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

Arturo Kohatsu-Higa* Antoine Lejay† Kazuhiro Yasuda‡

July 1, 2013

Abstract

We consider an Euler-Maruyama type approximation methods for a Stochastic Differential Equation (SDE) with a discontinuous drift and regular diffusion coefficient. The method regularizes the drift coefficient within a certain class of functions and then the Euler-Maruyama scheme for the regularized scheme is used as an approximation. This methodology gives two errors, the first is the error of regularization of the drift coefficient within a given class of functions and the second is the error of the regularized Euler-Maruyama scheme. This second error rate will be determined by the method of proof which takes into account that the regularized stochastic differential equation have coefficients in a given class of functions. After an optimization procedure with respect to this parameter we obtain various rates, which improve other known results but all of them of order less than one, which depend on the approximation class used.

Finally, we consider a particular example of a diffusion with an explicit discontinuous drift and we show that the weak rate of convergence of the Euler-Maruyama scheme is the same as in the classical Euler-Maruyama scheme with smooth coefficients.

Keywords. Stochastic Differential Equation, Euler-Maruyama scheme, discontinuous drift, weak rate of convergence, Malliavin calculus.

1 Introduction

The Euler-Maruyama scheme is a simple and efficient numerical scheme to simulate solutions of the multi-dimensional Stochastic Differential Equations (SDEs) defined by

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

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

*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.

of functions. For a fixed number n of steps, consider n independent Gaussian random vectors $\xi_1, \dots, \xi_n$ with zero mean and identity variance matrix. The Euler-Maruyama scheme consists in computing iteratively for $k = 0, \dots, n-1$ with $t_k = \frac{kT}{n}$,

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

Then $\mathbb{E}[f(X_T)]$ is approximated by $\mathbb{E}[f(\overline{X}_T)]$. This latter quantity is also approximated by an empirical mean over N samples of $\overline{X}_T$.

Notwithstanding the Monte Carlo error, the error induced by the use of this scheme may be assessed by two ways: the strong error and the weak error. The *strong error* consists in computing $\mathbb{E}[|X_T - \overline{X}_T|^2]^{1/2}$ or $\mathbb{E}[\sup_{t\in[0,T]} |X_t - \overline{X}_t|^2]^{1/2}$. This is used when one intends to analyze some path properties of X. With enough regularity on the coefficients b and σ, this error decreases to zero with the step n at rate $n^{-1/2}$ [22].

The *weak error* is defined as

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

for f in a given class of functions $\mathfrak{F}$. If the class $\mathfrak{F}$ is included in the class of Lipschitz functions then the weak error should be at least $n^{-1/2}$. At the price of higher complexity, using results on partial differential equations or Malliavin calculus, this rate of convergence could be however improved. With enough regularity on the coefficients σ and b and assuming that σ is uniformly elliptic, $d_f(X, \overline{X})$ converges to 0 at a rate n^{-1}, even if $\mathfrak{F}$ is the class of bounded measurable functions or Dirac distribution functions [5, 6, 29]. This fact is a non-trivial consequence of the regularity of the density of X_T.

This difference in convergence rate has many practical consequences in the design of various simulation methods.

Weakening the conditions on the coefficients b or σ naturally imply slower convergence rates. For example, for α-Hölder continuous coefficients, the strong error converges at a rate that decreases to zero as α decreases [17]. A similar property is known for weak rate of convergence (see [29] and section 8.2), which is summarized by the following theorem (For exact definitions, we refer to Section 2).

Theorem 1 (R. Mikulevicius and E. Platen [29]). *If for $\alpha \in (0,1) \cup (1,2) \cup (2,3)$, $a, b \in \mathrm{H}^{\alpha/2,\alpha}(\overline{H})$ and $f \in \mathrm{H}^{2+\alpha}(\mathbb{R}^d)$, then there exists a constant K such that*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)]| \leq \frac{K}{n^{E(\alpha)}}.$$

Here $E(\alpha) = \frac{\alpha}{2}\mathbf{1}(\alpha \in (0,1)) + \frac{1}{3-\alpha}\mathbf{1}(\alpha \in (1,2)) + \mathbf{1}(\alpha \in (2,3))$. Besides, the constant K is linear in $\|b\|_{\mathrm{H}^{\alpha/2,\alpha}}$ and $\|a\|_{\mathrm{H}^{\alpha/2,\alpha}}$.

In fact, we will prove that the above result is sub-optimal in many respects. Even more, few articles have been devoted to the convergence of the Euler-Maruyama scheme when the drift coefficient b presents some discontinuities [2, 4, 9, 18–20, 32, 36]. In these articles, a method of approximation is proposed, and then the convergence in strong or weak sense is studied but no rate of convergence is provided. In [11], P. Étoré and M. Martinez have constructed an exact simulation scheme for a SDE with a piecewise constant drift. But this approach is restricted to dimension one.

F. Bernardin *et al.* in [8] provides estimates on the mean passage time at small balls of a process which is observed discretely. As shown in the Appendix of [8], this result combined with the computations of [28] may lead to a weak rate of convergence for a particular type of SDE.

On the other hand, in Theorem 1, the order of Hölderianity of a, b and f have been carefully chosen so as to generate an “optimal” rate. In fact, as we will see later there is a strong interplay between the regularity of f and the regularity of the coefficients which then generates the rate of convergence beyond the statement of the above theorem.

Therefore the goal of this article is to provide bounds on the weak rate of convergence for the Euler-Maruyama scheme of the corresponding SDE 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.

Let us present our framework. The distance between the drift coefficient and its approximation b_ϵ in the class $\mathfrak{M}$ is measured by a parameter γ. When one uses the Euler scheme with a regularized drift in $\mathfrak{M}$ and a time step T/n, the weak rate of convergence is expressed by a parameter δ, to say that the rate of convergence is C_ϵ/n^δ. However, the constant C_ϵ depends itself of $\mathfrak{F}$ and of the drift in $\mathfrak{M}$. It explodes as $\epsilon \to 0$ as $b \notin \mathfrak{M}$. The rate of explosion is characterized by a parameter β.

Let $X(b_\epsilon)$ be the process with a regularized drift b_ϵ and $\overline{X}(b_\epsilon)$ is its Euler scheme (i.e. the process defined in (2) with b_ϵ instead of b).

Introducing a new constant q which depends on the regularity of $f \in \mathfrak{F}$ and some constants $K_q(f)$ and $\widehat{K}(f)$, we then write (See Figure 1)

$$d_f(X(b), X(b_\epsilon)) = \mathrm{O}(K_q(f)\epsilon^\gamma) \text{ and } d_f(X(b_\epsilon), \overline{X}(b_\epsilon)) = \mathrm{O}(\widehat{K}(f)\epsilon^{-\beta}n^{-\delta}).$$

Finally, carrying an optimization on the value of ϵ one obtains the best possible rate in this framework.

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.

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.

$$\mathrm{d}\overline{X}^\epsilon_t = \sigma(\phi(t), \overline{X}^\epsilon_{\phi(t)})\,\mathrm{d}B_t + b_\epsilon(\phi(t), \overline{X}^\epsilon_{\phi(t)})\,\mathrm{d}t \xrightarrow[n \nearrow \infty]{\mathrm{O}(\widehat{K}(f)\epsilon^{-\beta}n^{-\delta})} \mathrm{d}X^\epsilon_t = \sigma(t, X^\epsilon_t)\,\mathrm{d}B_t + b_\epsilon(t, X^\epsilon_t)\,\mathrm{d}t$$

$$\epsilon \searrow 0 \;\Big\downarrow\; \mathrm{O}(K_q(f)\epsilon^\gamma)$$

$$\mathrm{O}(n^{-\delta+\frac{\delta\beta}{\gamma+\beta}}) \text{ with } \epsilon = \mathrm{O}(n^{-\delta/(\gamma+\beta)}) \text{ as } n \nearrow \infty \dashrightarrow \mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t$$

Figure 1: Computation of the weak rate of convergence.

The key point of our analysis in order to measure γ is to compute the asymptotic distance between b and the class $\mathfrak{M}$. For this, using an equivalent of the perturbation formula for semi-groups in stochastic terms, we will get an inequality of type

$$d_f(X(b), X(b_\epsilon)) \leq C(p,q)K_q(f)d_p(b, b_\epsilon) \tag{3}$$

$$\text{with } d_p(b, b_\epsilon) := \mathbb{E}\left[\int_0^T |b(s, X_s(c)) - b_\epsilon(s, X_s(c))|^p \,\mathrm{d}s\right]^{1/p}$$

where $c = b$ or $c = 0$ (no drift) depending on the context. Here p and q are two positive real numbers such that they are *almost conjugate* in the sense that $p^{-1} + q^{-1} < 1$.

When p is fixed, $d_p(b, b_\epsilon)$ does not depend on f. However, the choice of p depends strongly on the choice of q and then of $\mathfrak{F}$. The smaller the value of p, the smaller the class of functions f that can be considered. For example, if f has a bounded derivative, then one may take $p > 1$ and $c = b$ in (3). If f is only continuous, one may only take $p > 2$ but $c = 0$ in (3). Using Gaussian or Krylov estimates when possible, one may compare $d_p(b, b_\epsilon)$ with the norm $b - b_\epsilon$ in some $\mathrm{L}^{r,q}$-space for a right choice of r and q. This distance is naturally related with a vast literature on approximations of functions. The rate of convergence of b_ϵ to b in the appropriate norm is a delicate issue with no unique answer as we explain in Section 6.

A similar situation appears when proving the rate of convergence of the Euler-Maruyama scheme with a regular enough drift. There is a strong interplay between the constant $\widehat{K}(f)$ and the parameters β and δ which is difficult to quantify as this depends strongly on the method of proof. That is, to find arguments where $\widehat{K}(f)$ is finite within a class $\mathfrak{F}$ and β is as small as possible and δ is as big as possible. There is no unique answer to this problem but we give some possible proofs here.

Putting these problems together we finally get results as in the table in Section 10. As we can already observe some of these results improve the ones in Theorem 1. Still one may wonder about the optimality of our results.

In trying to give a hint on this problem and the complex interplay of the parameters in the figure above, we also consider the particular case when σ is constant. Then we perform an analysis to study the distance between the Euler-Maruyama scheme $\overline{X}$ and the Euler-Maruyama scheme $\overline{X}^\epsilon$. This gives a bound on the weak rate of convergence of the Euler-Maruyama scheme even with a discontinuous coefficient. This is possible because $\overline{X}$, $\overline{X}^\epsilon$, X and X^ϵ can be reduced to a Wiener process using a Girsanov transformation. Still, the facts the change of measure involves a non-smooth integrand has to be dealt within the proofs. In many of these studies we find rates that improve the ones given in Table 1. Therefore pointing to a lack of optimality in our results. Indeed, in Section 9.4, 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.

As we use an approximation of the drift to establish the rate of convergence, one cannot say that the rate we give is optimal. Still the results presented here give estimates of rates of convergence that may help to design simulation methods for particular problems.

It also points towards the need of new techniques to deal with 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 and 7 provide us with estimates on $d_p(b, b_\epsilon)$ and $K_q(f)$. Section 8 provides a weak rates of convergence when the drift coefficient is smooth. Section 9 is devoted to extending our results when the diffusivity σ is constant. In Section 9.4, 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 9.5.

2 Notations, spaces and norms

Vectors in $\mathbb{R}^d$ are usually considered as column vectors. For a vector or matrix A, A^* denotes the transpose of A.

For $k,\ d \in \mathbb{N}$, we denote by $\mathcal{C}_{\mathrm{b}}^k(A)$ the space of real valued, bounded, continuous functions on an open set $A \subseteq \mathbb{R}^d$, with continuous derivatives up to order k which are also bounded. We denote the high order derivatives of $f \in \mathcal{C}_{\mathrm{b}}^k(A)$ for any multi-index $\alpha = (\alpha_1, \dots, \alpha_j) \in \{1, \dots, d\}^j$ of length $j \leq k$ as follows

$$\partial_\alpha f(x) = \frac{\partial^j}{\partial x_{\alpha_j} \cdots \partial x_{\alpha_1}} f(x).$$

Let $f = (f^1, ..., f^d) \in \mathcal{C}_{\mathrm{b}}^k(A)$, and we define the following semi-norm

$$\|\partial_x^k f\|_\infty := \max_{i=1,\dots,d,\ \alpha \in \{1,\dots,d\}^i} \sup_{x \in A} |\partial_\alpha f^i(x)|.$$

As it usual the norm for $f \in \mathcal{C}_{\mathrm{b}}^k(A)$ is defined as $\|f\|_{b,k} = \sum_{j=0}^k \|\partial_x^j f\|_\infty$. The extension of the above spaces to non-open sets A is taken as usual by extending the definition of continuity and differentiability to the boundary points through appropriate limits using elements of A.

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

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

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

We say that a 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 $\mathcal{C}_e(\mathbb{R}^d)$.

We denote by $\mathcal{C}_a^k(\mathbb{R}^d)$ 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 belong to the space $\mathcal{C}_a(\mathbb{R}^d)$.

Two particular domains will appear frequently. Let $H = [0, T) \times \mathbb{R}^d$ and $\overline{H} = [0, T] \times \mathbb{R}^d$ for $T > 0$. 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 $a = Sl$, we have that $\lim_{|x| \to \infty} \sup_{s \in [0,T)} |u(s, x)| e^{-c|x|^2} = 0$ for every $c > 0$ and $s \in [0, T)$. The extension of the norms from the spaces $\mathcal{C}_b^k(\mathbb{R}^d)$ to the spaces $\mathcal{C}_b^k(\overline{H})$ is done by taking supremums with respect to time. That is, $\|f\|_{b,k} := \sup_{t \in [0,T)} \|f(t, \cdot)\|_{b,k}$.

Now, we define in a similar way, Hölder type spaces of order $\alpha > 0$. Denote by $\mathrm{H}^\alpha(\mathbb{R}^d)$ the space of continuous, bounded functions with continuous, bounded derivatives up to order $\lfloor \alpha \rfloor$ and such that $\partial_x^{\lfloor \alpha \rfloor} f$ is $(\alpha - \lfloor \alpha \rfloor)$-Hölder continuous.

Let $\mathrm{H}^{\alpha/2,\alpha}(\overline{H})$ be the set of continuous functions with continuous derivatives $\partial_t^r \partial_x^s u$ for all $2r + s \leq \alpha$ and such that its corresponding norm

$$\begin{aligned} \|u\|_{\mathrm{H}^{\alpha/2,\alpha}} = \|u\|_{b,\lfloor \alpha \rfloor} &+ \sum_{2r+s=\lfloor \alpha \rfloor} \sup_{(t,x),\ (t,y) \in \overline{H},\ x \neq y} \frac{|\partial_t^r \partial_x^s u(t, x) - \partial_t^r \partial_x^s u(t, y)|}{|x - y|^{\alpha - \lfloor \alpha \rfloor}} \\ &+ \sum_{0 < \alpha - 2r - s < 2} \sup_{(t,x),\ (v,x) \overline{H},\ t \neq v} \frac{|\partial_t^r \partial_x^s u(t, x) - \partial_t^r \partial_x^s u(v, x)|}{|t - v|^{(\alpha - 2r - s)/2}} \end{aligned}$$

is finite.

