Numerische
Mathematik

CrossMark

Convergence of the Euler–Maruyama method for multidimensional SDEs with discontinuous drift and degenerate diffusion coefficient

Gunther Leobacher[1] · **Michaela Szölgyenyi**[2]

Abstract We prove strong convergence of order $1/4 - \epsilon$ for arbitrarily small $\epsilon > 0$ of the Euler–Maruyama method for multidimensional stochastic differential equations (SDEs) with discontinuous drift and degenerate diffusion coefficient. The proof is based on estimating the difference between the Euler–Maruyama scheme and another numerical method, which is constructed by applying the Euler–Maruyama scheme to a transformation of the SDE we aim to solve.

1 Introduction

We consider time-homogeneous stochastic differential equations (SDEs) of the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x, \tag{1}$$

✉ Michaela Szölgyenyi
michaela.szoelgyenyi@wu.ac.at

Gunther Leobacher
gunther.leobacher@uni-graz.at

1   Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, 8010 Graz, Austria

2   Institute for Statistics and Mathematics, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria

where $x \in \mathbb{R}^d$ is the initial value, $\mu : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ is the drift and $\sigma : \mathbb{R}^d \longrightarrow \mathbb{R}^{d \times d}$ is the diffusion coefficient.

The Euler–Maruyama approximation with step-size $\delta > 0$ of the solution to (1) is given by

$$X_t^\delta = x + \int_0^t \mu(X_{\underline{s}}^\delta)ds + \int_0^t \sigma(X_{\underline{s}}^\delta)dW_s \,, \tag{2}$$

with $\underline{s} = j\delta$ for $s \in [j\delta, (j+1)\delta)$, $j = 0, \ldots, (T - \delta)/\delta$. In particular, for $t \in \{j\delta : j = 0, \ldots, (T - \delta)/\delta\}$, we have

$$X_{t+\delta}^\delta = X_t^\delta + \mu(X_t^\delta)\delta + \sigma(X_t^\delta)(W_{t+\delta} - W_t) \,.$$

For μ, σ Lipschitz, Itô [9] proved existence and uniqueness of the solution of (1). In this case the Euler–Maruyama method (2) converges with strong order $1/2$ to the true solution, see [12, Theorem 10.2.2]. Higher order algorithms exist, but require stronger conditions on the coefficients.

In applications, frequently SDEs with less regular coefficients appear. For example in stochastic control theory, whenever the optimal control is of bang–bang type, meaning that the strategy is of the form $\mathbf{1}_S(X)$ for a measurable set $S \subseteq \mathbb{R}^d$, the drift of the controlled dynamical system is discontinuous. Furthermore, there are models which involve only noisy observations of a signal that has to be filtered. After applying filtering theory the diffusion coefficient typically is degenerate in the sense that $\|\sigma(x)^\top v\| = 0$, for some $x, v \in \mathbb{R}^d$. This motivates the study of SDEs with these kind of irregularities in the coefficients.

If μ is bounded and measurable, and σ is bounded, Lipschitz, and uniformly non-degenerate, i.e. if there exists a constant $c_0 > 0$ such that for all $x \in \mathbb{R}^d$ and all $v \in \mathbb{R}^d$ it holds that $\|\sigma(x)^\top v\| \geq c_0\|v\|$, Zvonkin [29] and Veretennikov [25,26] prove existence and uniqueness of a solution. Veretennikov [27] extends these results by allowing a part of the diffusion to be degenerate.

In [16] existence and uniqueness of a solution for the case where the drift is discontinuous at a hyperplane, or a special hypersurface and where the diffusion coefficient is degenerate is proven, and in [23] it is shown how these results extend to the non-homogeneous case.

Currently, research on numerical methods for SDEs with irregular coefficients is highly active. Hutzenthaler et al. [8] introduce the tamed Euler–Maruyama scheme and prove strong order $1/2$ convergence for SDEs with continuously differentiable and polynomially growing drift that satisfy a one-sided Lipschitz condition. Sabanis [22] proves strong convergence of the tamed Euler–Maruyama scheme from a different perspective and also considers the case of locally Lipschitz diffusion coefficient. Gyöngy [4] proves almost sure convergence of the Euler–Maruyama scheme for the case where the drift satisfies a monotonicity condition.

