0^T c^1_{i,j,k}(t) \left[\int_0^t c^2_{i,j,k}(s) \left(\int_{\phi(s)}^s c^3_{i,j,k}(u)\, \mathrm{d}W^i_u\right) \mathrm{d}W^j_s\right] \mathrm{d}W^k_t, \quad (16)$$

for some explicit adapted processes $\{(c^{i_1}_{i,j}(t), c^{i_2}_{i,j,k}(t))_{t\geq 0} : 0 \leq i, j, k \leq d, 0 \leq i_1 \leq 2, 0 \leq i_2 \leq 3\}$, where by abuse of notation, $\mathrm{d}W^0_t$ denotes $\mathrm{d}t$. In the above expansion, the terms $c^{i_1}_{i,j}, c^{i_2}_{i,j,k}$ are computed in the same spirit as in the proof of Lemma 4.2 in [19].

The explicit calculations are long and tedious[2]. The above expansion is obtained by using the mean value theorem for $(Y - \bar{Y}, V^\epsilon - \bar{V}^\epsilon)$ and solving explicitly the linear equation it satisfies. Therefore we only give here the dependence of each of the above terms on b_ϵ and its derivatives:

1) $c^0_{i,j}$ and $c^0_{i,j,k}$ depend linearly on $\exp\left(\theta \log(\mathcal{Z}(\Gamma^\epsilon)_T) + (1-\theta)\log(\mathcal{Z}(\widetilde{\Gamma}^\epsilon)_T)\right)$ with a random coefficient and no constant term. Therefore these terms depend on $b^*_\epsilon \sigma^{-1}$ and b^1_ϵ. Here θ denotes a uniform $[0,1]$ random variable independent of all the other random variables. To bound these terms, Lemma 10 is used.
2) $c^1_{i,j,k}$ linearly depends on $\nabla(b^*_\epsilon \sigma^{-1})$ with random coefficients, which are independent of b_ϵ.
3) $c^1_{i,j}$ are linear functions of $\nabla(b^*_\epsilon \sigma^{-1})$ $(j \neq 0)$, $\mathcal{S}(b^*_\epsilon \sigma^{-1})$ $(j \neq 0, i = 0)$, ∇b^1_ϵ $(j = 0)$ and $\mathcal{S}(b^1_\epsilon)$ $(i = 0, j = 0)$.
4) $c^2_{i,j}$, $c^2_{i,j,k}$ and $c^3_{i,j,k}$ do not depend on b_ϵ.

All the random coefficients mentioned above do not depend on b_ϵ but only depend upon σ, f and their derivatives up to order 3.

In fact, the so called *error process* $\mathcal{E}_t := Y_t - \overline{Y}_t$ may be written

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

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

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

The solution of the linear equation for $\mathcal{E}$ involves one stochastic integral. The mean value theorem applied to G gives a second stochastic integral linked to the difference $\overline{Y}_s - \overline{Y}_{\phi(s)} = \int_{\phi(s)}^s \sigma(\phi(s), \overline{Y}_{\phi(s)})\, \mathrm{d}W_s$. These two terms are associated with the two innermost integrals in (16). This explains why the corresponding inner coefficients do not depend on b_ϵ.

[2] A detailed calculation appears in [26] which is published as a side note. In particular, exact information about the set $\mathcal{A}(\ell)$ can be found through the proof.

In order to finish the argument and obtain the dependence on the derivatives of the drift function one follows the same steps as in the proof of Theorem 4.2 in section 4.5.1 in [19]. From there, one obtains the result by appropriately using Hölder's and Young's inequality together with Lemmas 1 and 2. Some of the needed Gaussian upper bound estimates have already been mentioned previously and others are obtained in Lemma 13 in the Appendix. In particular, the derivatives of b_ϵ appear due to the use of the iterative use of the duality relation of Malliavin Calculus in (16). For this reason, one needs the regularity assumptions on f, σ and b_ϵ.

The reason for the difference in restrictions for (q_1, r_1) and (q, r) arises because when duality is applied the stochastic differentiation of the term c^0 will give rise to stochastic integrals to which BDG inequalities are applied. For this reason all terms in the upper bound with norms $\|\cdot\|_{r_i,q_i,n}$ have the restriction that $i = 2$ when using Lemma 2. Otherwise, when possible, we try to use $L^1[0,T]$ norms (by using supremums in all processes that do not depend on the derivatives of $\mathbf{b}_\epsilon$), which corresponds to the unique case when the term c^1 is differentiated twice. For this reason, in the conclusion, we have the particular situation (described in 3.) with the indexes $\ell_1 + \cdots + \ell_4 = 4$ appearing in the final result. □

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 $d = 1$, $\sigma \in \mathfrak{D} := \mathcal{C}_b^{1,3}(\overline{H}) \cap \mathfrak{S}(\lambda, \Lambda)$, $\mathfrak{M} = \mathcal{C}_b^{1,3}(\overline{H})$, $\mathfrak{F} = \mathcal{C}_{Sl}^3(\mathbb{R})$.