The space $\mathrm{W}^{1,2}_p(H)$ is 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 [25, 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 $\|f\|_{\mathrm{L}^{q,q}(H)} = \|f\|_{\mathrm{L}^q(H)}$.

In general, we will not write the domains over which space integrals are evaluated unless there is the possibility of confusion. For a $d \times k$-matrix m, we define the matrix norm as $\|m\| = \left(\sum_{i,j=1}^{d,k} m_{i,j}^2 \right)^{1/2}$.

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 maybe denoted by the same symbol C or K and unless stated explicitly they may change values from one line to the next.

3 Hypotheses and estimates on an associated PDE

Let $\sigma : \overline{H} \to \mathbb{R}^{d\times d}$ be a measurable function with values in the space of symmetric $d \times d$-matrices. We define $a = \sigma\sigma^*$ and we assume that

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

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

Remark 1. Hypothesis (H1) gives a lower and upper bound on the eigenvalues of a. Therefore as σ is symmetric, the eigenvalues of σ are the square root of the corresponding eigenvalues of a and then (H1) holds with λ and Λ replaced by $\sqrt{\lambda}$ and $\sqrt{\Lambda}$.

Let us also consider a measurable function $b : \overline{H} \to \mathbb{R}^d$ such that

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

These conditions (H1)-(H3) are sufficient to ensure the existence of a unique weak solution in some probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ to the SDE

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

Here B is a d-dimensional Wiener process defined in the probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

Remark 2. As shown in [35], if (4) has a strong solution for $b \equiv 0$, then (4) also admits a strong solution.

Let L be the differential operator

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

The following results give properties of the PDE associated to (1) under weak conditions on the coefficients.

Theorem 2 (Theorem 3' and Corollary p. 401, [35]). *Let L be as above with σ and b satisfying* (H1)-(H3). *Fix $T > 0$ and assume that $f \in \mathcal{C}_{Sl}(\mathbb{R}^d)$. The Cauchy problem*

$$\begin{cases} \dfrac{\partial u(t,x)}{\partial t} + Lu(t,x) = 0, \\ u(T,x) = f(x) \end{cases} \tag{5}$$

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

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

and

$$\mathbb{E}\left[\int_0^T |\nabla u(s,X_s)\sigma(s,X_s)|^2\,\mathrm{d}s\right] = \mathbb{E}[f(X_T)^2] - \mathbb{E}[f(X_T)]^2 = \mathrm{Var}(f(X_T)). \tag{6}$$

Hypotheses (H1)-(H3) are assumed throughout the rest of the article without any further mentioning.

Similarly, in the particular case that $b \equiv 0$ in (5), we will denote by v, the unique solution in $\mathrm{W}^{1,2}_{p,\mathrm{loc}}(H) \cap \mathcal{C}_{Sl}(\overline{H})$ for $p \geq d+1$ to

$$\begin{cases} \dfrac{\partial v(t,x)}{\partial t} + \dfrac{1}{2}\displaystyle\sum_{i,j=1}^d a_{i,j}(t,x)\frac{\partial^2 v}{\partial x_i \partial x_j}(t,x) = 0, \\ v(T,x) = f(x). \end{cases} \tag{7}$$

Let X be the solution to (4) with general b. Then using Itô's formula,

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

Similarly, (6) is also valid where X is the process defined by (4) with $b \equiv 0$. In order to avoid further confusions we denote by Y the weak solution of the equation

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

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

4 The framework to obtain the weak rate of convergence

Before starting a general discussion with specific cases of regularity spaces for test functions and coefficients of stochastic differential equations, we prefer to discuss the general framework that will appear in the different cases so that the reader may easily follow the line of discussion. That is, we give a specific discussion of Figure 1.

Let $X \equiv X(b)$ be the random variable which we want to approximate. This random variable depends on a generally irregular function b. A first approximation is obtained by replacing the irregular function b by a "regularized" function $b_\epsilon \in \mathfrak{M}$ where $\mathfrak{M}$ is a class of functions with $b \notin \mathfrak{M}$ and $\lim_{\epsilon\to 0} d_p(b_\epsilon, b) = 0$ where d_p is some distance function in a space bigger than $\mathfrak{M}$. Using b_ϵ we construct a first approximation $X^\epsilon := X(b_\epsilon)$.

Furthermore $\overline{X}(b_\epsilon) \equiv \overline{X}(b_\epsilon, n)$ denotes an approximation of $X(b_\epsilon)$ where n is the parameter used in the construction of $\overline{X}$. In the particular case of the Euler scheme n denotes the number of steps.

The approximation quality of $X(b_\epsilon)$ and $\overline{X}(b_\epsilon)$ are measured through the class of test functions $\mathfrak{F}$. For this, we use a seminorm denoted by $d_f(\cdot,\cdot)$ for fixed $f \in \mathfrak{F}$. Then the distances that measure the quality of the approximations $X(b_\epsilon)$ and $\overline{X}(b_\epsilon)$ are $d_f(X(b), X(b_\epsilon))$ and $d_f(X(b_\epsilon), \overline{X}(b_\epsilon))$ respectively.

The general situation which will appear repeatedly is as follows.

Theorem 3. *Assume that there exist strictly positive constants q, γ, β and δ such that for every $f \in \mathfrak{F}$ and for every sequence $\{b_\epsilon, \epsilon \in (0,1)\} \subseteq \mathfrak{M}$ the following inequalities are satisfied:*

$$\begin{aligned} d_f(X(b), X(b_\epsilon)) &\leq K_q(f) d_p(b, b_\epsilon) \ \textit{with}\ d_p(b, b_\epsilon) = \mathrm{O}(\epsilon^\gamma) \\ \textit{and}\ d_f(X(b_\epsilon), \overline{X}(b_\epsilon)) &\leq \widehat{K}(f)\epsilon^{-\beta} n^{-\delta}, \end{aligned}$$

where $p > q/(q-1)$ and $K_q(f)$ and $\widehat{K}(f)$ are positive constants depending on f. Then the choice $\epsilon = O(n^{-\delta/(\gamma+\beta)})$ gives the optimal rate which is

$$d_f(X(b), \overline{X}(b_\epsilon)) = O(n^{-\delta\gamma/(\gamma+\beta)}).$$

The proof of the above statement is straightforward. Now we give more details of the precise framework of stochastic differential equations that we will be using from now on.

Let us fix $T > 0$, the time horizon. Let $b_\epsilon \in \mathfrak{M}$ be a family of measurable coefficients on $\overline{H}$ with $|b_\epsilon(t,x)| \leq \Lambda$ for $(t,x) \in \overline{H}$. Let us suppose that for any $b_\epsilon \in \mathfrak{M}$ there is a unique weak solution X^ϵ on $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ to

$$X_t^\epsilon = x + \int_0^t \sigma(s, X_s^\epsilon)\,\mathrm{d}B_s + \int_0^t b_\epsilon(s, X_s^\epsilon)\,\mathrm{d}s. \tag{8}$$

Since b_ϵ and b are bounded, the distribution of X^ϵ may be deduced from the distribution of X through a Girsanov transform.

In fact, this idea will appear recurrently throughout the paper. Let $\Gamma : \Omega \times [0,T] \to \mathbb{R}^d$, be a bounded, predictable function. Then $Z \equiv Z(\Gamma)$ denotes the exponential martingale in the direction Γ as

$$Z_t \equiv Z_t(\Gamma) = \exp\left(\int_0^t \Gamma(s, X_s)\,\mathrm{d}B_s - \frac{1}{2}\int_0^t \Gamma(s, X_s)^*\Gamma(s, X_s)\,\mathrm{d}s\right). \tag{9}$$

This martingale is the unique strong solution to the SDE $Z_t = 1 + \int_0^t Z_s\Gamma(s, X_s)\,\mathrm{d}B_s$.

The martingale Z 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 $\left.\frac{d\mathbb{Q}^\Gamma}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t;\ t \geq 0$. Various moment properties of Z are given in Appendix 11.1.

We let $\mathrm{d}W_t^\Gamma = dB_t - \Gamma(t, X_t)\,\mathrm{d}t$ denote the Wiener process on the probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q}^\Gamma)$. As it is well known, choosing the right definition for Γ and using Girsanov theorem, the equations (8) and (4) 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 will not make any distinction between them as we will only be considering their expectations.

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

With this introduction, we now give the values of the random variables to be used in Theorem 3.

We consider $d_f(X(b), X(b_\epsilon))$, where $X(b) := X_T$ and $X(b_\epsilon) = X_T^\epsilon$. As we will show below in Section 5, under rather general hypotheses,

$$d_f(X(b), X(b_\epsilon)) := |\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \le C(p,q)K_q(f)d_p(b, b_\epsilon) \tag{10}$$

for almost conjugate values of p and q, a constant $C(p,q)$ which explodes as $p^{-1}+q^{-1} \to 1$ and a suitable choice of $d_p(b, b_\epsilon)$ that does not depend on f once p is fixed (the choice of p depends on q which depends on the class $\mathfrak{F}$ of test functions). Assume that $d_p(b, b_\epsilon)$ decreases to 0 at the rate $\mathrm{O}(\epsilon^\gamma)$.

Let $\overline{X}^\epsilon$ be the continuous solution of the Euler-Maruyama scheme of step size T/n. That is, define for fixed $n \in \mathbb{N}$, $\phi(s) \equiv \phi_n(s) = \sup\{t \le s \,|\, t = k/n \text{ for } k \in \mathbb{N}\}$, then

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

When σ and b_ϵ belong to a "good" class of functions $\mathfrak{M}$ (for example $\mathfrak{M} = \mathrm{H}^{\alpha/2,\alpha}(\overline{H})$ for some $\alpha > 0$ or $\mathfrak{M} = \mathcal{C}_\mathrm{b}^{1,3}(\overline{H})$), and when f belongs to a proper class of functions $\mathfrak{F}$ (for example, $\mathfrak{F} = \mathrm{H}^{2+\alpha}(\mathbb{R}^d)$ or $\mathfrak{F} = \mathcal{C}^3(\mathbb{R}^d) \cap \mathcal{C}_{Sl}(\mathbb{R}^d)$ respectively), a weak rate of convergence of the Euler-Maruyama scheme $\overline{X}^\epsilon$ is known. This means that there exist $\delta > 0$ and some constant C_ϵ such that

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

Assume that $C_\epsilon = \mathrm{O}(\epsilon^{-\beta})$. 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 $\mathfrak{M}$ and $\mathfrak{F}$.

Optimizing over the choice of ϵ leads to the following corollary of Theorem 3, which we summarized in Figure 1.

Corollary 1. *Assume that for $f \in \mathfrak{F}$, $b_\epsilon \in \mathfrak{M}$, there exists positive constants γ, β and δ which only depend on the definition of the classes $\mathfrak{F}$ and $\mathfrak{M}$ such that*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| = \mathrm{O}(\epsilon^\gamma K_q(f)) \text{ and } |\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(\overline{X}_T^\epsilon)]| = \mathrm{O}\left(\epsilon^{-\beta} n^{-\delta} \widehat{K}(f)\right).$$

Then for $\epsilon = \mathrm{O}(n^{-\delta/(\gamma+\beta)})$, both errors above are of the same order $O(n^{-\kappa})$ with $\kappa = \delta - \frac{\delta\beta}{\gamma+\beta}$ and therefore

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

In the above result one clearly sees that, given (10), γ measures the irregularity of b with respect to the class $\mathfrak{M}$ and the law of Y. The parameter δ measures the weak error within the class $\mathfrak{M}$, for $f \in \mathfrak{F}$ and 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 $K_q(f)$ and $\widehat{K}(f)$ will determine the method of proof to be used in order to obtain the respective rates.

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

Our aim is now to give some general conditions for which the key inequality (10) holds.

5.1 An upper bound for two different drifts

For $p > 0$ and a measurable function $g : \overline{H} \to \mathbb{R}^k$, we say that $g \in \mathrm{L}^p(X)$ if its associated norm

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

is finite. The above norm is always considered for X defined as the weak solution of (4) on $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. We somewhat abuse the notation, using the convention $\|g\|_{X,\infty} := \|g\|_\infty$. In a similar fashion, we can define $\|g\|_{Y,p}$.

In this section, we give a perturbation formula which compares the laws of two diffusions with different drifts. This formula involves the gradient of the solution to (5) with $b = 0$.

In what follows, b_ϵ is a function approximating b such that it also satisfies hypothesis (H3).

Proposition 1. *Let us assume that $f \in \mathcal{C}_{Sl}(\mathbb{R}^d)$ such that for some $q > 1$, $\|\nabla v\|_{Y,q} < +\infty$, where v is the solution to (7) with terminal condition f. Then there exists a constant $C_2(p, q)$ depending only on T, Λ, λ, p and q such that for any p, q with*

$$2 \leq p < +\infty,\ 1 < q \leq \infty \text{ and } p^{-1} + q^{-1} < 1,$$

it holds that

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

with $C_2(p, q) \xrightarrow[p^{-1}+q^{-1}\to 1]{} +\infty$.

As mentioned in Section 4, recall that $d_f(X(b), X(b_\epsilon)) := |\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]|$ and $d_p(b, b_\epsilon) := \|b - b_\epsilon\|_{Y,p}$ with $K_q(f) := \|\nabla v\|_{Y,q}$.