Halidias and Kloeden [7] show that the Euler–Maruyama scheme converges strongly for SDEs with a discontinuous monotone drift coefficient. Kohatsu-Higa et al. [13] show weak convergence with rates smaller than 1 of a method where they first regularize the discontinuous drift and then apply the Euler–Maruyama scheme.

Étoré and Martinez [1,2] introduce an exact simulation algorithm for one-dimensional SDEs with a drift coefficient which is discontinuous in one point, but differentiable everywhere else. For one-dimensional SDEs with piecewise Lipschitz drift and possibly degenerate diffusion coefficient, in [14] an existence and uniqueness result is proven, and a numerical method, which is based on applying the Euler–Maruyama scheme to a transformation of (1), is presented. This method converges with strong order 1/2. In [15] a (non-trivial) extension of the method is introduced, which converges with strong order 1/2 also in the multidimensional case. The paper also contains an existence and uniqueness result for the multidimensional setting under more general conditions than, e.g., the ones stated in [16].

The method introduced in [15] is the first numerical method that is proven to converge with positive strong rate for multidimensional SDEs with discontinuous drift and degenerate diffusion coefficient. It requires application of a transformation and its numerical inverse in each step, which makes the method rather slow in practice. Furthermore, the method requires specific inputs about the geometry of the discontinuity of the drift to calculate this transformation. This is a drawback, if, e.g., the method shall be applied for solving control problems, since the control is usually not explicitly known. So a method is preferred that can deal with the discontinuities in the drift automatically.

First results in this direction are contained in a series of papers by Ngo and Taguchi. In [21] they show convergence of order up to 1/4 of the Euler–Maruyama method for multidimensional SDEs with discontinuous bounded drift that satisfies a one-sided Lipschitz condition and with Hölder continuous, bounded, and uniformly non-degenerate diffusion coefficient. In [19] they extend this result to cases where the drift is not necessarily one-sided Lipschitz for one-dimensional SDEs, and in [20] they extend the result for one-dimensional SDEs by allowing for discontinuities also in the diffusion coefficient. For many applications, their results fail to be applicable, since they only hold for one-dimensional SDEs and their method of proof relies on uniform non-degeneracy of the diffusion coefficient.

Contrasting the above, there are several delimiting results which state that even equations with infinitely often differentiable coefficients cannot always be solved approximately in finite time, even if the Euler–Maruyama method converges, cf. Hairer et al. [6], Jentzen et al. [10], Müller-Gronbah and Yaroslavtseva [18], Yaroslavtseva [28]. However, there is still a big gap between the assumptions on the coefficients under which convergence with strong convergence rate has been proven and the properties of the coefficients of the equation presented in [6].

In this paper we prove strong convergence of order $1/4 - \epsilon$ for arbitrarily small $\epsilon > 0$ of the Euler–Maruyama method for multidimensional SDEs with discontinuous drift satisfying a piecewise Lipschitz condition and with a degenerate diffusion coefficient. Note that we do not impose a one-sided Lipschitz condition on the drift. So even for SDEs with non-degenerate diffusion coefficient, which do not have a one-sided Lipschitz drift, this result is novel.

Our convergence proof is based on estimating the difference between the Euler–Maruyama scheme and the scheme presented in [15]. Close to the set of discontinuities of the drift, we have no tight estimate of this difference, so we need to study the occupation time of an Itô process with degenerate diffusion coefficient there. Away

from the set of discontinuities, it is essential to estimate the probability that during one step the distance between the interpolation of the Euler–Maruyama method and the previous Euler–Maruyama step becomes greater than some threshold.

This paper's result is the first one that gives strong convergence and also a strong convergence rate of a fully explicit scheme for multidimensional SDEs with discontinuous drift and degenerate diffusion coefficient, and the first one for multidimensional SDEs with discontinuous drift that does not satisfy a one-sided Lipschitz condition.

2 Preliminaries

In this section we first state the assumptions on the coefficients of SDE (1), under which the result of this paper is proven, then we study the occupation time of an Itô process close to a hypersurface, and finally we recall the transformation from [15], which is also essential for our proof.

2.1 Definitions and assumptions

We want to prove strong convergence of the Euler–Maruyama method for SDEs with discontinuous drift coefficient. Instead of the usual requirement of Lipschitz continuity we only assume that the drift is a piecewise Lipschitz function on the $\mathbb{R}^d$.