We follow the guideline stated in Theorem 1.

The first error is measured in Proposition 1. This leads to the study of $\|b - b_\epsilon\|_{Y,p}$ and $\mathbb{E}\left[|f(Y_T) - \mu|^{p'}\right]^{1/p'}$ for an almost conjugate pair (p, p').

This last quantity is bounded due to Lemma 3. Next, Lemma 1 in the case $i = 2$ deals with the rate for $\|b - b_\epsilon\|_{Y,p}$. Therefore by (H4) this error is bounded by $\epsilon^{\frac{1}{q}}$ with $q, r \in (p \vee 2, \infty]$ with $\frac{1}{2q} + \frac{1}{r} < \frac{1}{p\vee 2}$. Considering that the parameters p, q and r can be chosen within the above restrictions one finds that any rate close to $\gamma = 1/2$ can be achieved in Theorem 1. In fact, taking $r = \infty$, $p = 2$ and q close to 2 one obtains this asymptotic rate.

Now we study the second error in Theorem 1 with the help of Proposition 3 in the case that the drift does not depend on time.

Note that the norms $\|\partial_x^i \mathbf{b}_\epsilon\|_{L^{r_i,q_i}(H)}$ are bounded according to Hypothesis (H4). First, $\|b_\epsilon\|_\infty$ is uniformly bounded. Next, $\mathcal{S}(\mathbf{b}_\epsilon) \equiv 0$, therefore considering the structure of the bound in Proposition 3 we see that it is enough to measure the rate for the product $A := \prod_{i=1}^4 \|\partial_x^1 \mathbf{b}_\epsilon\|_{r^i,q^i,n}$. To make this the main source for the rate, we further assume that $\lim_{\epsilon\downarrow 0} |\mathbf{b}_\epsilon - \mathbf{b}|(0, x) = 0$. This is a hypothesis of approximation at the starting point of the stochastic differential equation.

Finally, the above product gives as rate $A \leq C\epsilon^{-2+\sum_{i=1}^4 \frac{1}{2q_i}}$ with the restrictions on the q_i's given in Proposition 3, provided that the quadruple $(p_1, \ldots, p_4)$ are almost conjugate. Now, we have to find the right choice of parameters for the index set $\mathcal{A}(\ell)$ that will make the total error the smallest which satisfies the conditions stated in Proposition 3.

As we may take $r_i = \infty$, we choose $2q_i > d(p_i \vee 2)$ and $q_i > p_i \vee 2$ for $i = 2, 3, 4$ with the exception that $2q_1 > dp_1$ and $q_1 > p_1$. Given that $d = 1$, we can only take $(q_1, \ldots, q_4)$ as close as possible to $(p_1, p_2 \vee 2, \ldots, p_4 \vee 2)$. For any $\eta > 0$, choose $p_1 \geq 1$ and $p_i \geq 2$, $i = 2, 3, 4$ so that $1/p_1 + \cdots + 1/p_4 = 1 - \eta$. Since we could choose $q_i = p_i(1 + \eta)$, we then obtain a rate of

$C(\eta)\epsilon^{-3/2-2\eta/(1+\eta)}n^{-1}$ with a constant $C(\eta)$ which explodes as $\eta \to 0$. Asymptotically, by letting η converging to 0, the best (inaccessible) rate is $\epsilon^{-3/2}n^{-1}$.

Then in asymptotic terms, we can take $\eta = 1/2, \beta = 3/2, \delta = 1$ in Theorem 1. Therefore asymptotically, the optimal rate becomes $\kappa = \frac{1}{4}$.

Note that this result may be taken as a first indication that the rate in [37] may not be optimal. In fact, as the drift becomes irregular the rate does not degenerate as may be inferred from the above calculation.

Remark 5. Various other conclusions of the type above can be reached using the above argument. For example, the case $d = 1$ can be also obtained from the above calculation. Also, we have not optimized the proof of Proposition 3 in order to avoid a longer proof. Proposition 2 could also be used in order to improve the above result. We do not do this in order to avoid longer proofs and because the above methods probably do not lead to an optimal rate.

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

In all the previous sections, we have considered situations where a smooth approximation for the irregular drift coefficient is used. Therefore naturally, control on the higher order derivatives of the approximation b_ϵ are needed. In this section, we provide an alternative method which limits the number of higher order derivatives being used. 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. We also need stronger regularity conditions on σ so that $\sigma \in \mathfrak{H}(\lambda, \Lambda)$ through all this section.

We still use the same definitions for X, $\overline{X}$, X^ϵ, $\overline{X}^\epsilon$, Y and $\overline{Y}$. Note that under the above conditions the upper Gaussian estimate (11) is satisfied for the processes stated in Lemma 13. Therefore we may freely apply Lemmas 1 and 2.