Proof. Let us define $Z = Z(\Gamma_1)$ with $\Gamma_1(s, x) := (b_\epsilon(s, x) - b(s, x))^* \sigma^{-1}(s, x)$. Note that $\hat{\Gamma}_1 := \sup_{(s,x)\in[0,T]\times\mathbb{R}^d} |\Gamma_1(s, x)| \leq 2\Lambda\lambda^{-1}$. Using Girsanov's theorem and then Itô's formula together with the estimates of Theorem 2 and Lemma 7 in Appendix,

$$\begin{aligned}
&\mathbb{E}[f(X_T^\epsilon)] - \mathbb{E}[f(X_T)] = \mathbb{E}[(Z_T - 1)f(X_T)]\\
&\quad = \mathbb{E}\left[\int_0^T Z_s(b_\epsilon - b)^*(s, X_s)\nabla v(s, X_s)\,\mathrm{d}s\right] + \mathbb{E}\left[(Z_T - 1)\int_0^T b^*(s, X_s)\nabla v(s, X_s)\,\mathrm{d}s\right].
\end{aligned}$$

For the first term, we use the Hölder inequality for some $p > \overline{p}$, $q > \overline{q}$ and $\overline{p}^{-1} + \overline{q}^{-1} < 1$ that

$$\begin{aligned}
\left|\mathbb{E}\left[\int_0^T Z_s(b_\epsilon - b)^*(s, X_s)\nabla v(s, X_s)\,\mathrm{d}s\right]\right| &\leq C(\overline{p}, \overline{q})\|b - b_\epsilon\|_{X,\overline{p}}\mathbb{E}\left[\int_0^T |\nabla v(s, X_s)|^{\overline{q}}\,\mathrm{d}s\right]^{1/\overline{q}}\\
&\leq C(\overline{p}, \overline{q})\varkappa(\hat{\Gamma}_1, \overline{p}, p)\|b - b_\epsilon\|_{Y,p}\varkappa(\hat{\Gamma}_1, \overline{q}, q)\|\nabla v\|_{Y,q}
\end{aligned}$$

where Lemma 7(II) has been used for the final step. Similarly, for the second term, Hölder's inequality yields using Lemma 8 that

$$\left|\mathbb{E}\left[(Z_T-1)\int_0^T b^*(s,X_s)\nabla v(s,X_s)\,\mathrm{d}s\right]\right| \leq \mathbb{E}[|Z_T-1|^{\overline{p}}]^{1/\overline{p}}\Lambda T^{1-1/\overline{q}}\mathbb{E}\left[\int_0^T |\nabla v(s,X_s)|^{\overline{q}}\,\mathrm{d}s\right]^{1/\overline{q}}$$
$$\leq C\varkappa(\hat{\Gamma}_1,\overline{p},p)\|b-b_\epsilon\|_{Y,p}\varkappa(\hat{\Gamma}_1,\overline{q},q)\|\nabla v\|_{Y,q},$$

where C is some positive constant which depends on p, q, T, Λ and λ. Hence the result. □

Although Proposition 1 seems to be the most natural as any consideration of the irregular drift has been taken away in the norm $\|\nabla v\|_{Y,q}$, the restriction $p \geq 2$ is a determinant problem in fixing the weak rate of convergence. This restriction appears due to the use of the BDG inequality in Lemma 8 in Appendix 11.1. In the next proposition, we take another option which retains the irregular drift and therefore the estimates on $\|\nabla u\|_{X,q}$ will be more difficult to obtain but we can use the resulting inequality for any $p > 1$.

Proposition 2. *Let us assume that $f \in \mathcal{C}_{Sl}(\mathbb{R}^d)$ is such that for some $1 < q \leq \infty$, $\|\nabla u\|_{X,q} < +\infty$, where u is the solution to (5) with terminal condition f. Then there exists a constant $C_1(p,q)$ depending only on T, Λ, λ, p and q such that for any p, q with*

$$1 < p < +\infty,\ 1 < q \leq \infty \text{ and } p^{-1}+q^{-1} < 1,$$

it satisfies that

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq C_1(p,q)\|b-b_\epsilon\|_{Y,p}\|\nabla u\|_{X,q}$$

with $C_1(p,q) \xrightarrow[p^{-1}+q^{-1}\to 1]{} +\infty$.

Proof. Set $\Gamma_2(s,x) = -b(s,x)^*\sigma^{-1}(s,x)$ and $\hat{\Gamma}_2 := \sup_{(s,x)\in[0,T]\times\mathbb{R}^d}|\Gamma_2(s,x)|$. From Girsanov's theorem,

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

As in the proof of Proposition 1, we have, for α_1, α_2 and q with $\alpha_1^{-1}+\alpha_2^{-1}+q^{-1}=1$,

$$|\Delta| \leq T^{1-\frac{1}{\alpha_1}}\mathbb{E}\left[|Z_T|^{\alpha_1}\right]^{\frac{1}{\alpha_1}}\mathbb{E}\left[\int_0^T |(b_\epsilon-b)(s,X_s)|^{\alpha_2}\,\mathrm{d}s\right]^{\frac{1}{\alpha_2}}\mathbb{E}\left[\int_0^T |\nabla u(s,X_s)|^q\,\mathrm{d}s\right]^{\frac{1}{q}}.$$

Finally the result follows from using Lemma 7(I) and Lemma 7(II) with $\Gamma_2(s,x)$ in order to obtain that $\|b_\epsilon-b\|_{X,\alpha_2} \leq \varkappa(\hat{\Gamma}_2,\alpha_2,p)\|b_\epsilon-b\|_{Y,p}$ with $p > \alpha_2$. □

6 Rate of convergence of smooth approximations. Considerations for $d(b,b_\epsilon)$

In the previous section we have analyzed the distance $d_f(X(b),X(b_\epsilon))$ and we have found in Propositions 1 and 2, that these distances depend on two factors. The first is the distance between b and b_ϵ in the L^p-norm. The second is a measure on the regularity of f. In this brief section we discuss the former. We note in this section that 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 will in general be smaller. Furthermore, the problem is reduced to a general problem of function approximation as the following reduction shows.

Lemma 1. *Let us assume that an upper Gaussian estimate holds for the transition density function $\mathsf{p}(t,x,y)$ of Y, that is, there exists some positive constants C_1 and C_2 such that*

$$\mathsf{p}(t,x,y) \;\le\; C_1 \mathsf{g}_{C_2 t}(x-y),\ \forall (t,x,y) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d. \tag{11}$$

Then using the above inequality and Hölder's inequality repeatedly, for any $q, r \in [p,\infty]$ with $\frac{d}{2q} + \frac{1}{r} < \frac{1}{q}$,

$$\|b - b_\epsilon\|_{Y,p} \le C_3 \left(\int_0^T \left(\int_{\mathbb{R}^d} |b(s,y) - b_\epsilon(s,y)|^q \,\mathrm{d}y \right)^{r/q} \mathrm{d}s \right)^{1/r} = C_3 \|b - b_\epsilon\|_{\mathrm{L}^{r,q}(H)}. \tag{12}$$

Remark 3. The estimate (11) holds for example if the diffusion coefficient a belongs to the Hölder space $\mathrm{H}^{\alpha/2,\alpha}(H)$ for some $\alpha > 0$ (see e.g. [25, § IV.13, p. 377]).

Even in absence of a Gaussian upper bounds, the *Krylov estimate* ([24] or [7, Theorem 7.6.2, p. 114]) could also be used with Hypothesis (H1) in order to get an estimate on $\|b - b_\epsilon\|_{Y,p}$. In this case of a homogeneous coefficient b, from the Krylov estimate, we have

$$\|b - b_\epsilon\|_{Y,p} \le C(\lambda, \Lambda) e^T \|b - b_\epsilon\|_{\mathrm{L}^{dp}(H)}.$$

In case of a time-inhomogeneous coefficient, a similar estimate could be obtained but on bounded domain and one should then estimate the exit time from such domains.

We now provide a standard example of convergence of regularized coefficients.

Example 1. Here we provide an example of order of ϵ in the case of the indicator function $b(t,x) = \mathbf{1}_{[\zeta_1,\zeta_2]}(x)$ for $x \in \mathbb{R}$ and $\zeta_1 < \zeta_2$. If we use the following b_ϵ for an approximation of b, b_ϵ has the Lipschitz continuity property. In fact, let for $\epsilon > 0$,

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

Then we have the following orders: for $p \ge 1$,

$$\left(\int_{-\infty}^{\infty} |b_\epsilon(x) - b(x)|^p \,\mathrm{d}x \right)^{\frac{1}{p}} = \left(\frac{4\epsilon}{p+1} \right)^{\frac{1}{p}} = O\left(\epsilon^{\frac{1}{p}} \right). \tag{13}$$

Similarly, if we use a mollifier with the Gaussian kernel, that is $b_\epsilon(x) := \int_{-\infty}^{+\infty} b(u) \mathsf{g}_\epsilon(x - u)\,\mathrm{d}u$, then we have the same order of the convergence as the above (13) with $b_\epsilon \in \mathcal{C}_b^\infty(\mathbb{R})$ and $\|b_\epsilon\|_{\mathrm{H}^\alpha(\mathbb{R})} \le C\epsilon^\alpha$.

The above example is part of a large literature 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. For this reason we will make the following assumption for the rest of the article.

$$\text{There exists } \gamma > 0 \text{ such that } \|b - b_\epsilon\|_{Y,p} \le C\epsilon^\gamma. \tag{H4}$$

For more references on this matter, see [1, Eq. (84)], [12], [13] or [34] for the relation of this problem with Besov spaces.

7 Finiteness of $K_q(f)$

We treat here with two general situations, where $q = 2$ and $q = +\infty$. Using the notion of *fractional smoothness* introduced in [14–16] it is possible to consider several classes $\mathfrak{F}$ of functions for which $K_q(f)$ is finite for various values of q. These classes may include even discontinuous functions. This will be subject to future works.

We recall that hypotheses (H1), (H2) and (H3) are assumed. The next corollary immediately follows from Proposition 2, Theorem 2, Hypothesis (H2), with (H1) together with (6) applied to $\|\nabla v\|_{X,q}$.

Corollary 2. *For $f \in \mathcal{C}_{Sl}(\mathbb{R}^d)$, $K_q(f) = \|\nabla v\|_{Y,q} \leq \lambda^{-1/2}\sqrt{\operatorname{Var} f(Y_T)}$. Thus, for $p > 2$,*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^\epsilon)]| \leq \lambda^{-1/2} C_1(p,2)\|b - b_\epsilon\|_{Y,p}\sqrt{\operatorname{Var} f(Y_T)}.$$

There is a trade-off between the value of p and the constants in Propositions 1 and 2. When considering only the order, it is suitable to take p as small as possible. In Proposition 1, we have that $p > 2 \vee \frac{q}{q-1}$. For $q > 2$, this gives that anyway $p > 2$. However, with Proposition 2, the only constraint is that $p > \frac{q}{q-1}$. We now give a condition for which one may take $q = +\infty$ so that p may be taken as close to 1 as possible.

Proposition 3. *Assume that $d = 1$ and that $f \in \mathcal{C}_{Sl}(\mathbb{R}) \cap \mathcal{C}^1(\mathbb{R})$ with a bounded derivative. Then we have*

$$\|\nabla u\|_{X,\infty} \leq C\|\nabla f\|_\infty, \tag{14}$$

where C depends only on T, λ and Λ.

When $d > 1$ and $\sigma = \operatorname{Id}$, (14) *also holds for $f \in \mathcal{C}_{Sl}(\mathbb{R}^d) \cap \mathcal{C}^1(\mathbb{R}^d)$ with bounded derivatives.*

Proof. The proof is similar to the one of Proposition 2 but we use the fact that for u solution to (5), ∇u is globally bounded to get the result. We now focus on this last point.

Let us consider a smooth family of approximations a^n, b^n and f^n of a, b and f, where a^n and b^n satisfy Hypotheses (H1), (H2) and (H3) and $|f^n(0)| \leq |f(0)|$ as well as $\|\nabla f^n\|_\infty \leq \|\nabla f\|_\infty$. In fact, first define an approximation $\hat{f}_n$ of f constructed by smooth convolution. Then, define $f_n = \hat{f}_n - \hat{f}_n(0) + f(0)$. This will satisfy the required properties.

The solution u^n to (5) with (a, b, f) replaced by (a^n, b^n, f^n) is a classical solution to (5) on H and is smooth on $\overline{H}$.

∘ *Case $d = 1$.* The function $v^n(t,x) = \nabla u^n(t,x)$ is then solution to

$$\begin{cases} \dfrac{\partial v^n(t,x)}{\partial t} + \dfrac{1}{2}\dfrac{\partial}{\partial x}\left(a^n(t,x)\dfrac{\partial}{\partial x}\right) v^n(t,x) + \dfrac{\partial}{\partial x}\Big(b^n(t,x)v^n(t,x)\Big) = 0, \\ v^n(T,x) = \nabla f^n(x). \end{cases} \tag{15}$$

With the Aronson estimates, the fundamental solution $\mathsf{q}^n(s,t,x,y)$ of $\partial_t + \frac{1}{2}\partial_x(a^n\partial_x) + \partial_x(b^n\cdot)$ satisfies a Gaussian inequality of type

$$\mathsf{q}^n(s,t,x,y) \leq C_1 \mathsf{g}_{C_2(t-s)}(x-y) \text{ on } [0,T]^2 \times \mathbb{R}^2, \tag{16}$$

where C_1 and C_2 depend only on the constants λ and Λ [3,33]. In addition,

$$v^n(t,x) = \int_{\mathbb{R}} \mathsf{q}^n(t,T,x,y)\nabla f^n(y)\,\mathrm{d}y.$$

In particular, (16) implies that by

$$|v^n(t,x)| \leq C_3 \|\nabla f^n\|_\infty, \tag{17}$$

where C_3 depends only on λ and Λ.

Let v be the weak solution to (15) with $(a^n, b^n, \nabla f^n)$ replaced by $(a, b, \nabla f)$, for which a locally Hölder continuous version exists on $[0,T)\times\mathbb{R}$ (See *e.g.* [33, Theorem II.2.12, p. 340]) It is standard that $v^n(t,x)$ converges locally uniformly to $v(t,x)$ on H (See *e.g.* [33, Theorem II.3.8, p. 344]).

It remains to identify v with ∇u.

From Theorem 2, $u(t,x) = \mathbb{E}_{t,x}[f(X_T)]$, where $\mathbb{P}_{t,x}$ is the distribution of $(X_s)_{s\geq t}$ with $X_t = x$ almost surely. It is easily established that

$$\mathbb{E}_{t,x}[|X_T|^2] \leq C(1+x^2), \tag{18}$$

where C depends only on Λ and T. Hence,

$$|u(t,x)| \leq |f(0)| + \|\nabla f\|_\infty \sqrt{C_4(1+x^2)}. \tag{19}$$

Let X^n be the process generated by $L^n = \frac{1}{2}\sum_{i,j} a^n_{i,j}\partial^2_{x_i x_j} + \sum_i b^n_i \partial_{x_i}$. Using the results in [31], X^n converges in distribution to X in the space of continuous functions. As (18) is also valid for X^n, it is easily obtained that $u^n(t,x) = \mathbb{E}_{t,x}[f^n(X^n_T)]$ is solution to $(\partial_t + L^n)u^n = 0$ with $u^n(T,x) = f^n(x)$. Moreover, $u^n(t,x)$ converges to $u(t,x) = \mathbb{E}_{t,x}[f(X_T)]$ for any $(t,x)\in H$.