Definition 2.1 ([15, Definitions 3.1 and 3.2]) Let $A \subseteq \mathbb{R}^d$.

1. For a continuous curve $\gamma : [0, 1] \longrightarrow \mathbb{R}^d$, let $\ell(\gamma)$ denote its length,

$$\ell(\gamma) = \sup_{n, 0 \leq t_1 < \cdots < t_n \leq 1} \sum_{k=1}^{n} \|\gamma(t_k) - \gamma(t_{k-1})\|.$$

The *intrinsic metric* ρ on A is given by

$$\rho(x, y) := \inf\{\ell(\gamma) : \gamma : [0, 1] \longrightarrow A \text{ is a continuous curve satisfying } \gamma(0) = x, \ \gamma(1) = y\},$$

where $\rho(x, y) := \infty$, if there is no continuous curve from x to y.

2. Let $f : A \longrightarrow \mathbb{R}^m$ be a function. We say that f is *intrinsic Lipschitz*, if it is Lipschitz w.r.t. the intrinsic metric on A, i.e. if there exists a constant L such that

$$\forall x, y \in A : \|f(x) - f(y)\| \leq L\rho(x, y).$$

The prototypical examples for intrinsic Lipschitz function are given, like in the one-dimensional case, by differentiable functions with bounded derivative.

Lemma 2.2 ([15, Lemma 3.8]) *Let $A \subseteq \mathbb{R}^d$ be open and let $f : A \longrightarrow \mathbb{R}^m$ be a differentiable function with $\|f'\| < \infty$. Then f is intrinsic Lipschitz with Lipschitz constant $\|f'\|$.*

Definition 2.3 ([15, Definition 3.4]) A function $f: \mathbb{R}^d \longrightarrow \mathbb{R}^m$ is *piecewise Lipschitz*, if there exists a hypersurface Θ with finitely many connected components and with the property, that the restriction $f|_{\mathbb{R}^d \setminus \Theta}$ is intrinsic Lipschitz. We call Θ an *exceptional set* for f, and we call

$$\sup_{x, y \in \mathbb{R}^d \setminus \Theta} \frac{\|f(x) - f(y)\|}{\rho(x, y)}$$

the *piecewise Lipschitz constant* of f.

In this paper Θ will be a fixed C^3-hypersurface, and we will only consider piecewise Lipschitz functions with exceptional set Θ. In the following, L_f denotes the piecewise Lipschitz constant of a function f, if f is piecewise Lipschitz, and it denotes the Lipschitz constant, if f is Lipschitz.

We define the distance $d(x, \Theta)$ between a point x and the hypersurface Θ by $d(x, \Theta) := \inf\{\|x - y\| : y \in \Theta\}$, and for every $\varepsilon > 0$ we define $\Theta^\varepsilon := \{x \in \mathbb{R}^d : d(x, \Theta) < \varepsilon\}$.

Recall that, since $\Theta \in C^3$, for every $\xi \in \Theta$ there exists an open environment $U \subseteq \Theta$ of ξ and a continuously differentiable function $n: U \longrightarrow \mathbb{R}^d$ such that for every $\zeta \in U$ the vector $n(\zeta)$ has length 1 and is orthogonal to the tangent space of Θ in ζ. On a given connected open subset of Θ the local unit normal vector n is unique up to a factor ± 1.

We recall a definition from differential geometry.

Definition 2.4 Let $\Theta \in \mathbb{R}^d$ be any set.

1. An environment Θ^ε is said to have the *unique closest point property*, if for every $x \in \mathbb{R}^d$ with $d(x, \Theta) < \varepsilon$ there is a unique $p \in \Theta$ with $d(x, \Theta) = \|x - p\|$. Therefore, we can define a mapping $p: \Theta^\varepsilon \longrightarrow \Theta$ assigning to each x the point $p(x)$ in Θ closest to x.
2. Θ is said to be of *positive reach*, if there exists $\varepsilon > 0$ such that Θ^ε has the unique closest point property. The *reach* of Θ is the supremum over all such ε if such an ε exists, and 0 otherwise.

Now, we give assumptions which are sufficient for the results in [15] to hold and which we need to prove the main result here.