In order the make of of Girsanov theorem, we introduce the additional intermediate approximation process

$$\widetilde{X}^\epsilon_t = x + \int_0^t \sigma(s, \widetilde{X}^\epsilon_s)\,\mathrm{d}B_s + \int_0^t b_\epsilon(\phi(s), \widetilde{X}^\epsilon_{\phi(s)})\,\mathrm{d}s.$$

When σ is constant, then $\widetilde{X}^\epsilon = \overline{X}^\epsilon$.

Our class of test functions is $\mathfrak{F} := \mathfrak{C}_{\mathrm{Sl}}(\mathbb{R}^d)$. We set $K_q(f) := \mathbb{E}[|f(Y_T) - \mathbb{E}[f(Y_T)]|^q]^{1/q} < \infty$ for $f \in \mathfrak{F}$ according to Lemma 3. For any two processes X^1 and X^2 we define

$$D_{g,p}(X^1, X^2_\phi) := \mathbb{E}\left[\left(\int_0^T |g(s, X^1_s) - g(\phi(s), X^2_{\phi(s)})|^2\,\mathrm{d}s\right)^{p/2}\right]^{1/p},$$

$$\bar{D}_{g,p}(X^1, X^2_\phi) := \mathbb{E}\left[\left(\int_0^T |g(\phi(s), X^1_s) - g(\phi(s), X^2_{\phi(s)})|^2\,\mathrm{d}s\right)^{p/2}\right]^{1/p}.$$

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,

$$|\mathbb{E}[f(Z^{(1,j)})] - \mathbb{E}[f(Z^{(2,j)})]| \leq C_1(p, p')\mathbb{E}\left[\left(\int_0^T |\beta^{(1,j)}(s) - \beta^{(2,j)}(s)|^2 \, \mathrm{d}s\right)^{p/2}\right]^{1/p} K_{p'}(f), \tag{17}$$

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

These estimates are used in 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 (p, p') an almost conjugate pair. Then there exists a positive constant C depending only on p, Λ and λ such that for $f \in \mathcal{C}_{\mathrm{Sl}}(\mathbb{R}^d)$*

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

Remark 6. The rate of the last term $\sqrt{\mathrm{Var}\,(f(Y_T) - f(\overline{Y}_T))}$ 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 [4] if f is a non-Lipschitz function.

The proof of this proposition is divided into 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, fix $p > 1$ so that (p, p') is an almost conjugate pair. Furthermore, let $q, r \in (p \vee 2, \infty]$ so that $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p\vee 2}$. Then there exists a constant C which depends only on p, q, r and the constants C_1 and C_2 appearing in* (11) *such that*

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

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^ϵ 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 $f \in \mathcal{C}_{\mathrm{Sl}}(\mathbb{R}^d)$, there exists a constant $C = C(p, p', \lambda, \Lambda)$ such that*

$$|\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\widetilde{X}_T^\epsilon)]| \leq C D_{b_\epsilon, p}(Y, Y_\phi) K_{p'}(f), \tag{18}$$

$$|\mathbb{E}[f(\widetilde{X}_T^\epsilon)] - \mathbb{E}[f(\overline{X}_T^\epsilon)]| \leq C \left(\left(\bar{D}_{\sigma^{-1} b_\epsilon, p}(Y_\phi, \overline{Y}_\phi) + D_{\sigma^{-1}, p}(Y, Y_\phi) \right) K_{p'}(f) + \sqrt{\mathrm{Var}\left(f(Y_T) - f(\overline{Y}_T)\right)} \right). \tag{19}$$

Proof. Eq. (18) is just a restatement of (17). To obtain (19), consider $\Gamma_1(s) := \sigma^{-1}(s, Y_s) b_\epsilon(\phi(s), Y_{\phi(s)})$ and $\Gamma_2(s) = \sigma^{-1}(\phi(s), \overline{Y}_{\phi(s)}) b_\epsilon(\phi(s), \overline{Y}_{\phi(s)})$. Therefore,

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

For an almost conjugate pair (p, p'), Lemma 11 yields

$$A_1 \leq C \mathbb{E}\left[\left(\int_0^T |\sigma^{-1}(s, Y_s) b_\epsilon(\phi(s), Y_{\phi(s)}) - \sigma^{-1}(\phi(s), \overline{Y}_{\phi(s)}) b_\epsilon(\phi(s), \overline{Y}_{\phi(s)})|^2 \, \mathrm{d}s \right)^{p/2} \right]^{1/p} K_{p'}(f).$$

Combining (20) and (21) applied on σ^{-1} which belongs to $\mathrm{H}^{1/2,1}(\overline{H})$ thanks to Lemma 12,