Let D and D' be two open domains such that $\overline{D}\subset D'$ and $\overline{D'}\subset H$. Set $\mathcal{L}^n = \partial_t + L^n$ and ζ be a smooth function equal to 1 on D and 0 outside D'. Then

$$\mathcal{L}^n(u^n\zeta) = \zeta\mathcal{L}^n u^n + u^n\mathcal{L}^n\zeta + \langle a\nabla u^n, \nabla\zeta\rangle = u^n\mathcal{L}^n\zeta + \langle a^n\nabla u^n, \nabla\zeta\rangle.$$

Since (19) is valid for u^n and ∇u^n also satisfies (17), the function $u^n\mathcal{L}^n\zeta + \langle a^n\nabla u^n, \nabla\zeta\rangle$ with support in $\overline{D}'$ is uniformly bounded in n. It follows from Theorem IV.9.1 in [25, p. 341], $u^n\zeta$ is uniformly bounded in $\mathrm{W}^{1,2}_p(D')$ for any $p\geq 1$. With the Corollary of Theorem IV.9.1, p. 341, we also recover the fact that ∇u^n is uniformly locally Hölder continuous.

For a smooth function ϕ on H with a compact support,

$$\int_0^T\int_{\mathbb{R}} \nabla\phi(t,x)u^n(t,x)\,\mathrm{d}x\,\mathrm{d}t = -\int_0^T\int_{\mathbb{R}}\phi(t,x)\nabla u^n(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

Passing to the limit, as u^n converges to u and ∇u^n converges to v, this shows that $\nabla u(t,x) = v(t,x)$ on H and then that $|v(t,x)|\leq C_3\|\nabla f\|_\infty$ for any $(t,x)\in H$.

∘ *Case* $d>1$, $\sigma = \mathrm{Id}$. Let b^n and f^n as above and let us assume moreover that f^n is bounded with bounded derivatives up to order 4. From Theorem IV.5.1 in [25, p. 320], u^n and its first (in space and time) and second derivatives (in space) are bounded and Hölder continuous. In particular, ∇u^n is globally bounded.

We now show that $\|\nabla u^n\|_\infty \leq C\|\nabla f\|_\infty$ for a constant C that depends only on Λ, T and d.

Set $v^n_i = \partial_{x_i}u^n$. For $j=1,\dots,d$,

$$\frac{\partial v^n_j(t,x)}{\partial t} + \frac{1}{2}\sum_{i=1}^d \frac{\partial^2 v^n_j(t,x)}{\partial x_i^2} + \sum_{i=1}^d \frac{\partial}{\partial x_j}(b^n_i(t,x)v^n_i(t,x)) = 0 \tag{20}$$

with $v_j^n(T,x) = \partial_{x_j} f^n(x)$. Using the results in [3, Theorem 5, p. 656], for a Brownian motion B,

$$v_j^n(t,x) = \mathbb{E}[\partial_{x_j} f^n(x+B_{T-t})] + \sum_{i=1}^d \int_t^T \int_{\mathbb{R}^d} \mathsf{G}_j(s-t,x,y) b_i^n(s,y) v_i^n(s,y) \,\mathrm{d}y \,\mathrm{d}s \qquad (21)$$

$$\text{with } \mathsf{G}_j(t,x,y) = \partial_{x_j} \mathsf{g}_t(y-x) = \frac{-(x_j-y_j)}{(2\pi)^{d/2} t^{(d+2)/2}} \exp\left(\frac{-|x-y|^2}{2t}\right).$$

Hence,

$$\sup_{s\in[t,T]} \sup_{x\in\mathbb{R}^d} \sup_{j=1,\dots,d} |v_j^n(s,x)| \leq \|\nabla f\|_\infty + C\sqrt{T-t} \sup_{s\in[t,T]} \sup_{x\in\mathbb{R}^d} \sup_{j=1,\dots,d} |v_j^n(s,x)|,$$

where C is a constant that depends only on Λ and d. For $T-t \leq \theta$ with $C\sqrt{\theta} \leq 1/2$,

$$\sup_{s\in[t,T]} \sup_{x\in\mathbb{R}^d} \sup_{j=1,\dots,d} |v_j^n(s,x)| \leq 2\|\nabla f\|_\infty.$$

Since v_j^n is solution to (20) on $[T-2\theta, T-\theta]$ with $v_j^n(T-\theta,x)$ which is continuous and bounded $2\|\nabla f\|_\infty$, one deduces iteratively that

$$\sup_{j=1,\dots,d} \|v_j^n\|_\infty \leq C'\|\nabla f\|_\infty \qquad (22)$$

for some constant C' that depends only on d, T and Λ.

Using the representation (21), it is easily seen that v_j^n is locally Hölder continuous with a modulus of continuity which is uniform in n. The Ascoli theorem asserts that $(v_j^n)_{n\in\mathbb{N}}$ has a convergent subsequence. Using the integrability of Γ_j, the derivative of the Gaussian kernel, one gets easily that any possible limit of $(v_j^n)_{n\in\mathbb{N}}$ satisfies (21) with b^n replaced by b and $\partial_{x_j} f^n$ replaced by $\partial_{x_j} f$. Hence, for $j=1,\dots,d$, $(v_j^n)_{n\in\mathbb{N}}$ is globally convergent to a single limit v_j which satisfies (22).

The remainder of the proof is as above and allows one to identify $\partial_{x_j} u$ with v_j. □

8 Consideration for $d(X^\epsilon, \overline{X}^\epsilon)$: Rates of convergence of the Euler-Maruyama scheme with regular coefficients

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

8.1 Case of smooth coefficients

Theorem 1 requires the coefficients to be Hölder continuous. Of course, the convergence rate is better for smooth coefficients. But in order to achieve a rate equal to 1, it requires a to be in $\mathrm{H}^{\alpha/2,\alpha}(\overline{H})$ with $\alpha > 2$ and a terminal conditions in $\mathrm{H}^{2+\alpha}(\mathbb{R}^d)$ and then with a better regularity than $\mathcal{C}_p^4$.

With a bit more regularity on a and b (if we use mollifiers for the approximation, b_ϵ has enough regularity), we will see that we may achieve a convergence rate equal to 1 with a less stringent condition on f.

Assume that $\sigma \in \mathrm{H}^{1/2,1}(\overline{H})$ satisfies (H1). Let B be a Wiener process and denote by $\overline{Y}$ the Euler-Maruyama scheme associated to Y with step size T/n. That is, $\overline{Y}_t = x + \int_0^t \sigma(\phi(s), \overline{Y}_{\phi(s)})\,\mathrm{d}W_s$.

It follows from the results of [26] that $\overline{Y}_T$ has a density $\mathsf{p}_n(T,x,y)$ for which Gaussian estimates holds. Hence, there exists some constants c_1 and c_2 such that for all $n \in \mathbb{N}$,

$$\mathsf{p}_n(T,x,y) \leq c_1 \mathsf{g}_{c_2 T}(x-y). \tag{23}$$

The process Y has also a density $\mathsf{p}(t,x,y)$ satisfying (23) (which follows from (23) after passing to the limit or from [25, (13.1), p. 376]).

Lemma 2. *Let f be a function in $\mathcal{C}_{\mathrm{p}}(\mathbb{R}^d)$ with $|f(x)| \leq C(1+|x|^p)$. Then for any integer k, there exists a function $B_{\mathrm{p}}(f,x)$ such that*

$$|\mathbb{E}[f(\overline{Y}_T)^k]|^{1/k} + |\mathbb{E}[f(Y_T)^k]|^{1/k} \leq B_{\mathrm{p}}(f,x) \text{ with } \sup_{x\in\mathbb{R}^d} \frac{B_p(f,x)}{1+|x|^p} \leq \varkappa < +\infty,$$

where $\varkappa$ depends on C, T, k, p, d and $\|\sigma\|_{(1/2,1)}$.

Let f be a function in $\mathcal{C}_{\mathrm{e}}(\mathbb{R}^d)$ with $|f(x)| \leq C_1 e^{C_2|x|}$. Then for any integer k, there exists a function $B_{\mathrm{e}}(f,x)$ depending on C_1 and C_2 such that

$$|\mathbb{E}[f(\overline{Y}_T)^k]|^{1/k} + |\mathbb{E}[f(Y_T)^k]|^{1/k} \leq B_{\mathrm{e}}(f,x) \text{ with } \sup_{x\in\mathbb{R}^d} B_e(f,x)e^{-C_2|x|} \leq \varkappa < +\infty$$

where $\varkappa$ depends only C_1, C_2, T, k, d and $\|\sigma\|_{(1/2,1)}$.

Proof. These inequalities follows immediately from (23) applied to the densities $\mathsf{p}_n(t,x,y)$ and to $\mathsf{p}(t,x,y)$. □

Theorem 4. *Assume that f in $\mathcal{C}^3_{\mathrm{p}}(\mathbb{R}^d)$, $b_\epsilon \in \mathcal{C}^{1,3}_{\mathrm{b}}(\overline{H})$ and $\sigma \in \mathcal{C}^{1,3}_{\mathrm{b}}(\overline{H})$. Then there exist some positive constants C and C', which do not depend on the function b_ϵ and depend on d, T, f and σ, such that*

$$|\mathbb{E}[f(X^\epsilon_T)] - \mathbb{E}[f(\overline{X}^\epsilon_T)]| \leq \frac{C}{n}\exp\left(C'\|b_\epsilon\|_\infty\right) P_4\left(\|\partial_t b_\epsilon\|_\infty, \|b_\epsilon\|_\infty, \|\partial^1_x b_\epsilon\|_\infty, \|\partial^2_x b_\epsilon\|_\infty, \|\partial^3_x b_\epsilon\|_\infty\right).$$

Here $P_4(x_0,...,x_4)$ is a polynomial of degree 4 such that if any term of the polynomial is denoted by $\prod_{i=0}^4 x_i^{n_i}$ then the powers satisfy the restriction $2n_0 + n_1 + \cdots + n_4 \leq 4$.

Remark 4. This approach, namely using the Malliavin calculus, has a possibility to be extended to non Markovian processes like delayed stochastic differential equations. It also clarifies the need of certain regularity of f, b and σ and that these conditions may be interchanged requiring less regularity of f at the expense of demanding more regularity of b and σ. Readers can find more details in [10].

Proof. As we need to put both the solution process and the approximation process in the same probability space we re-define the processes as follows: Let $\Gamma(s,x) := b^*\sigma^{-1}(s,x)$, $\Gamma^\epsilon(s,x) := b^*_\epsilon\sigma^{-1}(s,x)$ and $\tilde{\Gamma}^\epsilon(s) := b^*_\epsilon\sigma^{-1}(\phi(s), \overline{Y}_{\phi(s)})$. Accordingly, we define the measures $\mathbb{Q}^\Gamma$, $\mathbb{Q}^{\Gamma^\epsilon}$ and $\mathbb{Q}^{\tilde{\Gamma}^\epsilon}$ as in (9). Then under $\mathbb{Q}^\Gamma$ have that $Y \stackrel{\mathcal{L}}{=} X$. Similarly, under the change of measure $\mathbb{Q}^{\Gamma^\epsilon}$, $Y \overset{\mathcal{L}}{=} X^\epsilon$ and under $\mathbb{Q}^{\tilde{\Gamma}^\epsilon}$, we have that $\overline{Y} \overset{\mathcal{L}}{=} \overline{X}^\epsilon$. Therefore we consider the following decomposition:

$$\mathbb{E}\left[f(X_T^\epsilon)\right] - \mathbb{E}\left[f(\overline{X}_T^\epsilon)\right] = \mathbb{E}\left[Z_T(\Gamma^\epsilon)f(Y_T)\right] - \mathbb{E}\left[Z_T(\tilde{\Gamma}^\epsilon)f\left(\overline{Y}_T\right)\right].$$

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

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

and let the system to approximate be $(Y, \log(Z(\Gamma^\epsilon)))$. This system is approximated using $(\overline{Y}, \log(Z_T(\tilde{\Gamma}^\epsilon)))$. The theory developed in [10] allows us to write the error in general. As this methodology has been also applied in other problems, we only briefly sketch the proof here. In the present case we have to be careful with the term involving b_ϵ as its derivatives may blow up as $\epsilon \to 0$. Therefore for $[0,1]$-uniform random variables $\theta, \theta_i', \dot{\theta}$, $i = 1, \cdots, d$, which are independent of W and also independent between themselves,

$$\mathbb{E}\left[g\left(Y_T, Z_T(\Gamma^\epsilon)\right) - g\left(\overline{Y}_T, Z_T(\tilde{\Gamma}^\epsilon)\right)\right] = \mathbb{E}[F^*\mathcal{E}_T + F_{d+1}(A_1 + A_2 + A_3 + A_4)] \tag{24}$$

with

$$\begin{aligned}
A_1 &:= \sum_{i=1}^d \int_0^T \left(\nabla(b_\epsilon^*\sigma^{-1})^i(s, \theta_i' Y_s + (1-\theta_i')\overline{Y}_s)\right)^* \mathcal{E}_s \, \mathrm{d}W_s^i, \\
A_2 &:= \int_0^T \left(b_\epsilon^*\sigma^{-1}(s, \overline{Y}_s) - b_\epsilon^*\sigma^{-1}(\phi(s), \overline{Y}_{\phi(s)})\right) \mathrm{d}W_s \\
A_3 &:= -\frac{1}{2}\int_0^T \left(\nabla(b_\epsilon^* a^{-1} b_\epsilon)(s, \dot{\theta} Y_s + (1-\dot{\theta})\overline{Y}_s)\right)^* \mathcal{E}_s \, \mathrm{d}s \\
A_4 &:= -\frac{1}{2}\int_0^T \left(b_\epsilon^* a^{-1} b_\epsilon(s, \overline{Y}_s) - b_\epsilon^* a^{-1} b_\epsilon(\phi(s), \overline{Y}_{\phi(s)})\right) \mathrm{d}s,
\end{aligned}$$

where for a vector x, x^i is the ith-component of the vector x, $\mathcal{E}_t = Y_t - \overline{Y}_t$ is the *error process*, and

$$\begin{aligned}
F &= \nabla g\left(\theta\left(Y_T, \log(Z_T(\Gamma^\epsilon))\right) + (1-\theta)\left(\overline{Y}_T, \log(Z_T(\tilde{\Gamma}^\epsilon))\right)\right), \\
F_{d+1} &= \partial_{d+1} g\left(\theta\left(Y_T, \log(Z_T(\Gamma^\epsilon))\right) + (1-\theta)\left(\overline{Y}_T, \log(Z_T(\tilde{\Gamma}^\epsilon))\right)\right) \\
&= g\left(\theta\left(Y_T, \log(Z_T(\Gamma^\epsilon))\right) + (1-\theta)\left(\overline{Y}_T, \log(Z_T(\tilde{\Gamma}^\epsilon))\right)\right),
\end{aligned}$$

where $\partial_{d+1} = \frac{\partial}{\partial z}$ denotes the partial derivative with respect to the component z.