Assumption 2.1 We assume the following for the coefficients of (1):

1. μ and σ are bounded;
2. the diffusion coefficient σ is Lipschitz;
3. the drift coefficient μ is a piecewise Lipschitz function $\mathbb{R}^d \longrightarrow \mathbb{R}^d$. Its exceptional set Θ is a C^3-hypersurface of positive reach;
4. *non-parallelity condition:* there exists a constant $c_0 > 0$ such that $\|\sigma(\xi)^\top n(\xi)\| \geq c_0$ for all $\xi \in \Theta$;
5. the function $\alpha: \Theta \longrightarrow \mathbb{R}^d$ defined by

$$\alpha(\xi) := \lim_{h \to 0} \frac{\mu(\xi - hn(\xi)) - \mu(\xi + hn(\xi))}{2\|\sigma(\xi)^\top n(\xi)\|^2} \tag{3}$$

is C^3 and all derivatives up to order three are bounded.

Theorem 2.5 ([15, Theorem 3.21]) *Let Assumption* 2.1 *hold. Then SDE* (1) *has a unique strong solution.*

Remark on Assumption 2.1:

1. For existence and uniqueness of a solution to (1), in [15, Theorem 3.21] instead of Assumption 2.1.1 only boundedness in an ε-environment of Θ is needed. However, for the proof of our convergence result we require global boundedness. Note that other results in the literature on numerical methods for SDEs with discontinuous drift also rely on boundedness of the coefficients, cf. [19–21].
2. Assumption 2.1.2 is a technical condition; the focus in this paper is on other types of irregularities in the coefficients. There are results in the literature, where the authors deal with a non-globally Lipschitz diffusion coefficient, see, e.g., [5], but in contributions where only Hölder continuity is required for σ, usually uniform non-degeneracy is assumed.
3. Assumption 2.1.3 is a geometrical condition which we require in order to locally flatten Θ, i.e. to map Θ to a hyperplane in a regular way. This is crucial in many places in [15] and here, in particular for the proof of Theorem 2.7 below. In addition to that, Assumption 2.1.3 implies that there exists a constant c_1 such that $\|n'(\xi)\| \leq c_1$ for every $\xi \in \Theta$ and every orthonormal vector n on Θ, see [15, Lemma 3.10].
4. Assumption 2.1.4 means that the diffusion coefficient must have a component orthogonal to Θ in all $\xi \in \Theta$. This condition is significantly weaker than uniform non-degeneracy, and it is essential: in [16] we give a counterexample for the case where the non-parallelity condition does not hold. Then, even existence of a solution is not guaranteed.
5. Assumption 2.1.5 is a technical condition, which is required for our transformation method to work. Boundedness of α and α' is needed for proving the local invertibility of our transform. Existence and boundedness of α'' and α''' is required for the multidimensional version of Itô's formula to hold for the transform, see [15]. Moreover, it has been shown in [15, Proposition 3.13] that α is a well-defined function on Θ, i.e. it does not depend on the choice of the normal vector n and, in particular, on its sign.

Example 2.6 Suppose Θ is the finite and disjoint union of orientable compact C^3-manifolds. Then Θ is of positive reach by the lemma in [3], and each connected component of Θ separates the $\mathbb{R}^n$ into two open connected components by the Jordan–Brouwer separation theorem, see [17].

Thus $\mathbb{R}^d \setminus \Theta$ is the union of finitely many disjoint open connected subsets of $\mathbb{R}^d$; we can write $\mathbb{R}^d \setminus \Theta = A_1 \cup \cdots \cup A_n$.

Suppose there exist bounded and Lipschitz C^3-functions $\mu_1, \ldots, \mu_n : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ such that $\mu = \sum_{k=1}^n \mathbf{1}_{A_k} \mu_k$, and suppose that $\sigma : \mathbb{R}^d \longrightarrow \mathbb{R}^{d \times d}$ is bounded, Lipschitz, and C^3 with $\sigma(\xi)^\top n(\xi) \neq 0$ for every $\xi \in \Theta$.

Then it is readily checked that μ and σ satisfy Assumption 2.1.

In Sect. 4 we present a number of concrete examples which satisfy Assumption 2.1 and we perform numerical tests on the associated SDEs.