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

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

$$A_2 \leq C \sqrt{\mathrm{Var}\left(f(Y_T) - f(\overline{Y}_T)\right)}.$$

From here the result follows. □

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 $\sigma \in \mathfrak{H}(\lambda, \Lambda)$. For any $p \in (1, \infty)$, $r, q \in (p \vee 2, \infty]$ such that $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p \vee 2}$, then there exists some positive constant $C = C(p, q, r, \lambda, \Lambda, \rho, T)$ such that for any $\rho \in (0, 1/2)$,*
(I) *For every function g with $\|S(g)\|_{\mathrm{L}^{r,q}(H)} < \infty$ and $\|\partial_x g(\phi(\cdot), \cdot)\|_{\mathrm{L}^{r,q}(H)} < \infty$, it holds that*

$$D_{g,p}(Y, Y_\phi) \leq \frac{C}{n^{1/2-\rho}} \left(\|S(g)\|_{\mathrm{L}^{r,q}(H)} + \|\partial_x g(\phi(\cdot), \cdot)\|_{\mathrm{L}^{r,q}(H)} \right) \tag{20}$$

$$\text{and } D_{g,p}(\overline{Y}, \overline{Y}_\phi) \leq \frac{C}{n^{1/2-\rho}} \left(\|S(g)\|_{r,q,n,2} + \|\partial_x g(\phi(\cdot), \cdot)\|_{r,q,n,2} \right).$$

(II) *Suppose that for any $\alpha \in [0,1]$, $Y_t^\alpha = \alpha Y_t + (1-\alpha)\overline{Y}_t$ satisfies the Gaussian upper bound* (11). *Then for every function g with $\|\partial_x g(\phi(\cdot),\cdot)\|_{\mathrm{L}^{r,q}(H)} < \infty$,*

$$\bar{D}_{g,p}(Y,\overline{Y}) \le \frac{C\|\partial_x g(\phi(\cdot),\cdot)\|_{\mathrm{L}^{r,q}(H)}}{\sqrt{n}}. \tag{21}$$

Proof. Without loss of generality, we may assume that $p > 2$. We do the proof using a general process Z which can be later applied for $Z = Y$ and $Z = \bar{Y}$.

(I) Using Minkowski's and Hölder's inequalities with a conjugate pair (a, a') together with Lemma 1,

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T |g(s,Z_s) - g(\phi(s),Z_{\phi(s)})|^2 \,\mathrm{d}s\right)^{p/2}\right]^{1/p} &\le Cn^{-1/2}\mathbb{E}\left[\left(\int_0^T \frac{|g(s,Z_s)-g(\phi(s),Z_s)|^2}{s-\phi(s)}\,\mathrm{d}s\right)^{p/2}\right]^{1/p} \\
+ \mathbb{E}\left[\left(\int_0^T \left(\frac{|g(\phi(s),Z_s)-g(\phi(s),Z_{\phi(s)})|}{|Z_s - Z_{\phi(s)}|}\right)^2 \mathrm{d}s\right)^{ap/2}\right]^{1/ap} &\mathbb{E}\left[\sup_{s\in[0,T]} |Z_s - Z_{\phi(s)}|^{a'p}\right]^{1/a'p} \\
\le C\left(n^{-1/2}\|S(g)\|_{\mathrm{L}^{r_1,q_1}(H)} + \|\partial_x g(\phi(\cdot),\cdot)\|_{\mathrm{L}^{r_2,q_2}(H)}\mathbb{E}\left[\sup_{s\in[0,T]} |Z_s - Z_{\phi(s)}|^{a'p}\right]^{1/a'p}\right).
\end{aligned} \tag{22}$$

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

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 $\overline{F}(x) := \mathbb{P}\left[\sup_{s\in[0,T]} |Z_s - Z_{\phi(s)}|^b > x\right]$. Then for any $\rho \in (0, \frac{1}{2})$,

$$\mathbb{E}\left[\sup_{s\in[0,T]} |Z_s - Z_{\phi(s)}|^b\right] = \int_0^{n^{-b(\frac{1}{2}-\rho)}} \overline{F}(x)\,\mathrm{d}x + \int_{n^{-b(\frac{1}{2}-\rho)}}^\infty \overline{F}(x)\,\mathrm{d}x \le n^{-b(\frac{1}{2}-\rho)} + \int_{n^{-b(\frac{1}{2}-\rho)}}^\infty \overline{F}(x)\,\mathrm{d}x. \tag{23}$$

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

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

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

$$\int_{n^{-b(\frac{1}{2}-\rho)}}^\infty \overline{F}(x)\,\mathrm{d}x \le C\int_{n^{-b(\frac{1}{2}-\rho)}}^\infty ne^{-cx^{2/b}n}\,\mathrm{d}x \le Cn^{1-\frac{b}{2}}\int_{n^{b\rho}}^\infty e^{-cu^{2/b}}\,\mathrm{d}u.$$

The above integral on the right hand side converges to zero faster than $n^{-b(\frac{1}{2}-\rho)}$ as $n \to \infty$, so that from (23), we have that the second term in equation (22) converges to zero with order $n^{-b(\frac{1}{2}-\rho)}$ and the conclusion follows.