The error process may be written as follows:

$$\mathcal{E}_t = \sum_{i=1}^d \int_0^t \alpha_i(s)\mathcal{E}_s \, \mathrm{d}W_s^i + \int_0^t G_s \, \mathrm{d}W_s, \tag{25}$$

where

$$G_s = \sigma(s, \overline{Y}_s) - \sigma(\phi(s), \overline{Y}_{\phi(s)}), \tag{26}$$

and for $i = 1, \cdots, d$, $\alpha_i(s)$ is a $d \times d$-matrix whose (j,k)th-component is

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

Then we can rewrite

$$\mathcal{E}_t = U_t \int_0^t U_s^{-1} G_s \, \mathrm{d}W_s - \sum_{i=1}^d U_t \int_0^t U_s^{-1} \alpha_i(s) (G_s)_{\cdot,i} \mathrm{d}s,$$

where for a matrix A, $A_{\cdot,i}$ denotes the ith-column of the matrix A and U and U^{-1} are the $d \times d$ matrices which are solutions of the stochastic differential equation

$$U_t = \mathrm{Id} + \sum_{i=1}^d \int_0^t \alpha_i(s) U_s \, \mathrm{d}W_s^i \text{ and } U_t^{-1} = \mathrm{Id} - \sum_{i=1}^d \int_0^t U_s^{-1} \alpha_i(s) \, \mathrm{d}W_s^i,$$

where Id is the $d \times d$-unit matrix. Note that U is stable in the sense that all $L^p(\Omega)$-sup norms are bounded. In the sense of expectations, we have $\mathcal{E} \approx O(n^{-1})$ using duality two times. This will also require two derivatives of all terms multiplying it. In this sense we can compute the orders of derivatives of b_ϵ. That is,

$$\mathbb{E}\left[F^* \mathcal{E}_T\right] \approx \|b_\epsilon\|_{2,\infty} n^{-1},$$

and similarly for all the other terms in (24). As the calculations are long we just show how to deal with the term above:

$$\mathbb{E}\left[F^* \mathcal{E}_T\right] = \mathbb{E}\left[F^* \left\{ U_T \int_0^T U_s^{-1} G_s \, \mathrm{d}W_s - \sum_{i=1}^d U_T \int_0^T U_s^{-1} \alpha_i(s) (G_s)_{\cdot,i} \mathrm{d}s \right\}\right]. \tag{27}$$

The above leads to two terms, we show how to bound the first, as the second is similar. For random variables β_{ij} uniform on $[0,1]$, $i, j = 1, \cdots, d$, we can replace the difference $G_s = \sigma(s, \overline{Y}_s) - \sigma(\phi(s), \overline{Y}_{\phi(s)})$ in the above expression by

$$\int_{\phi(s)}^s \partial_t \sigma(u, \overline{Y}_s) \, \mathrm{d}u + \left\{ \int_0^1 \left(\nabla \sigma_{i,j}(\beta_{ij} \overline{Y}_s + (1-\beta_{ij}) \overline{Y}_{\phi(s)}) \right)^* \sigma(\phi(s), \overline{Y}_{\phi(s)}) \int_{\phi(s)}^s \mathrm{d}W_u \, \mathrm{d}\beta_{ij} \right\}_{i,j=1,\cdots,d}.$$

The first term in brackets leads to a term of order n^{-1}. We only need to analyze what happens when the second term is replaced in (27), that is for

$$\mathbb{E}\left[F^* U_T \int_0^T \left\{ \sum_{l=1}^d (U_s^{-1} \alpha_i(s))_{j,l} \left(\int_0^1 \nabla \sigma_{l,k}(\beta_{lk} \overline{Y}_s + (1-\beta_{lk}) \overline{Y}_{\phi(s)}) \, \mathrm{d}\beta_{lk} \right)^* \sigma(\phi(s), \overline{Y}_{\phi(s)}) \int_{\phi(s)}^s \mathrm{d}W_u \right\}_{j,k=1,\cdots,d} \mathrm{d}W_s \right].$$

To analyze this term, we will repeatedly use the following duality formula of Malliavin Calculus

$$\mathbb{E}\left[H^* \int_a^b u_s \mathrm{d}W_s \right] = \sum_{i,j=1}^d \mathbb{E}\left[\int_a^b D_s^j H^i \cdot u_{ij}(s) \, \mathrm{d}s \right]$$

where a d-dimensional random vector $H \in \mathbb{D}^{1,2}$ and a $d \times d$-matrix adapted process $u \in \mathbb{L}^{1,2}$ (for notations, see Nualart [30]). Applying this formula two times, the first term in (27) can be rewritten as

$$\sum_{j,k=1}^d \sum_{l=1}^d \mathbb{E}\left[\int_0^T D_s^k (F^* U_T)_j \cdot (U_s^{-1} \alpha_i(s))_{j,l} \left(\nabla \sigma_{l,k}(\beta_{lk} \overline{Y}_s + (1-\beta_{lk}) \overline{Y}_{\phi(s)}) \right)^* \sigma(\phi(s), \overline{Y}_{\phi(s)}) \int_{\phi(s)}^s \mathrm{d}W_u \, \mathrm{d}s \right]$$

$$= \sum_{j,k=1}^d \sum_{l=1}^d \sum_{m=1}^d \int_0^T \mathbb{E}\left[\int_{\phi(s)}^s D_u^m \left(D_s^k (F^* U_T)_j \cdot (U_s^{-1} \alpha_i(s))_{j,l} \left(\nabla \sigma_{l,k}(\beta_{lk} \overline{Y}_s + (1-\beta_{lk}) \overline{Y}_{\phi(s)}) \right)^* \sigma_{\cdot,m}(\phi(s), \overline{Y}_{\phi(s)}) \right) \mathrm{d}u \right] \mathrm{d}s$$

The last part of the calculation corresponds to find a uniform bound for the integrand above. Note that all stochastic derivatives (up to order 3) of Y and $\overline{Y}$ are bounded. Similarly for U (up to order 2). Every derivative of $\log(Z_T(\Gamma^\epsilon))$ and $\log(Z_T(\tilde{\Gamma}^\epsilon))$ will introduce a derivative of b_ϵ.

To get the required bound, we use the chain rule for stochastic derivatives on F, together with the Cauchy-Schwarz inequality and apply Lemma 9 in Appendix in order to separate the terms.

To show this, using the chain rule for the Malliavin derivative we get that

$$\left|\mathbb{E}\left[\int_{\phi(s)}^{s} D_u^m\Big(D_s^k(F^*U_T)_j\cdot(U_s^{-1}\alpha_i(s))_{j,l}\Big(\nabla\sigma_{l,k}(\beta_{lk}\overline{Y}_s+(1-\beta_{lk})\overline{Y}_{\phi(s)})\Big)^*\sigma_{\cdot,m}(\phi(s),\overline{Y}_{\phi(s)})\Big)\,\mathrm{d}u\right]\right|$$

$$\leq \mathbb{E}\left[\sup_{k=1,2,3}|\nabla^k f(\theta Y_T+(1-\theta)\overline{Y}_T)|\exp(\theta\log(Z_T(\Gamma^\epsilon))+(1-\theta)\log(Z_T(\tilde{\Gamma}^\epsilon)))\int_0^T\int_{\phi(s)}^{s} h(u,s,T)\,\mathrm{d}u\,\mathrm{d}s\right],$$

where $h(u,s,T)$ depends on products of the three processes U, Y, $\overline{Y}$, σ and their derivatives up to order 2. As all these three processes belong to the space of processes A such that $\mathbb{E}[\sup_{t\in[0,T]}|A_t|^k]<+\infty$ for any $k\geq 0$, applying the polynomial growth property of of f and its derivatives together with Lemma 9 leads to a bound of type

$$\leq C_1\exp(C_2\|b_\epsilon\|_\infty^2)\int_0^T\int_{\phi(s)}^{s}\mathrm{d}u\,\mathrm{d}s$$

Note that the terms $\|\partial_x b_\epsilon\|_\infty^4$, $\|\partial_x b_\epsilon\|_\infty^2\|\partial_x^2 b_\epsilon\|_\infty$, $\|\partial_x^2 b_\epsilon\|_\infty^2$ and $\|\partial_x b_\epsilon\|_\infty\|\partial_x^3 b_\epsilon\|_\infty$, whose terms give the first divergence order $1/\epsilon^4$ as $\epsilon\to 0$, appear due to the term

$$F_{d+1}\int_0^T\Big(\nabla(b_\epsilon^*\sigma^{-1})^i(s,\theta_i'Y_s+(1-\theta_i')\overline{Y}_s)\Big)^*\mathcal{E}_s\,\mathrm{d}W_s$$

in (24).

We have deliberately avoided writing all the arguments as they are long. One can do it using Malliavin derivatives or just Itô formula. Either way is a series of long calculations. □

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

Conclusion 1. Let $\sigma\in C_b^{1,3}(\overline{H})$, $\mathfrak{M}=C_b^{1,3}(\overline{H})$, $\mathfrak{F}=C_p^3(\mathbb{R}^d)$ then $\beta=4$ and $\delta=1$. Under (H4) then the weak rate is at least $n^{-1+\frac{4}{\gamma+4}}$ in general. In the particular case that $d=1$ and $b(t,x)=\mathbf{1}_{[\zeta_1,\zeta_2]}(x)$ then according Proposition 2, we need to consider a bound for $d_p(b,b_\epsilon)=\|b-b_\epsilon\|_{Y,p}$ for $p>1$. Furthermore, assume that the Aronson estimates (11) are satisfied and then (12) can be used as an upper bound for $\|b-b_\epsilon\|_{Y,q}$ with $q\geq p>1$. We therefore obtain $\gamma=\frac{1}{q}$ with $q>1$. With these considerations we find an upper bound of the weak rate and therefore our arguments give a rate of at least $n^{-\frac{1}{5}+r}$ for any small $r>0$. Note in particular that this result improves the result in Theorem 1 in the case that $\alpha<\frac{2}{5}$.

Similar calculations can be used using Proposition 1 which will lead to somewhat worse rates. On the other hand, the values of q appearing in $K_q(f)$ will be lower. In the next conclusions, we will also use similar arguments and therefore where we will not give as much detail as in the case above.

Now we weaken the conditions on the coefficients and the rate of approximation of b, at the price of a lower rate of convergence.

Proposition 4. *Let us assume that* $\sigma \in \mathrm{H}^{1/2,1}(\overline{H})$ *satisfying* (H1), $b_\epsilon \in \mathrm{H}^{1/2,1}(\overline{H})$ *and* f *in* $\mathcal{C}^1_{\mathrm{p}}(\mathbb{R}^d)$ *or* $\mathcal{C}^1_{\mathrm{e}}(\mathbb{R}^d)$. *Then*

$$|\mathbb{E}[f(X^\epsilon_T)]-\mathbb{E}[f(\overline{X}^\epsilon_T)]| \leq (B(f,x)+B(\nabla f,x))\frac{C(T,\|b_\epsilon\|_\infty,\|\sigma\|_{(1/2,1)})(1+\|b_\epsilon\|_{(1/2,1)})}{\sqrt{n}} \quad (28)$$

with $B = B_{\mathrm{p}}$ *if* $f \in \mathcal{C}^1_{\mathrm{p}}(\mathbb{R}^d)$ *and* $B = B_{\mathrm{e}}$ *if* $f \in \mathcal{C}^1_{\mathrm{e}}(\mathbb{R}^d)$. *Besides, if* $\sigma(t,x) = \mathrm{Id}$ *for any* $(t,x) \in \overline{H}$, *then* $C(T,\|b_\epsilon\|_\infty,\|\sigma\|_{(1/2,1)}) = C(T,\|b_\epsilon\|_\infty) < \infty$.

Remark 5. Inequality (28) shall be compared with Theorem 1, where the Hölder exponent is not an integer. Our hypotheses correspond then to the "limit case" $\alpha = 1$. Yet our conditions on f are less stringent.

Proof. Since $\mathbb{E}[|W_s - W_{\phi(s)}|^2] = s - \phi(s)$, we obtain that for $g \in \mathrm{H}^{1/2,1}(\overline{H})$,

$$\int_0^T \mathbb{E}[(g(s,\overline{Y}_s) - g(\phi(s),\overline{Y}_{\phi(s)}))^2]\,\mathrm{d}s \leq C\int_0^T (s-\phi(s))\,\mathrm{d}s = \frac{CT^2}{2n},$$

where C depends on $\|g\|_{(1/2,1)}$ and $\|\sigma\|_\infty$.

From (26), and $\sigma \in \mathrm{H}^{1/2,1}(\overline{H})$,

$$\mathbb{E}[\|G_s\|^2] \leq C_1(s-\phi(s)), \quad (29)$$

where C_1 depends only on $\|\sigma\|_{(1/2,1)}$.

Note that $\mathcal{E}$ is a square integrable martingale. With the Itô formula applied to $\mathcal{E}_t^2$ where $\mathcal{E}$ has been defined in (25),

$$\mathbb{E}[|\mathcal{E}_t|^2] \leq \int_0^t C_2\mathbb{E}[|\mathcal{E}_s|^2]\,\mathrm{d}s + \int_0^t \mathbb{E}[\|G_s\|^2]\,\mathrm{d}s,$$

were C_2 depends on $\|\nabla\sigma\|_\infty$ and d. With the Gronwall Lemma and (29),

$$\mathbb{E}[|\mathcal{E}_T|^2] \leq C_1 e^{C_2 T}\frac{T^2}{n}.$$

Applying the Burkholder-Davies-Gundy inequality to (25) and using this expression, one gets that

$$\mathbb{E}\left[\sup_{t\in[0,T]} |\mathcal{E}_t|^2\right] \leq \frac{C(T,\|\sigma\|_{(1/2,1)})}{n}. \quad (30)$$

Using the Cauchy-Schwarz inequality, Lemma 7 with $\alpha = 4$, and Lemma 2 with $k = 4$,

$$\mathbb{E}[|F|^2]^{1/2} + \mathbb{E}[(F_{d+1})^2]^{1/2} \leq C(\|b\|_\infty,\lambda,T)(B(f,x)+B(\nabla f,x)).$$

Using the Cauchy-Schwarz inequality on (24) to separate F from $\mathcal{E}_T$ and F_{d+1} from the terms $A_1,\dots,A_4$, Lemma 10 and (30), (28) is easily obtained.