2.2 Occupation time close to a hypersurface

In this section we study the occupation time of an Itô process close to a C^3-hypersurface. In the proof of our main theorem, the Euler–Maruyama approximation X^δ in equation (2) will play the role of that Itô process.

Theorem 2.7 *Let Θ be a C^3-hypersurface of positive reach and let $\varepsilon_0 > 0$ be such that the closure of Θ^{ε_0} has the unique closest point property. Let further $X = (X_t)_{t \geq 0}$ be an $\mathbb{R}^d$-valued Itô process*

$$X_t = X_0 + \int_0^t A_s ds + \int_0^t B_s dW_s \,,$$

with progressively measurable processes $A = (A_t)_{t \geq 0}$, $B = (B_t)_{t \geq 0}$, where A is $\mathbb{R}^d$-valued and B is $\mathbb{R}^{d \times d}$-valued. Let the coefficients A, B be such that

1. *there exists a constant c_{AB} such that for almost all $\omega \in \Omega$ it holds that*

$$\forall t \in [0, T] : X_t(\omega) \in \Theta^{\varepsilon_0} \implies \max(\|A_t(\omega)\|, \|B_t(\omega)\|) \leq c_{AB} \,;$$

2. *there exists a constant c_0 such that for almost all $\omega \in \Omega$ it holds that*

$$\forall t \in [0, T] : X_t(\omega) \in \Theta^{\varepsilon_0} \implies n(p(X_t(\omega)))^\top B_t(\omega) B_t(\omega)^\top n(p(X_t(\omega))) \geq c_0 \,.$$

Then there exists a constant C such that for all $0 < \varepsilon < \varepsilon_0/2$,

$$\int_0^T \mathbb{P}\left(\{X_s \in \Theta^\varepsilon\}\right) ds \leq C\varepsilon \,.$$

For the proof we will construct a one-dimensional Itô process Y with the property that Y is close to 0, if and only if X is close to Θ. For the construction of Y we decompose the path of X into pieces close to Θ and pieces farther away. These pieces are then mapped to $\mathbb{R}$ by using a signed distance of X from Θ and pasted together in a continuous way.

A signed distance to Θ is locally given by $D(x) := n(p(x))^\top (x - p(x))$, where n is a local unit normal vector.

Lemma 2.8 *For all $x \in \Theta^{\varepsilon_0}$ it holds that $D'(x) = n(p(x))^\top$.*

Proof Fix $x \in \Theta^{\varepsilon_0} \backslash \Theta$ and consider the function h defined by $h(b) := \|x - p(x+b)\|^2$. By definition of the projection map p, h has a minimum in $b = 0$, such that $h'(0) = 0$. Hence from $h'(b) = -2(x - p(x+b))^\top p'(x+b)$, we get $(x - p(x))^\top p'(x) = 0$. This implies $n(p(x))^\top p'(x) = 0$, since $(x - p(x))$ is a scalar multiple of $n(p(x))$.

Using that $D(x) = a\|x - p(x)\|$ for an $a \in \{-1, 1\}$, we compute

$$D'(x) = a\|x - p(x)\|^{-1}(x - p(x))^\top (\mathrm{id}_{\mathbb{R}^d} - p'(x)) = a\,n(p(x))^\top (\mathrm{id}_{\mathbb{R}^d} - p'(x))$$
$$= a\big(n(p(x))^\top - n(p(x))^\top p'(x)\big) = an(p(x))^\top \,. \tag{4}$$

For $\psi \in \mathbb{R}$ with $|\psi|$ small we get

$$D(x + \psi n(p(x))) = n\Big(p\big(x + \psi n(p(x))\big)\Big)^{\top}\Big(x + \psi n(p(x)) - p\big(x + \psi n(p(x))\big)\Big)$$
$$= n(p(x))^{\top}(x + \psi n(p(x)) - p(x)) = D(x) + \psi ,$$

such that the directional derivative of D in direction $n(p(x))$ in x is 1. From this and from (4) it follows that $D'(x) = n(p(x))^{\top}$. This also holds for $x \in \Theta$ by the continuity of D'.   □

The following lemma states that for any continuous curve γ in Θ^{ε_0} there is a continuous path of unit normal vectors, such that to every point of γ we can assign a signed distance in a continuous way.