(II) The proof is similar as the above proof except that one uses the strong rate of convergence (see, e.g., [24]) which yields that $\sup_{t\in[0,T]}\mathbb{E}[|Y_t - \overline{Y}_t|^2]^{1/2} \le C/\sqrt{n}$. Hence the result. □

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

Theorem 3. *Fix constants (p, q, r) such that $p \in (1, \infty)$ with $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p\vee 2}$ and $q, r \in (p \vee 2, \infty)$. Under the same hypothesis of Proposition 4 and $f \in \mathcal{C}^1_{\mathrm{Sl}}(\mathbb{R}^d)$, there exists a positive constant C which depends only on q, r, λ, Λ, T and the constants C_1 and C_2 appearing in* (11) *such that for any $\rho \in (0, 1/2)$,*

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

Proof. Given the conditions on σ,

$$
\frac{\|\partial_x(\sigma^{-1}b_\epsilon)(\phi(\cdot),\cdot)\|_{\mathrm{L}^{r,q}(H)}}{\sqrt{n}} \leq \frac{C(\lambda,\Lambda)}{n^{1/2-\rho}}\left(\|b_\epsilon(\phi(\cdot),\cdot)\|_{\mathrm{L}^{r,q}(H)} + \|\partial_x b_\epsilon(\phi(\cdot),\cdot)\|_{\mathrm{L}^{r,q}(H)}\right).
$$

Hence the result. □

Conclusion 2. Let $\sigma \in \mathfrak{H}(\lambda,\Lambda)$, $\mathfrak{M} = H^{1/2,1}(\overline{H})$, $\mathfrak{F} = \mathcal{C}_{\mathrm{Lip}}(\mathbb{R}^d)$, 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 $\epsilon^{-\left(\frac{1}{2}(1-\frac{1}{q})\right)}n^{-(\frac{1}{2}-\rho)} + \epsilon^{1/q}$.

Applying Theorem 1 with $\gamma = 1/q$, $\beta = \frac{1}{2}(1-\frac{1}{q})$ and $\delta = \frac{1}{2} - \rho$ we achieve a rate close to $n^{-1/3}$ by taking q close to 2 and ρ close to zero.

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

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

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

$$
|\mathbb{E}[f(\overline{X}_T)] - \mathbb{E}[f(\overline{X}^\epsilon_T)]| \leq CK_{p'}(f)\left(\|b - b_\epsilon\|_{\mathrm{L}^{r,q}(H)} + \frac{|b - b_\epsilon|(0,x)}{n^{1/2}}\right).
$$

Proof. We apply (17) with $Z^{(1,2)} = \overline{X}_T$ and $Z^{(2,2)} = \overline{X}^\epsilon_T$ so that

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

Therefore the result is a direct consequence of Lemma 2 for $i = 2$. □

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 $\sigma \in \mathfrak{H}(\lambda, \Lambda)$ and $b \in \mathfrak{B}(\Lambda)$ are time homogeneous. 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 $b_\epsilon \in \mathfrak{B}(\Lambda)$ is a Lipschitz function on $G(\epsilon)^c$ with Lipschitz constant independent of ϵ, the Lipschitz constant of b_ϵ on $G(\epsilon)$ is ϵ^{-1} and $\mathrm{meas}(G(\epsilon)) = \mathrm{O}(\epsilon^\kappa)$ for some $\kappa > 0$. Then for any $f \in \mathfrak{C}_{\mathrm{Sl}}(\mathbb{R}^d)$, $p \in (1, \infty)$, $q \in (p \vee 2, \infty)$ such that $\frac{d}{2q} < \frac{1}{p\vee 2}$, there exists a positive constant C which depends on p, q, λ, Λ, T and κ such that*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)]| \le C\left(K_p(f)\epsilon^{\kappa/q}\left(\frac{1}{\epsilon n^{1/2-\rho}} + 1\right) + \sqrt{\mathrm{Var}\left(f(Y_T) - f(\overline{Y}_T)\right)}\right).$$

Proof. We combine Lemma 6 with Theorem 3. For this, we compute each error term as follows.