If σ is constant, then $\overline{Y} = Y$ and $\mathcal{E} = 0$. Hence the simplification. □

Conclusion 2. Let $\sigma \in \mathrm{H}^{1/2,1}(\overline{H})$, $\mathfrak{M} = \mathrm{H}^{1/2,1}(\overline{H})$, $\mathfrak{F} = C^1_p(\mathbb{R}^d)$. Then $\beta = 1$, $\delta = 1/2$ and under (H4), the weak rate of convergence is at least of the order $n^{-\frac{1}{2}+\frac{1}{2(\gamma+1)}}$. As in Conclusion 1, in the particular case that $d = 1$ and $b(t,x) = \mathbf{1}_{[\zeta_1,\zeta_2]}(x)$, we achieve a rate of $n^{-\frac{1}{4}+r}$ for any small $r > 0$.

8.2 Case of Hölder continuous coefficients

The weak rate of convergence of the Euler-Maruyama scheme when the coefficients of the PDE are Hölder continuous has been given in Theorem 1. As mentioned before this result seems to be non-optimal. Still putting all the previous results together we obtain the following conclusion.

Conclusion 3. Let $\sigma \in \mathrm{H}^{\alpha/2,\alpha}(\overline{H})$, $\mathfrak{M} = \mathrm{H}^{\alpha/2,\alpha}(\overline{H})$, $\mathfrak{F} = H^{2+\alpha}(\mathbb{R}^d)$. Then $\beta = \alpha$, $\delta = E(\alpha)$ and under (H4), the weak rate of convergence is at least of the order $n^{-E(\alpha)+\frac{\alpha E(\alpha)}{\gamma+\alpha}}$. As in Conclusion 1, in the particular case that $d = 1$ and $b(t,x) = \mathbf{1}_{[\zeta_1,\zeta_2]}(x)$, we achieve a rate of $n^{-\frac{E(\alpha)}{1+\alpha}+r}$ for any small $r > 0$.

9 Rate of convergence of the Euler-Maruyama scheme with a constant diffusion coefficient

In this section, we use Proposition 1, so that we have $p > 2$. We have then always used $q = 2$.

In order to test our results in particular cases and make the different terms as explicit as possible, we will deal in this section with two cases. First, the simple case of a constant diffusion coefficient and homogeneous drift coefficient. Later we will further give an explicit example of discontinuous drift coefficient b. To keep it simple, we assume that σ is the identity matrix Id and therefore X is solution to

$$X_t = x + B_t + \int_0^t b(X_s)\,\mathrm{d}s \tag{31}$$

and $Y_t = x + B_t$. Let $b_\epsilon \in \mathfrak{M}$ be a family of approximations of b satisfying (H3) and (H4).

Similarly, let $\overline{X}$ and $\overline{X}^\epsilon$ be the continuous Euler-Maruyama schemes

$$\overline{X}_t = x + B_t + \int_0^t b(\overline{X}_{\phi(s)})\,\mathrm{d}s \text{ and } \overline{X}^\epsilon_t = x + B_t + \int_0^t b_\epsilon(\overline{X}^\epsilon_{\phi(s)})\,\mathrm{d}s.$$

In order to make the calculations as explicit as possible we center our error estimates on v, the solution to the PDE

$$\frac{\partial v}{\partial t}(t,x) + \frac{1}{2}\triangle v(t,x) = 0 \text{ with } v(T,x) = f(x)$$

where the terminal condition f belongs to $\mathcal{C}_{Sl}(\mathbb{R}^d)$ which is our most general case. From Theorem 2 and the isometry property of stochastic integrals,

$$K_2(f) := \|\nabla v\|_{Y,2} \leq \sqrt{\mathrm{Var}(f(x+B_T))} < +\infty.$$

As we make use of Proposition 1, there is no restriction to consider only the case $q = 2$ as in this proposition, $p > \frac{q}{q-1} \vee 2$ and $q/(q-1) < 2$ as soon as $q > 2$.

Proposition 1 may be applied even for a non-anticipative drift. Under the above conditions, for $r \geq 2$,

$$|\mathbb{E}[f(Z^{(1)})] - \mathbb{E}[f(Z^{(2)})]| \leq C_2(r)\mathbb{E}\left[\int_0^T |\beta^{(1)}(s) - \beta^{(2)}(s)|^r\,\mathrm{d}s\right]^{1/r} K_2(f) \tag{32}$$

where $(Z^{(i)}, \beta^{(i)})$ for $i = 1, 2$ is one of the following:

(i) $Z^{(i)} = X_T$ and $\beta^{(i)}(s) = b(x + B_s)$.
(ii) $Z^{(i)} = \overline{X}_T$ and $\beta^{(i)}(s) = b(x + B_{\phi(s)})$.
(iii) $Z^{(i)} = X^\epsilon_T$ and $\beta^{(i)}(s) = b_\epsilon(x + B_s)$.
(iv) $Z^{(i)} = \overline{X}^\epsilon_T$ and $\beta^{(i)}(s) = b_\epsilon(x + B_{\phi(s)})$.

9.1 Rate of convergence of smooth approximations of the drift

The next lemma is a direct consequence of Proposition 1 and the Hölder inequality of the Gaussian density.

Lemma 3. *For $p > 2 \vee d$, there exists a constant C such that*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X^\epsilon_T)]| \le CK_2(f)\|b - b_\epsilon\|_{\mathrm{L}^p(H)},$$

where C depends only on p, T and the constants C_1 and C_2 in Lemma 1.

Proof. This follows from (32) with $Z^{(1)}_T = X_T$ and $Z^{(2)}_T = X^\epsilon_T$ and Lemma 1 to compare $\|b - b_\epsilon\|_{Y,r} \le C_3\|b - b_\epsilon\|_{\mathrm{L}^{\infty,p}}$ when $d/2p < 1/r$, $p > r$. □

9.2 Rate of convergence of the Euler scheme with smooth approximations of the drift

The weak rate of convergence of the Euler-Maruyama scheme to the solution to (31) has been studied by V. Mackevičius in [27] for a drift coefficient which is Lipschitz continuous.

Theorem 5 (R. Mackevičius, [27, Theorem 1])**.** *If b_ϵ is bounded Lipschitz continuous with constant* $\mathrm{Lip}(b_\epsilon)$ *and $f \in \mathcal{C}^3_{\mathrm{p}}(\mathbb{R}^d)$, then there exists a constant $C(T, \Lambda, f)$ such that*

$$|\mathbb{E}[f(X^\epsilon_T)] - \mathbb{E}[f(\overline{X}^\epsilon_T)]| \le \frac{C(T, \Lambda, f)}{n} \mathrm{Lip}(b_\epsilon).$$

Remark 6. The proof is given for the dimension $d = 1$, but it is remarked in the article that it is suitable whatever the dimension (See Remark below Theorem 1 in [27]). The statement of Theorem 1 in [27] is slightly different since b is not assumed to be bounded. Yet it is clear from the proof that the rate of convergence is linear in $\mathrm{Lip}(b_\epsilon)$ if b_ϵ is also bounded.

Conclusion 4. Let $\mathfrak{M} = \mathrm{Lip}(\mathbb{R}^d)$, $\mathfrak{F} = \mathcal{C}^3_p(\mathbb{R}^d)$. Then $\beta = 1$, $\delta = 1$ and under (H4), the weak rate of convergence is at least of the order $n^{-1+\frac{1}{\gamma+1}}$. As in Conclusion 1, in the particular case that $d = 1$ and $b(t, x) = \mathbf{1}_{[\zeta_1,\zeta_2]}(x)$, we achieve a rate of $n^{-\frac{1}{2}+r}$ for any small $r > 0$.

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. *For $p > 2$ and for $f \in \mathcal{C}_{Sl}(\mathbb{R}^d)$, there exists a constant $C(p, T)$ such that*

$$|\mathbb{E}[f(X^\epsilon_T)] - \mathbb{E}[f(\overline{X}^\epsilon_T)]| \le \frac{C(p, T)}{\sqrt{n}} \mathrm{Lip}(b_\epsilon)K_2(f)$$

Proof. Using (32) with $Z^{(1)} = X^\epsilon_T$ and $Z^{(2)} = \overline{X}^\epsilon_T$ and $r = p$, this follows from

$$\mathbb{E}\left[\int_0^T |b_\epsilon(Y_s) - b_\epsilon(Y_{\phi(s)})|^p \,\mathrm{d}s\right]^{1/p} \leq \mathrm{Lip}(b_\epsilon) K(p)^{1/p} \left(\int_0^T |s-\phi(s)|^{p/2}\,\mathrm{d}s\right)^{1/p}$$
$$= \mathrm{Lip}(b_\epsilon) K(p)^{1/p} \frac{T^{1/2+1/p}}{\sqrt{n}} \left(\frac{p}{2}+1\right)^{-1/p}$$

where $\mathbb{E}[|Y_t - Y_s|^p] \leq K(p)|t-s|^{p/2}$. □

A generalization of the above result may be given as follows.

Proposition 6. *For $f \in \mathcal{C}_{Sl}(\mathbb{R}^d)$, $p > 2$ and for $b_\epsilon \in \mathcal{C}^2_{\mathrm{b}}(\mathbb{R}^d)$, there exists a constant $C_7(p,\Lambda,T)$ such that*

$$|\mathbb{E}[f(X^\epsilon_T)] - \mathbb{E}[f(\overline{X}^\epsilon_T)]| \leq \frac{C(p,\Lambda,T)}{n^{1/p}} \max\{\|\triangle b_\epsilon\|^{1/p}_\infty, \|\nabla b_\epsilon\|^{2/p}_\infty\} K_2(f).$$

Proof. The proof is similar to the one of Proposition 5 but where the Itô formula is used on $g(y, B_t) = |b_\epsilon(y+B_t) - b_\epsilon(y)|^p$. Note that

$$\|\triangle g(y,\cdot)\|_\infty \leq 2^{p-2}p(p-1)\Lambda^{p-2}\|\nabla b_\epsilon\|^2_\infty + 2^{p-1}p\Lambda^{p-1}\|\triangle b_\epsilon\|_\infty.$$

Hence the result. □

Conclusion 5. Let $\mathfrak{M} = \mathcal{C}^2_b(\mathbb{R}^d)$, $\mathfrak{F} = \mathcal{C}_{Sl}(\mathbb{R}^d)$. Then $\beta = 2/p$, $\delta = 1/p$ for $p > 2$ and under (H4), the weak rate of convergence is at least of the order $n^{-\frac{1}{p}+\frac{2}{p(p\gamma+2)}}$. As in Conclusion 1, in the particular case that $d = 1$ and $b(t,x) = \mathbf{1}_{[\zeta_1,\zeta_2]}(x)$ (with the further assumption that $\|\nabla v\|_{Y,q} < \infty$ for any q big enough), we achieve a rate of $n^{-\frac{1}{4}+r}$ for any small $r > 0$.

9.3 The weak error of the Euler scheme for X

In this particular case, one can also study the distance $d(\overline{X}(b_\epsilon), \overline{X})$.

Lemma 4. *For $p > d \vee 2$, there exists a constant C depending on p, Λ, λ and T such that*

$$|\mathbb{E}[f(\overline{X}_T)] - \mathbb{E}[f(\overline{X}^\epsilon_T)]| \leq CK_2(f)\|b - b_\epsilon\|_{\mathrm{L}^p(H)}.$$

Proof. Fix $p > 2$ and choose $\gamma' \in (1 \vee d/2, p/2)$. We apply (32) with $Z^{(1)} = \overline{X}_T$, $Z^{(2)} = \overline{X}^\epsilon_T$ and $r = p/\gamma' > 2$, so that

$$|\mathbb{E}[f(\overline{X}_T)] - \mathbb{E}[f(\overline{X}^\epsilon_T)]| \leq C_2(p,2)K_2(f)\mathbb{E}\left[\int_0^T |\beta(s) - \beta_\epsilon(s)|^{p/\gamma'}\,\mathrm{d}s\right]^{\gamma'/p}$$

with $\beta(s) = b(x + B_{\phi(s)})$ and $\beta_\epsilon(s) = b_\epsilon(x + B_{\phi(s)})$. The proof is now a variation of the one of Lemma 1. Note that for $\gamma \geq 1$,

$$\left(\int_{\mathbb{R}^d} \mathrm{g}_t(x-y)^\gamma\,\mathrm{d}y\right)^{1/\gamma} \leq \frac{1}{\gamma^{d/(2\gamma)}} \cdot \frac{1}{t^{(\gamma-1)d/(2\gamma)}}.$$

Let $\gamma = \gamma'/(\gamma'-1)$. Then there exists some constants $C(\gamma) = \gamma^{-\frac{d}{2\gamma}}$ and $C'(T,\gamma) = \frac{2\gamma' C(\gamma)}{(2\gamma'-d)T^{d/2\gamma'-1}}$ such that

$$\mathbb{E}\left[\int_0^T |\beta(s)-\beta_\epsilon(s)|^{p/\gamma'}\,\mathrm{d}s\right] = \frac{T}{n}\sum_{k=1}^n \int_{\mathbb{R}^d} \mathrm{g}_{t_k}(x-y)|b(y)-b_\epsilon(y)|^{p/\gamma'}\,\mathrm{d}y$$
$$\leq \frac{C(\gamma)}{T^{d/2\gamma'-1}}\cdot\frac{1}{n}\sum_{k=1}^n \frac{n^{d/2\gamma'}}{k^{d/2\gamma'}}\left(\int_{\mathbb{R}^d}|b(y)-b_\epsilon(y)|^p dy\right)^{1/\gamma'} \leq C'(T,\gamma)\|b-b_\epsilon\|_{\mathrm{L}^p}^{p/\gamma'}.$$

The last inequality follows from the identity

$$\frac{1}{n}\sum_{k=1}^n \frac{n^{d/2\gamma'}}{k^{d/2\gamma'}} \leq \int_0^1 \frac{1}{s^{d/2\gamma'}}\,\mathrm{d}s \leq \frac{2\gamma'}{2\gamma'-d}$$

for any $n>1$. Hence the result. □

In order to be even more explicit we make a hypothesis on the irregularity of the function b. 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 6. *Let b be a bounded function on $\mathbb{R}^d$ which is Lipschitz except on a set G such that the Lebesgue* $\mathrm{meas}(G(\epsilon)) = \mathrm{O}(\epsilon^\kappa)$ *for some $\kappa\geq 1$.*

(I) *For any $f\in\mathcal{C}^3_{\mathrm{p}}(\mathbb{R}^d)$,*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)]| = \mathrm{O}\left(n^{-1+\frac{p}{p+\kappa}}\right) \text{ for } p > d\vee 2. \tag{33}$$

(II) *For any $f\in\mathcal{C}_{Sl}(\mathbb{R}^d)$,*

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)]| = \mathrm{O}\left(n^{\frac{1}{2}\left(-1+\frac{p}{p+\kappa}\right)}\right) \text{ for } p > d\vee 2.$$

Proof. (I) Combining Theorem 5 with Lemmas 3 and 4, then for a choice of $p > d\vee 2$, there exists constant $C(p,\Lambda,T)$ and $C'(\Lambda,T,f)$ such that

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)]| \leq C(p,\Lambda,T)K_2(f)\|b-b_\epsilon\|_{\mathrm{L}^p(H)} + \frac{C'(\Lambda,T,f)}{n}\mathrm{Lip}(b_\epsilon).$$

Hence for any $\epsilon>0$, we consider a family of smooth approximations b_ϵ of b such that $b=b_\epsilon$ outside the set $G(\epsilon)=\{x\in\mathbb{R}^d|d(x,G)\leq\epsilon\}$.