Lemma 2.9 *Let $\gamma : [a, b] \longrightarrow \Theta^{\varepsilon_0}$ be a continuous function. Then there exists $m : [a, b] \longrightarrow \mathbb{R}^d$ such that*

1. *m is continuous;*
2. *$\|m(t)\| = 1$ for all $t \in [a, b]$;*
3. *$m(t)$ is orthogonal to Θ in the point $p(\gamma(t))$ for all $t \in [a, b]$.*

Proof For $\xi \in \Theta$ we denote the tangent space to Θ in ξ by tang_{ξ}. Let

$$S := \{a \leq s \leq b : \exists m : [a, s] \longrightarrow \mathbb{R}^d \text{ continuous}, \ \|m(t)\|$$
$$= 1, \ m(t) \bot \text{tang}_{p(\gamma(t))} \ \forall t \in [a, s]\}.$$

The set S is nonempty and its elements are bounded by b. Let $s_1 := \sup S$. There exists an open and connected subset $U \subseteq \Theta$ such that $p(\gamma(s_1)) \in U$, and a unit normal vector $n_1 : U \longrightarrow \mathbb{R}^d$.

Since U is open and $p \circ \gamma$ is continuous, there exists $\eta > 0$ such that $p(\gamma([s_1 - \eta, s_1])) \subseteq U$. By the definition of s_1 there exists $s \in (s_1 - \eta, s_1)$ and $m : [a, s] \longrightarrow \mathbb{R}^d$ continuous, with $\|m(t)\| = 1$ and $m(t) \bot \text{tang}_{p(\gamma(t))}$ for all $t \in [a, s]$.

Since n_1 is unique up to a factor ± 1, the mapping $n_1 \circ p \circ \gamma$ either coincides with m or $-m$ on $(s_1 - \eta, s)$. Without loss of generality we may assume that the former is the case. Thus we can extend m continuously to $[a, s_1]$ by defining $m(t) := n_1(p(\gamma(t)))$ for all $t \in (s, s_1]$.

Now, if s_1 was strictly smaller than b, then we could use the same mapping $n_1 \circ p \circ \gamma$ to extend m continuously beyond s_1, contradicting the definition of s_1.   □

We will need the following estimate on the local time of a one-dimensional Itô process.

Lemma 2.10 *Let $Y = (Y_t)_{t \geq 0}$ be an Itô process with bounded and progressively measurable coefficients $\hat{A} = (\hat{A}_t)_{t \geq 0}$, $\hat{B} = (\hat{B}_t)_{t \geq 0}$.*

Then $\sup_{y \in \mathbb{R}} \mathbb{E}(L_T^y(Y)) \leq \left(3T^2 \|\hat{A}\|_{\infty}^2 + \frac{3}{2}T \|\hat{B}_s\|_{\infty}^2\right)^{1/2}$.

The claim is a special case of [19, Lemma 3.2]. We give a proof for the convenience of the reader.

Proof From the Meyer–Tanaka formula [11, Section 3.7, Eq. (7.9)] we have

$$
\begin{aligned}
2L_T^y(Y) &= |Y_T - y| - |Y_0 - y| - \int_0^T \left(\mathbf{1}_{\{Y_s > y\}} - \mathbf{1}_{\{Y_s < y\}}\right) dY_s \\
&\leq |Y_T - Y_0| + \left|\int_0^T \left(\mathbf{1}_{\{Y_s > y\}} - \mathbf{1}_{\{Y_s < y\}}\right) dY_s\right| \\
&\leq \left|\int_0^T \hat{A}_s ds\right| + \left|\int_0^T \hat{B}_s dW_s\right| + \left|\int_0^T \left(\mathbf{1}_{\{Y_s > y\}} - \mathbf{1}_{\{Y_s < y\}}\right) \hat{A}_s ds\right| \\
&\quad + \left|\int_0^T \left(\mathbf{1}_{\{Y_s > y\}} - \mathbf{1}_{\{Y_s < y\}}\right) \hat{B}_s dW_s\right| \\
&\leq 2\int_0^T |\hat{A}_s| ds + \left|\int_0^T \hat{B}_s dW_s\right| + \left|\int_0^T \left(\mathbf{1}_{\{Y_s > y\}} - \mathbf{1}_{\{Y_s < y\}}\right) \hat{B}_s dW_s\right|.
\end{aligned}
$$