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

Next, we need to compute $\|b_\epsilon(\phi(\cdot), \cdot)\|_{\mathrm{L}^{r,q}(H)}$ and $\|\partial_x b_\epsilon(\phi(\cdot), \cdot)\|_{\mathrm{L}^{r,q}(H)}$. As they are similar we only compute the second. We obtain

$$\|\partial_x b_\epsilon(\phi(\cdot), \cdot)\|_{\mathrm{L}^{r,q}(H)} \le C\epsilon^{-1}\epsilon^{\kappa/q}.$$

Putting all the estimates together, for any choice of an almost conjugate pair (p, p'), $q, r \in (p \vee 2, \infty]$ such that $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p\vee 2}$, there exists constant $C = C(p, p', q, r, \lambda, \Lambda)$ such that for any $\rho \in (0, 1/2)$,

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)]| \le CK_p(f)\epsilon^{\kappa/q}\left(\frac{1}{\epsilon n^{1/2-\rho}} + 1\right) + C\sqrt{\mathrm{Var}(f(Y_T) - f(\overline{Y}_T))} + CK_{p'}(f)\frac{|b - b_\epsilon|(0, x)}{n^{1/2}}.$$

Hence the result. □

Conclusion 3. Considering the proof carefully, one may say that the approximations of the drift are taken on the space $\mathfrak{M} = \mathrm{H}^1(\mathbb{R}^d)$. If $\sigma \in \mathfrak{S} = \mathfrak{H}(\lambda, \Lambda)$ and $\mathfrak{F} = \mathfrak{C}^1_{\mathrm{Sl}}(\mathbb{R}^d)$ then for $\kappa > q$, for some q satisfying the hypothesis of the above theorem then the convergence rate is $n^{-1/2+\rho}$.

In the case that $\kappa < q$ for any q satisfying the hypothesis of the above Theorem, we have that the rate is $n^{-(\frac{1}{2}-\rho)\frac{\kappa}{q}}$.

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

$$\mathrm{d}X_t(x) = b(X_t(x))\,\mathrm{d}t + \mathrm{d}W_t,\ X_0(x) = x,\ \text{with } b(x) := \begin{cases} -\theta, & x > 0, \\ 0, & x = 0, \\ \theta, & x < 0 \end{cases}$$

for $\theta > 0$. As before $\overline{X}(x)$ 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, see [23, Section 6.5]. From [23, Section 6.5, (5.14), p. 441], the transition density function of $X_t(x)$, $x \geq 0$ is

$$\mathsf{p}_t(x,z) = \begin{cases} h_-(x-z,t) + \dfrac{\theta g_-(z)}{\sqrt{2\pi t}} \displaystyle\int_{x+z}^{\infty} \exp\left(-\frac{(v-\theta t)^2}{2t}\right) \mathrm{d}v \ \text{ for } \ z > 0, \\ g_+(x)h_+(x-z,t) + \dfrac{\theta g_+(z)}{\sqrt{2\pi t}} \displaystyle\int_{x-z}^{\infty} \exp\left(-\frac{(v-\theta t)^2}{2t}\right) \mathrm{d}v \ \text{ for } \ z \leq 0. \end{cases} \tag{24}$$

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 $f \in \mathcal{C}^3_{\mathrm{p}}(\mathbb{R})$ an* even *function, the weak error satisfies*

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

where C is a positive constant.

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 $f \in \mathcal{C}^2_{\mathrm{p}}(\mathbb{R};\mathbb{R})$ 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 $k = 0,1,2$,*

$$\sup_{t\in[0,T]} \sup_{x\in\mathbb{R}} |\partial_x^k u(t,x)| \leq C, \tag{25}$$

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

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x) + b(x)\frac{\partial u}{\partial x}(t,x) = 0, \ u(T,x) = f(x). \tag{26}$$

(ii). Assume that $f \in \mathcal{C}^3_{\mathrm{p}}(\mathbb{R};\mathbb{R})$ 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

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

where C is a positive constant and ∂_{x+}, ∂_{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,

$$u(t,x) = \int f(z)\mathsf{p}_{T-t}(x,z)\,\mathrm{d}z.$$

Here $\mathsf{p}_t(x, z)$ denotes the density associated with the process $X_t(x)$ at the point z whose formula was given in (24).

From this formula one can verify that the density $\mathsf{p}_t(x, z)$ 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 $x \in \mathbb{R} \setminus \{0\}$ and any $t < T$ due to the successive application of the integration by parts formula. In particular, u satisfies the PDE (26). □

Let us denote by $\overline{\mathsf{p}}_t(x, y)$ 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. *For any $x \in \mathbb{R}$, $\overline{\mathsf{p}}_{\phi(t)}(0, x) = \overline{\mathsf{p}}_{\phi(t)}(0, -x)$. 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 $\overline{\mathsf{p}}_{\phi(t)}(0, x) \leq C/\sqrt{\phi(t)}$.*

The proof of Theorem 5 is achieved by using a Taylor expansion theorem on the solution of the partial differential equation (26) taking into account the fact that the space variable may be on the positive or negative real axis. Then Lemma 8 is used to show the integrability of each term.

8.4 Numerical Results

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

$$X_t = x + \int_0^t b(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s,$$

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 θ_0, θ_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 [25].