Hence, $\mathrm{Lip}(b_\epsilon) = \mathrm{O}(\epsilon^{-1})$ and $b(x) = b_\epsilon(x)$ outside the set $G(\epsilon)$ of Lebesgue measure $|G(\epsilon)| = \mathrm{O}(\epsilon^\kappa)$. If follows that $\|b-b_\epsilon\|_{\mathrm{L}^p} = \mathrm{O}(\epsilon^{\kappa/p})$. Hence, the weak error of the Euler-Maruyama scheme is bounded by $\frac{C}{n\epsilon} + C'\epsilon^{\kappa/p}$ for some constants C and C' that depend only on Λ, T, p and f. Optimizing over the choice of ϵ, this leads to (33) using Corollary 1.

The proof of (II) is similar when one uses Proposition 5. □

Conclusion 6. In (I), letting $p\to d\vee 2$, the weak rate is at least of order $n^{-\frac{\kappa}{d+\kappa}+r}$ (resp. $n^{-\frac{\kappa}{2+\kappa}+r}$) when $d>2$ (resp. $d=1$) with r arbitrary small. However, the constant hidden in the $\mathrm{O}(n^{-\kappa/(\kappa+p)})$ explodes as $p\to d\vee 2$. With our estimates, a better rate of convergence is obtained at the cost of a bigger constant in front of the rate. In (II), the rate is at least of the order $n^{-\frac{1}{2}\frac{\kappa}{d\vee 2+\kappa}+r}$ for any small $r>0$ with a constant exploding as $r\downarrow 0$.

9.4 A particular case with a weak approximation rate of n^{-1}

In this section, we show that a rate of convergence of order 1 could be achieved. For this, we consider the following family of SDE

$$\mathrm{d}X_t(x) = b(X_t(x))\,\mathrm{d}t + \mathrm{d}W_t,\ X_0 = x,$$

where for $\theta > 0$,

$$b(x) := \begin{cases} -\theta, & x > 0, \\ 0, & x = 0, \\ \theta, & x < 0. \end{cases}$$

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. From Karatzas and Shreve [21, Section 6.5, (5.14), p. 441], the transition density function of a Brownian motion of $X_t(x)$, $x \geq 0$ is

$$\mathsf{p}_t(x,z) = \begin{cases} \dfrac{1}{\sqrt{2\pi t}}\left[\exp(-\dfrac{(z-x+\theta t)^2}{2t}) + \theta e^{-2\theta z}\displaystyle\int_{x+z}^{\infty}\exp(-\dfrac{(v-\theta t)^2}{2t})dv\right] & \text{for } z > 0, \\ \dfrac{1}{\sqrt{2\pi t}}\left[\exp(-\dfrac{2\theta x-(z-x-\theta t)^2}{2t}) + \theta e^{2\theta z}\displaystyle\int_{x-z}^{\infty}\exp(-\dfrac{(v-\theta t)^2}{2t})dv\right] & \text{for } z \leq 0. \end{cases} \tag{34}$$

Theorem 7. *For $f \in \mathcal{C}^3_{\mathrm{p}}(\mathbb{R})$, we have the following weak error:*

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

where C is a positive constant.

Although in this particular case, one may feel that there is no need to study the weak rate as there are other more direct simulation methods, we perform this study in order to show that in general the optimal rate is probably $\frac{1}{n}$. Also we remark that from the analytic point of view the fact that the starting point for the diffusion is zero is strongly used through the symmetry of the density of the associated Euler-Maruyama scheme.

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

Lemma 5. *Assume that $f \in \mathcal{C}^3_{\mathrm{p}}(\mathbb{R};\mathbb{R})$. Define $u(t,x) = \mathbb{E}[f(X_{T-t}(x))]$, $0 \leq t \leq T$ and $x \in \mathbb{R}$. Then for $k = 0, \ldots, 3$,*

$$\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}}|\partial_x^k u(t,x)| \leq C. \tag{35}$$

Furthermore, u satisfies the following PDE

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x) + b(x)\frac{\partial u}{\partial x}(t,x) = 0,\ u(T,x) = f(x). \tag{36}$$

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 (34). From this formula, one can verify the various needed properties. In particular, we define the functions $h_\pm(x,t) = \exp(-\frac{(x\pm\theta t)^2}{2t})/\sqrt{2\pi t}$ and $g_\pm(x) = \exp(\pm 2\theta x)$. It holds that

$$\partial_x \mathsf{p}_t(x,z) = \begin{cases} -\partial_z h_-(x-z,t) - \theta g_+(x)h_+(x+z,t), & x \geq 0, z > 0, \\ -\partial_z h_+(x-z,t)g_+(x) + \theta g_+(x)h_+(x-z,t), & x \geq 0, z \leq 0, \\ -\partial_z h_+(x-z,t) + \theta g_-(x)h_-(x+z,t), & x < 0, z < 0, \\ -\theta g_-(x)h_-(x-z,t) - \partial_z h_-(x-z,t)g_-(x), & x < 0, z > 0. \end{cases}$$

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, u is differentiable three times for all x due to the successive application of the integration by parts formula. In particular, u satisfies the PDE (36). $\square$

Remark 7. In [21, Exercise 5.3, p. 441], it is verified the above theorem for a particular function f. In fact, without regularity conditions on f one may prove existence and uniqueness of u and that it satisfies (36) if f is an even function. In order for the PDE to be satisfied at $x=0$ one needs to have the symmetry of f. In fact, by a direct calculation one finds that

$$\partial_x^2 u(t,0+) - \partial_x^2 u(t,0-) = \frac{4\theta}{\sqrt{2\pi t}t}\int_0^\infty z\exp\left(-\frac{(z+\theta t)^2}{2t}\right)(f(z)-f(-z))\,\mathrm{d}z.$$

In what follows we need that the derivatives are uniformly bounded as stated in (35) and therefore stronger hypotheses on f are needed.

Let us denote by $\overline{\mathsf{p}}_t(x,y)$ the density transition function of the Euler scheme of step n.

Lemma 6. *Denote $\overline{\mathsf{p}}_{\phi(t)}(x) = \overline{\mathsf{p}}_{\phi(t)}(0,x)$, then we have that $\overline{\mathsf{p}}_{\phi(t)}(x) = \overline{\mathsf{p}}_{\phi(t)}(-x)$. Furthermore for any $n \in \mathbb{N}$ and any $t \geq \frac{T}{n}$ there exists a positive constant independent of n and x such that*

$$\overline{\mathsf{p}}_{\phi(t)}(x) \leq \frac{C}{\sqrt{\phi(t)}}. \tag{37}$$

Proof. The density $\overline{\mathsf{p}}_{T/n}(x,y)$ of $\overline{X}_{T/n}(x)$ is given by

$$\overline{\mathsf{p}}_{T/n}(x,y) = \exp\left(-\frac{(y-x+\mathrm{sgn}(x)\theta T/n)^2}{2T/n}\right).$$

Hence,

$$\overline{\mathsf{p}}_{kT/n}(x,y) = \iint_{\mathbb{R}^{k-1}} \overline{\mathsf{p}}_{T/n}(x,z_1)\cdots\overline{\mathsf{p}}_{T/n}(z_{k-1},y)\,\mathrm{d}z_1\cdots\mathrm{d}z_{k-1}. \tag{38}$$

As $\overline{\mathsf{p}}_{T/n}(x,y) = \overline{\mathsf{p}}_{T/n}(-x,-y)$ and $\overline{\mathsf{p}}_{T/n}(0,x) = \overline{\mathsf{p}}_{T/n}(0,-x)$, the symmetry of $\overline{\mathsf{p}}_{T/n}(0,\cdot)$ follows from (38).

Finally, (37) follows from the results in [26], where a Gaussian upper bound is given for the density $\overline{\mathsf{p}}_{\phi(t)}(x,y)$ of the Euler-Maruyama scheme. $\square$

Proof of Theorem 7. From the definition of u, Itô's formula, Equation (36) and the definition of $\overline{X}_t(0)$, we have

$$\begin{aligned}\mathbb{E}[f(X_T(0))] - \mathbb{E}[f(\overline{X}_T(0))] &= u(0,0) - \mathbb{E}[u(T, \overline{X}_T(0))] \\ &= \int_0^T \mathbb{E}\left[\frac{\partial u}{\partial x}(t, \overline{X}_t(0))\left\{b(\overline{X}_t(0)) - b(\overline{X}_{\phi(t)}(0))\right\}\right] \mathrm{d}t \\ &= \int_0^T \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial u}{\partial x}(t, x + b(x)(t-\phi(t)) + \sqrt{t-\phi(t)}z) \\ \times &\left\{\underline{b(x + b(x)(t-\phi(t)) + \sqrt{t-\phi(t)}z) - b(x)}\right\} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \overline{\mathsf{p}}_{\phi(t)}(x)\, \mathrm{d}x\, \mathrm{d}t. \quad (39)\end{aligned}$$

The underlined-part of equation (39) equals

$$\begin{cases} 2\mathbf{1}\left(z < \dfrac{-x+\theta(t-\phi(t))}{\sqrt{t-\phi(t)}}\right) & \text{for } \ x > 0 \\ -2\mathbf{1}\left(z > \dfrac{-x-\theta(t-\phi(t))}{\sqrt{t-\phi(t)}}\right) & \text{for } \ x < 0. \end{cases}$$

Then we perform the change of variables $y = \frac{x}{\sqrt{t-\phi(t)}}$. Here, we assume that this change is done for $t \neq \phi(t)$ which is a set of measure zero. The integrability follows from the argument we give here. We rewrite

$$\begin{aligned}(39) = 2\theta \int_0^T \sqrt{t-\phi(t)} &\left(\int_0^\infty dy \int_{-\infty}^{-y+\theta\sqrt{t-\phi(t)}} \mathrm{d}z \frac{\partial u}{\partial x}(t, \sqrt{t-\phi(t)}(y - \theta\sqrt{t-\phi(t)} + z))\right. \\ -\int_{-\infty}^0 \mathrm{d}y \int_{-y-\theta\sqrt{t-\phi(t)}}^{\infty} dz &\left.\frac{\partial u}{\partial x}(t, \sqrt{t-\phi(t)}(y + \theta\sqrt{t-\phi(t)} + z))\right) \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \overline{\mathsf{p}}_{\phi(t)}(\sqrt{t-\phi(t)}y)\, \mathrm{d}t. \end{aligned} \quad (40)$$

We apply Taylor's expansion with integral residue to $\frac{\partial u}{\partial x}(t, \sqrt{t-\phi(t)}y \mp \theta(t-\phi(t)) + \sqrt{t-\phi(t)}z)$:

$$\begin{aligned}\frac{\partial u}{\partial x}(t, \sqrt{t-\phi(t)}(y \mp \theta\sqrt{t-\phi(t)} + z)) &= \frac{\partial u}{\partial x}(t, y\sqrt{t-\phi(t)} \mp \theta(t-\phi(t))) \\ &+ \frac{\partial^2 u}{\partial x^2}(t, y\sqrt{t-\phi(t)} \mp \theta(t-\phi(t)))\sqrt{t-\phi(t)}z \\ + \int_0^1 \frac{\partial^3 u}{\partial x^3}(t, y\sqrt{t-\phi(t)} \mp \theta(t-\phi(t)) &+ \alpha\sqrt{t-\phi(t)}z)(1-\alpha)d\alpha(t-\phi(t))z^2. \quad (41)\end{aligned}$$

Due to Lemma 5, u is smooth enough and therefore $\frac{\partial^3 u}{\partial x^3}$ has a uniform upper estimation. The third term is already of order $1/n$.

With this expansion, we consider the first order terms generated in (40) using the first order terms in (41).

$$\begin{aligned}2\theta \int_0^T \sqrt{t-\phi(t)} &\left(\int_0^\infty \mathrm{d}y\Phi\left(-y + \theta\sqrt{t-\phi(t)}\right) \frac{\partial u}{\partial x}(t, y\sqrt{t-\phi(t)} - \theta(t-\phi(t)))\right. \\ -\int_{-\infty}^0 \mathrm{d}y &\left.\left\{1 - \Phi\left(-y - \theta\sqrt{t-\phi(t)}\right)\right\} \frac{\partial u}{\partial x}(t, y\sqrt{t-\phi(t)} + \theta(t-\phi(t)))\right) \overline{\mathsf{p}}_{\phi(t)}(y\sqrt{t-\phi(t)})\, \mathrm{d}t,\end{aligned} \quad (42)$$

where $\Phi(x)$ is the distribution function of the standard normal distribution. Using the symmetric property of the standard normal density function, changing variables as $z = -y$ and using the symmetry (see Lemma 6) $\bar{\mathsf{p}}_{\phi(t)}(x) = \bar{\mathsf{p}}_{\phi(t)}(-x)$, the second term of (42) is

$$-2\theta \int_0^T \sqrt{t-\phi(t)} \int_0^\infty \Phi\left(-y+\theta\sqrt{t-\phi(t)}\right) \frac{\partial u}{\partial x}(t, -y\sqrt{t-\phi(t)}+\theta(t-\phi(t)))\bar{\mathsf{p}}_{\phi(t)}(y\sqrt{t-\phi(t)})\,\mathrm{d}y\,\mathrm{d}t.$$

And then we can write (42) as

$$(42) = 2\theta \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \sqrt{t-\phi(t)} \int_0^\infty \left\{ \frac{\partial u}{\partial x}(t, y\sqrt{t-\phi(t)} - \theta(t-t_i)) - \frac{\partial u}{\partial x}(t, -y\sqrt{t-\phi(t)} + \theta(t-t_i)) \right\} \times \Phi\left(-y+\theta\sqrt{t-t_i}\right) \bar{\mathsf{p}}_{t_i}(y\sqrt{t-\phi(t)})\,\mathrm{d}y\,\mathrm{d}t. \quad (43)$$