Using the inequality $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$ we get

$$
4L_T^y(Y)^2 \leq 12\|\hat{A}\|_\infty^2 T^2 + 3\left|\int_0^T \hat{B}_s dW_s\right|^2 + 3\left|\int_0^T \left(\mathbf{1}_{\{Y_s > y\}} - \mathbf{1}_{\{Y_s < y\}}\right) \hat{B}_s dW_s\right|^2,
$$

and, using Itô's L^2-isometry,

$$
4\mathbb{E}\left(L_T^y(Y)^2\right) \leq 12\|\hat{A}\|_\infty^2 T^2 + 6\int_0^T \hat{B}_s^2 ds \leq 12\|\hat{A}\|_\infty^2 T^2 + 6T\|\hat{B}\|_\infty^2.
$$

The claim now follows by applying the Cauchy–Schwarz-inequality and taking the supremum over all $y \in \mathbb{R}$. $\qquad\square$

We are ready to prove the result of this section.

Proof of Theorem 2.7 Let $\varepsilon_1 = \varepsilon_0/2$. Define a mapping $\lambda : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$
\lambda(z) = \begin{cases} z - \frac{2}{3\varepsilon_1^2}z^3 + \frac{1}{5\varepsilon_1^4}z^5 & |z| \leq \varepsilon_1 \\ \frac{8\varepsilon_1}{15} & z > \varepsilon_1 \\ -\frac{8\varepsilon_1}{15} & z < -\varepsilon_1. \end{cases}
$$

Note that $\lambda'(0) = 1$ and $\lambda'(\pm\varepsilon_1) = \lambda''(\pm\varepsilon_1) = 0$, so that $\lambda \in C^2$.

Next we decompose the path of X: let $\tau_0 := \inf\{t \geq 0 : X_t \in \Theta^{\varepsilon_1}\}$. In particular we have $\tau_0 = 0$, if $X_0 \in \Theta^{\varepsilon_1}$. For $k \in \mathbb{N}_0$, define

$$
\begin{aligned}
\kappa_{k+1} &:= \inf\{t \geq \tau_k : X_t \notin \Theta^{2\varepsilon_1}\} \wedge T, \\
\tau_{k+1} &:= \inf\{t \geq \kappa_{k+1} : X_t \in \Theta^{\varepsilon_1}\} \wedge T.
\end{aligned}
$$

By Lemma 2.9 there exist continuous $m_k : [\tau_k, \kappa_{k+1}] \longrightarrow \mathbb{R}^d$, with $\|m_k(t)\| = 1$ and $m_k(t) \perp \mathrm{tang}_{p(X_t)}$ for all $t \in [\tau_k, \kappa_{k+1}]$. Without loss of generality m_0 can be chosen

such that $m_0(\tau_0)^\top(X_{\tau_0} - p(X_{\tau_0})) \geq 0$. We construct a one-dimensional process Y as follows:

$$Y_t = \begin{cases} \lambda(m_0(\tau_0)^\top(X_{\tau_0} - p(X_{\tau_0}))) & t \leq \tau_0 \\ \lambda(m_k(t)^\top(X_t - p(X_t))) & t \in [\tau_k, \kappa_{k+1}] \\ \lambda(m_k(\kappa_k)^\top(X_{\kappa_k} - p(X_{\kappa_k}))) & t \in [\kappa_k, \tau_k], \end{cases}$$

where without loss of generality the m_k are chosen such that

$$\lambda(m_{k+1}(\tau_{k+1})^\top(X_{\tau_{k+1}} - p(X_{\tau_{k+1}}))) = \lambda(m_k(\kappa_k)^\top(X_{\kappa_k} - p(X_{\kappa_k}))) . \qquad (5)$$

Note that by construction both sides of (5) can only take the values $\pm\lambda(\varepsilon_1)$.

We have thus constructed a continuous $[\lambda(-\varepsilon_1), \lambda(\varepsilon_1)]$-valued process Y with the property that the occupation time of Y in an environment of 0 is the same as the occupation time of X in an environment of Θ, i.e. $Y \in (-\lambda(\varepsilon), \lambda(\varepsilon))$, iff $X \in \Theta^\varepsilon$ for all $0 < \varepsilon < \varepsilon_1$.