9 Conclusions

We have given several weak orders for the approximation problem $|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}^{\epsilon}_T)]|$ with non-smooth drift coefficients by analyzing the errors of $|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X^{\epsilon}_T)]|$ (Propositions 1 and 2 as well as Corollary 1), and $|\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\overline{X}_T^\epsilon)]|$ (Propositions 3 and Proposition 5) under various assumptions to b, b_ϵ, σ and f.

Two types of proof have been given. One is based on the proof of weak approximation and another one is based on the proof of strong approximation. From the *Conclusions* we have drawn, one can hint that the rates given in [24] for the case of non-smooth coefficients are not optimal. We have tried to show this using the above two methods of proof.

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 finished 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.

In fact, in the recent months a number of articles dealing with this matter have appeared in the form of preprints which assure the interest in this problem (see *e.g.* [34, 38, 39, 41, 42], ...). Among them, let us note [28] which uses a related approach to control the error on the densities.

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 $(\omega, s) \in \Omega \times [0, T]$. Then*
(I) *For any $\alpha \in \mathbb{R}$, and $t \in [0, T]$,*

$$\mathbb{E}[\mathcal{Z}(\Gamma)_t^\alpha] \leq \exp\left(\left|\alpha\left(\alpha - \frac{1}{2}\right)\right| t\hat{\Gamma}^2\right). \tag{27}$$

(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 $g \in \mathrm{L}^p(Y)$ for some $p < +\infty$. For any $1 \leq q < p$,*

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

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

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

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

Since Γ 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 (27).
(II) We use the Burkholder-Davis-Gundy (BDG) inequality, Hölder's inequality and (I) to obtain

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

(III) For $\alpha = p/q > 1$ and its conjugate $\alpha' = \alpha/(\alpha-1)$, Hölder's inequality and (27) yields

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

This achieves the proof of the lemma. □

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

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

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

The same inequality holds for b_ϵ instead of b.

Lemma 11. *Let Γ and Γ' 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 $p > \alpha$,*

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

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

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

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 $t \in [0, T]$

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

Choosing appropriately the values of q and a one obtains the result. □

10.2 A lemma on Hölder functions

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

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

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

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

The required continuity follows easily since $\|\mathrm{Id} - \Lambda^{-1}a(t,x)\|^k \le (1 - \Lambda^{-1}\lambda)^k$ for any $(t,x) \in \overline{H}$ and $k \ge 0$. The same decomposition holds for σ^{-1}. □

10.3 Some Gaussian upper bound estimates

The next lemma establishes the upper bound for the random variables $Y_t^\alpha = \alpha Y_t + (1-\alpha)\overline{Y}_t$ and $\overline{Y}_t^\alpha = \alpha\overline{Y}_t + (1-\alpha)\overline{Y}_{\phi(t)}$.

Lemma 13. *Under the assumptions of Proposition 3, the density of* Y_t^α *exists for all* $\alpha \in [0,1]$ *and satisfies the Gaussian upper bound estimate* (11) *uniformly in* α*. Similarly, for the variable* $\overline{Y}_t^\alpha$*, for* $t \ge t_1$ *its density* $\overline{\mathsf{p}}(t,x,y)$ *exists and*

$$\overline{\mathsf{p}}(t,x,y) \le C_1 \mathsf{g}_{C_2\phi(t)}(x-y),\ \forall (t,x,y) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d,$$

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

Proof. The proof in the case of Y_t^α is somewhat standard in Malliavin Calculus theory. One first uses the Lemma 7.1 in [27] to estimate the Malliavin covariance matrix of Y_t^α, together with classical martingale exponential inequalities, such as Lemma A.5 in [40].

For the second case ($\bar{Y}_t^\alpha$), one needs to use the results of [33] for $\overline{Y}_{\phi(t)}$ and explicitly convolute this upper bound with the explicit transition density of $\alpha(\overline{Y}_t - \overline{Y}_{\phi(t)})$. Finally the uniform ellipticity condition together with explicit bounds on the ratio $\frac{\phi(t)}{\phi(t)+\alpha(t-\phi(t))}$ gives the result. □

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] 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.

[4] R. Avikainen, *On irregular functionals of SDEs and the Euler scheme*, Finance Stoch. **13** (2009), no. 1, 381–401.

[5] 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.

[6] ———, *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.

[7] R.F. Bass, *Diffusions and Elliptic Operators*, Springer, 1998.

[8] 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.

[9] 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.

[10] David Dereudre, Sara Mazzonetto, and Sylvie Roelly, *An explicit representation of the transition densities of the skew Brownian motion with drift and two semipermeable barriers*, Monte Carlo Methods Appl. **22** (2016), no. 1, 1–23, DOI 10.1515/mcma-2016-0100.

[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 solutions of one-dimensional stochastic differential equations with discontinuous drift*, ESAIM Probab. Stat. **18** (2014), 686–702, DOI 10.1051/ps/2013053.