We use again Taylor's expansion to $\frac{\partial u}{\partial x}(t, \pm(y\sqrt{t-t_i} - \theta(t-t_i)))$ around zero, obtaining

$$\frac{\partial u}{\partial x}(t, y\sqrt{t-t_i} - \theta(t-t_i)) - \frac{\partial u}{\partial x}(t, -(y\sqrt{t-t_i} - \theta(t-t_i))) = 2\frac{\partial^2 u}{\partial x^2}(t,0)y\sqrt{t-t_i} - 2\theta\frac{\partial^2 u}{\partial x^2}(t,0)(t-t_i) + 2\int_0^1 \left[\frac{\partial^3 u}{\partial x^3}(t, \alpha(y\sqrt{t-t_i} - \theta(t-t_i))) - \frac{\partial^3 u}{\partial x^3}(t, -\alpha(y\sqrt{t-t_i} - \theta(t-t_i))) \right] \times (1-\alpha)\,\mathrm{d}\alpha (y\sqrt{t-t_i} - \theta(t-t_i))^2. \quad (44)$$

We consider (43) with the first term of (44).

$$\left| 2\int_{t_i}^{t_{i+1}} \int_0^\infty \frac{\partial^2 u}{\partial x^2}(t,0)y(t-t_i)\Phi\left(-y+\theta\sqrt{t-t_i}\right) \bar{\mathsf{p}}_{t_i}(y\sqrt{t-t_i})\,\mathrm{d}y\,\mathrm{d}t \right| \leq \frac{2T}{n} \int_{t_i}^{t_{i+1}} \left| \frac{\partial^2 u}{\partial x^2}(t,0) \right| \int_0^\infty y\Phi\left(-y+\theta\sqrt{t-t_i}\right) \bar{\mathsf{p}}_{t_i}(y\sqrt{t-t_i})\,\mathrm{d}y\,\mathrm{d}t. \quad (45)$$

Therefore the above integral is uniformly bounded due to Lemma 6. The second order terms in (41) are treated in a similar fashion with a Taylor expansion of order one for $\frac{\partial^2 u}{\partial x^2}(t, y\sqrt{t-\phi(t)} \mp \theta(t-\phi(t)))$ around 0 which will give the needed estimates. In fact, one obtains for the first term in (40) corresponding to this case

$$- 2\theta \int_0^T (t-\phi(t))^{3/2} \int_0^\infty \frac{\partial^2 u}{\partial x^2}(t,0)(y-\theta\sqrt{t-\phi(t)}) \frac{e^{-\frac{(y-\theta\sqrt{t-\phi(t)})^2}{2}}}{\sqrt{2\pi}} \bar{\mathsf{p}}_{\phi(t)}(\sqrt{t-\phi(t)}y)\,\mathrm{d}y\,\mathrm{d}t$$
$$- 2\theta \int_0^T (t-\phi(t))^{3/2} \int_{-\infty}^0 \frac{\partial^2 u}{\partial x^2}(t,0)(y+\theta\sqrt{t-\phi(t)}) \frac{e^{-\frac{(y+\theta\sqrt{t-\phi(t)})^2}{2}}}{\sqrt{2\pi}} \bar{\mathsf{p}}_{\phi(t)}(\sqrt{t-\phi(t)}y)\,\mathrm{d}y\,\mathrm{d}t. \quad (46)$$

We remark that for the residues the explicit bound found in Lemma 6 is not needed as the only property used is then that the density integrates to one.

Also, as it can be seen from the signs in (46), the two applications of Taylor's expansions can not be achieved at once as the symmetries and evaluations carried in the previous steps only apply to the first term of the expansion in (41). □

9.5 Numerical Results

In this section, we report on some simulations done in various situations. In order to save space, we will not give any graphs of actual simulations. Simulations where 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

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

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 3000 where carried out. The number of Monte Carlo simulations is n^2. Then $n = 3000$ 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 9.4 is much more general.

A part of the simulation experiments can be found in [23]. Other simulation experiments will appear elsewhere.

10 Conclusions

We have given several weak orders with respect to $|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T^{\epsilon})]|$ with discontinuous drift coefficients by analyzing the errors of $|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^{\epsilon})]|$ (Propositions 1 and 2 as well as Corollary 2), and $|\mathbb{E}[f(X_T^{\epsilon})] - \mathbb{E}[f(\overline{X}_T^{\epsilon})]|$ (Theorem 1, 4, and Proposition 4) under various assumptions to b, b_ϵ, σ and f. By combining those theorems, propositions and lemmas, we can derive a short table. In the table, we assume that the error $\|b - b_\epsilon\|_{Y,p}$ has an order $O(\epsilon^\gamma)$ for $\gamma > 0$. Note that γ depends on the dimension d and the moment p. In the last two columns we consider the particular case treated in Example 1. That is,

Case 1(C1): $d = 1$, $\gamma < 1/2$, Prop. 1

Case 2(C2): $d = 1$, $\gamma < 1$, Prop. 2.

The value given within parenthesis is the best rate possible. Then the values for α for which our result is better than Theorem 1 is given. Clearly C2 requires less integrability condition on f if one uses Corollary 2.

σ	$\mathfrak{M}$	$\mathfrak{F}$	β	δ	rate	C1	C2
$\mathcal{C}_b^{1,3}(\overline{H})$	$\mathcal{C}_b^{1,3}(\overline{H})$	$\mathcal{C}_p^3(\mathbb{R}^d)$	4	1	$-1 + \frac{4}{\gamma+4}$	$(-\frac{1}{9})$, $\alpha < \frac{2}{9}$	$(-\frac{1}{5})$, $\alpha < \frac{2}{5}$
$\mathrm{H}^{\frac{1}{2},1}(\overline{H})$	$\mathrm{H}^{\frac{1}{2},1}(\overline{H})$	$\mathcal{C}_p^1(\mathbb{R}^d)$	1	$\frac{1}{2}$	$-\frac{1}{2} + \frac{1}{2(\gamma+1)}$	$(-\frac{1}{6})$, $\alpha < \frac{1}{3}$	$(-\frac{1}{4})$, $\alpha < \frac{1}{2}$
$\mathrm{H}^{\frac{\alpha}{2},\alpha}(\overline{H})$	$\mathrm{H}^{\frac{\alpha}{2},\alpha}(\overline{H})$	$\mathrm{H}^{2+\alpha}(\mathbb{R}^d)$	α	$E(\alpha)$	$-E(\alpha) + \frac{\alpha E(\alpha)}{\gamma+\alpha}$	$(-\frac{E(\alpha)}{1+2\alpha})$	$(-\frac{E(\alpha)}{1+\alpha})$
const.	$\mathcal{C}_b^2(\mathbb{R}^d)$	$\mathcal{C}_{Sl}(\mathbb{R}^d)$	$\frac{2}{p}$	$\frac{1}{p}$	$-\frac{1}{p} + \frac{2}{p(\gamma p+2)}$	$(-\frac{1}{6})$, $\alpha < \frac{1}{3}$	

Table 1: Weak orders of $|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T^{\epsilon})]|$ using the results in Conclusions 1-4.

This table does not fully characterize all the cases treated in this article. In fact, we have also obtained some additional results where some of the conditions for $\mathfrak{M}$ or $\mathfrak{F}$ are weakened but they do not lead to a final change in the rate. For reasons of space, some of these results are not quoted here.

This article should be considered as a first attempt at understanding the approximation quality of the Euler scheme for stochastic differential equations with irregular coefficients. We have not obtained their optimal rates and there are clearly various questions that remain open and that should be treated in the near future.

11 Appendix

11.1 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 7. (I) *For any* $\alpha > 1$,

$$\mathbb{E}[Z_T^\alpha]^{1/\alpha} \leq \exp\left(\left(\alpha - \frac{1}{2}\right) T\hat{\Gamma}^2\right), \tag{47}$$

with $\hat{\Gamma} := \sup_{(s,x)\in[0,T]\times\mathbb{R}^d} |\Gamma(s,x)|$.

(II) *Let* $g \in \mathrm{L}^p(Y)$ *for some* $p < +\infty$. *Then there exists some constant* $\varkappa(\hat{\Gamma}, p, q)$ *such that for any* $q < p$,

$$\|g\|_{X,q} \leq \varkappa(\hat{\Gamma}, p, q)\|g\|_{Y,p}$$

with $\varkappa(\hat{\Gamma}, q, p) \xrightarrow[q\to p]{} +\infty$.

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

$$\mathbb{E}[Z_T^\alpha] \leq \mathbb{E}\left[\exp\left(2\alpha M_T - \frac{4\alpha^2}{2}\langle M\rangle_T\right)\right]^{1/2} \mathbb{E}[\exp\left((2\alpha^2 - \alpha)\langle M\rangle_T\right)]^{1/2}.$$

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 (47).
(II) Let $\mathbb{Q}^\Gamma$ be such that $\Gamma(t,x) := b(s,x)^*\sigma^{-1}(s,x)$. For $\alpha = p/q > 1$ and $\alpha' = \alpha/(\alpha-1)$, Hölder's inequality and (47) with (α, α') yields

$$\begin{aligned}\|g\|_{X,q} = \mathbb{E}\left[\int_0^T |g(s,X_s)|^q\,\mathrm{d}s\right]^{1/q} &= \mathbb{E}\left[Z_T\int_0^T |g(s,Y_s)|^q\,\mathrm{d}s\right]^{1/q} \\ &\leq T^{(p-q)/(pq)}\exp\left(T\hat{\Gamma}^2\frac{p+q}{2(p-q)q}\right)\|g\|_{Y,p}.\end{aligned}$$

□

Lemma 8. *Let* Z *be defined by* (9). *Then for any* $p \geq 2$ *and any* $\overline{p} > p$, *there exists some constant* $C(p, \overline{p})$ *depending only on* p *and* $\overline{p}$ *such that*

$$\mathbb{E}[|Z_T - 1|^p] \leq C(p,\overline{p})T^{p/2-2/\overline{p}}\exp\left(p\left(\frac{p\overline{p}}{\overline{p}-p} - \frac{1}{2}\right)T\hat{\Gamma}^2\right)\mathbb{E}\left[\int_0^T |\Gamma(s,X_s)|^{\overline{p}}\,\mathrm{d}s\right]^{p/\overline{p}}.$$

Proof. Since Γ is bounded, we have already noted that Z is a martingale. With the Burkholder-Davies-Gundy inequality (with constant $C(p)$) and the Hölder inequality applied twice,

$$\begin{aligned}\mathbb{E}[|Z_T - 1|^p] &\leq C(p)\mathbb{E}\left[\left(\int_0^T Z_s^2 |\Gamma(s,X_s)|^2 \,\mathrm{d}s\right)^{p/2}\right] \leq C(p)T^{p/2-1}\mathbb{E}\left[\int_0^T |Z_s|^p |\Gamma(s,X_s)|^p \,\mathrm{d}s\right] \\ &\leq C(p)T^{p/2-1-(\alpha-1)/\alpha}\mathbb{E}\left[\sup_{s\in[0,T]} |Z_s|^{p\alpha'}\right]^{1/\alpha'} \mathbb{E}\left[\int_0^T |\Gamma(s,X_s)|^{p\alpha} \,\mathrm{d}s\right]^{1/\alpha}\end{aligned}$$

for $\alpha = \overline{p}/p > 1$. The conclusion stems from Doob's inequality on Z and then Lemma 7(I). □

Lemma 9. *Let θ be a random variable with the uniform distribution on $(0,1)$, which is independent of W_t. Then for any $\alpha > 1$, we have*

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

And also the inequality holds for b_ϵ.

We can prove this lemma using similar arguments to Lemma 7.

11.2 A lemma on Hölder functions

Lemma 10. *The product of two functions in $\mathrm{H}^{1/2,1}(\overline{H})$ remains in $\mathrm{H}^{1/2,1}(\overline{H})$. Besides, if $\sigma \in \mathrm{H}^{1/2,1}(\overline{H})$ satisfies (H1), then $a = \sigma\sigma^*$ and a^{-1} belongs 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, $|f(t)g(t) - f(s)g(s)| \leq |f(t) - f(s)| \cdot \|g\|_\infty + |g(t) - g(s)| \cdot \|f\|_\infty$.

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

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

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

References

[1] F. Arandiga, A. Cohen, R. Donat, and N. Dyn, *Interpolation and approximation of piecewise smooth functions*, SIAM J. Numer. Anal. **43** (2005), no. 1, 41–57, DOI 10.1137/S0036142903426245.

[2] S. Arnold, *Approximation schemes for SDEs with discontinuous coefficients*, Ph.D. thesis, ETH, Zürich, 2006.

[3] D. G. Aronson, *Non-negative solutions of linear parabolic equations*, Ann. Scuola Norm. Sup. Pisa (3) **22** (1968), 607–694.

[4] S. Attanasio, *Stochastic flows of diffeomorphisms for one-dimensional SDE with discontinuous drift*, Electron. Commun. Probab. **15** (2010), 213–226, DOI 10.1214/ECP.v15-1545.

[5] 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] E. Clément, A. Kohatsu-Higa, and D. Lamberton, *A duality approach for the weak approximation of stochastic differential equations*, Ann. Appl. Probab. **16** (2006), no. 3, 1124–1154, DOI 10.1214/105051606000000060.

[11] P. Étoré and M. Martinez, *Exact simulation for SDE with discontinuous drift* (2013), available at `arXiv:1301.3019v1`.

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

[13] 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.

[14] 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.

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

[16] 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.

[17] I. Gyongy and M. Rasonyi, *A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients*, Stochastic processes and their applications **121** (2011), no. 10, 2189–2200, DOI 10.1016/j.spa.2011.06.008.

[18] 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.

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

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

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

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

[23] 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.

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

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

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

[27] V. Mackevičius, *On the convergence rate of Euler scheme for SDE with Lipschitz drift and constant diffusion*, Proceedings of the Eighth Vilnius Conference on Probability Theory and Mathematical Statistics, Part I (2002), 2003, pp. 301–310, DOI 10.1023/A:1025754020469.

[28] M. Martinez and D. Talay, *Time discretization of one dimensional Markov processes with generators under divergence form with discontinuous coefficients*, Electronic Journal Probab. **17** (2012), no. 27, DOI 10.1214/EJP.v17-1905.

[29] 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.

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

[31] A. Rozkosz and L. Słomiński, *On existence and stability of weak solutions of multidimensional stochastic differential equations with measurable coefficients*, Stochastic Process. Appl. **37** (1991), no. 2, 187–197, DOI 10.1016/0304-4149(91)90042-B.

[32] A. Semrau, *Discrete approximations of strong solutions of reflecting SDEs with discontinuous coefficients*, Bull. Pol. Acad. Sci. Math. **57** (2009), no. 2, 169–180, DOI 10.4064/ba57-2-10.

[33] D. W. Stroock, *Diffusion semigroups corresponding to uniformly elliptic divergence form operators*, Séminaire de Probabilités, XXII, Lecture Notes in Math., vol. 1321, Springer, Berlin, 1988, pp. 316–347.

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

[35] 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.

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