[13] E. B. Fabes and C. E. Kenig, *Examples of singular parabolic measures and singular transition probability densities*, Duke Math. J. **48** (1981), no. 4, 845–856, DOI 10.1215/S0012-7094-81-04846-8.

[14] M. Frazier and B. Jawerth, *Decomposition of Besov spaces*, Indiana Univ. Math. Journ. **34** (1985), 777-799.

[15] 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.

[16] C. Geiss, S. Geiss, and E. Gobet, *Generalized fractional smoothness and L_p-variation of Bsdes with non-Lipschitz terminal condition*, Stochastic Proces. Appl. **122** (2012), 2078-2116.

[17] S. Geiss and E. Gobet, *Fractional smoothness and applications in finance*, Advanced mathematical methods for finance, Springer, Heidelberg, 2011, pp. 313–331.

[18] 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.

[19] 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.

[20] 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.

[21] ———, *A note on the Euler-Maruyama scheme for stochastic differential equations with a discontinuous monotone drift coefficient*, BIT **48** (2008), no. 1, 51–59.

[22] R. Janssen, *Difference methods for stochastic differential equations with discontinuous coefficients*, Stochastics **13** (1984), no. 3, 199–212.

[23] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus (2nd)*, Springer, 1998.

[24] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Applications of Mathematics (New York), vol. 23, Springer-Verlag, Berlin, 1992.

[25] 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.

[26] ———, *Detailed proof of Proposition 3 for a weak error result with non regular drift* (2016).

[27] A. Kohatsu-Higa, *Weak Approximations: A Malliavin Calculus approach*, Mathematics of Computation **70** (2001), 135–172.

[28] V. Konakov and S. Menozzi, *Weak error for the Euler scheme approximation of diffusion with non-smooth coefficient* (2016), available at `arxiv:1604.00771`.

[29] N. Krylov, *An inequality in the theory of stochastic integrals*, Th. Probab. Applic. **16** (1971), 438–448.

[30] S. Kusuoka, *Hölder continuity and bounds for fundamental solutions to nondivergence form parabolic equations*, Anal. PDE **8** (2015), no. 1, 1–32, DOI 10.2140/apde.2015.8.1.

[31] G.C. Kyriazis, *Approximations from Shift-Invariant Spaces*, Constr. Approx. **11** (1995), 141–164.

[32] O.A. Ladyženskaja, V.A. Solonnikov, and N.N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type*, American Mathematical Society, 1967.

[33] V. Lemaire and S. Menozzi, *On some non asymptotic bounds for the Euler scheme*, Electron. J. Probab. **15** (2010), no. 53, 1645–1681.

[34] Gunther Leobacher and Michaela Szölgyenyi, *A numerical method for SDEs with discontinuous drift*, BIT **56** (2016), no. 1, 151–162, DOI 10.1007/s10543-015-0549-x.

[35] M. Martinez and D. Talay, *One-dimensional parabolic diffraction equations: pointwise estimates and discretization of related stochastic differential equations with weighted local times*, Electronic Journal Probab. **17** (2012), no. 27, DOI 10.1214/EJP.v17-1905.

[36] R. Mikulevičius and C. Zhang, *Weak Euler Approximation for Itô Diffusion and Jump Processes*, Stoch. Anal. Appl. **33** (April 2015), no. 3, 549–571, DOI 10.1080/07362994.2015.1014102.

[37] 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.

[38] H.-L. Ngo and D. Taguchi, *Strong rate of convergence for the Euler-Maruyama approximation of stochastic differential equations with irregular coefficients*, Math. Comp. **85** (2016), no. 300, 1793–1819, DOI 10.1090/mcom3042.

[39] ———, *Approximation for non-smooth functionals of stochastic differential equations with irregular drift* (May 2015), 1–20 pp., available at `arxiv:1505.03600`.

[40] D. Nualart, *The Malliavin calculus and related topics*, 2nd ed., Probability and its Applications (New York), Springer-Verlag, 2006.

[41] O. Papaspiliopoulos, G. O. Roberts, and K. B. Taylor, *Exact sampling of diffusions with a discontinuity in the drift*, Adv. in Appl. Probab. **48** (2016), no. A, 249–259, DOI 10.1017/apr.2016.54.

[42] P. 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.

[43] 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.

[44] A. I. Stepanets, *Methods of Approximation Theory*, VSP, 2005.

[45] 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.

[46] D.W. Stroock and S.R.S. Varadhan, *Multidimensional Diffusion Processes*, Springer, 1979.

[47] H. Triebel, *Theory of function spaces II*, Birkhauser, Basel, 1992.

[48] 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.

[49] L. Yan, *The Euler scheme with irregular coefficients*, Ann. Probab. **30** (2002), no. 3, 1172–1194.