To show that Y is an Itô process, we want to use Itô's formula. For this we recognize that Y, depending on its proximity to Θ, is either constant or locally of the form $Y_t = \lambda(n(p(X_t))^\top(X_t - p(X_t)))$ for a suitable choice of the unit normal vector. Denote $D(x) = n(p(x))^\top(x - p(x))$. The function D is locally a signed distance to Θ and $D \in C^2$. This can be seen by following the proof of [3, Theorem 1]. Hence, we may apply Itô's formula to get

$$dY_t = \lambda'(D(X_t))D'(X_t)A_t dt + \lambda'(D(X_t))D'(X_t)B_t dW_t + \frac{1}{2}\,\mathrm{tr}\left(B_t^\top\lambda''(D(X_t))B_t\right)dt .$$

By Lemma 2.8 we have $D'(x) = n(p(x))^\top$, and hence

$$(\lambda(D(x)))'' = (\lambda'(D(x))n(p(x))^\top)' = \lambda''(D(x))n(p(x))n(p(x))^\top + \lambda'(D(x))n'(p(x)) .$$

Since λ' and λ'' are bounded by construction, $\|n(p(x))n(p(x))^\top\| = 1$, $\|n'\|$ is bounded (c.f. the remark on Assumption 2.1.3), and by Assumption 1 of the theorem, the coefficients of Y are uniformly bounded. Therefore $dY_t = \hat{A}_t dt + \hat{B}_t dW_t$, with bounded and progressively measurable $\hat{A}, \hat{B}$.

Let $0 < \varepsilon \leq \varepsilon_1/2$. For all $|z| \leq \varepsilon$, we have $\lambda'(z) \geq \left(\frac{3}{4}\right)^2$. Thus by Assumption 2 of the theorem,

$$\left(\frac{3}{4}\right)^2 c_0^2 \int_0^t \mathbf{1}_{\{X_s \in \Theta^\varepsilon\}}ds = \left(\frac{3}{4}\right)^2 c_0^2 \int_0^t \mathbf{1}_{\{Y_s \in (-\lambda(\varepsilon), \lambda(\varepsilon))\}}ds$$

$$\leq \int_0^t \mathbf{1}_{\{Y_s \in (-\lambda(\varepsilon), \lambda(\varepsilon))\}}\lambda'(D(X_s))^2 n(p(X_s))^\top B_s B_s^\top n(p(X_s))ds$$

$$= \int_0^t \mathbf{1}_{\{Y_s \in (-\lambda(\varepsilon), \lambda(\varepsilon))\}}d[Y]_s .$$

By the occupation time formula [11, Chapter 3, 7.1 Theorem] for one-dimensional continuous semimartingales, we get

$$\int_0^T \mathbb{P}\left(\{X_s \in \Theta^\varepsilon\}\right) ds \leq \left(\frac{4}{3c_0}\right)^2 \mathbb{E}\left(\int_0^T \mathbf{1}_{\{Y_s \in (-\lambda(\varepsilon), \lambda(\varepsilon))\}} d\,[Y]_s\right)$$

$$= 2\left(\frac{4}{3c_0}\right)^2 \mathbb{E}\left(\int_{\mathbb{R}} \mathbf{1}_{(-\lambda(\varepsilon), \lambda(\varepsilon))}(y) L_T^y(Y)\, dy\right)$$

$$\leq \frac{4^3}{3^2 c_0^2} \sup_{y \in \mathbb{R}} \mathbb{E}\left(L_T^y(Y)\right) \varepsilon .$$

$\square$

2.3 The transformation

The proof of convergence is based on a transformation that removes the discontinuity from the drift and makes the drift Lipschitz while preserving the Lipschitz property of the diffusion coefficient. A suitable transform is presented in [15]. We recall it here.

Define $G : \mathbb{R}^d \longrightarrow \mathbb{R}^d$,

$$G(x) = \begin{cases} x + \tilde{\phi}(x)\alpha(p(x)) & x \in \Theta^{\varepsilon_0} \\ x & x \in \mathbb{R}^d \setminus \Theta^{\varepsilon_0} , \end{cases}$$

where $\varepsilon_0 > 0$ is smaller than the reach of Θ, see Assumption 2.1.3, α is the function defined in Assumption 2.1.5, and

$$\tilde{\phi}(x) = n(p(x))^\top (x - p(x))\|x - p(x)\|\phi\left(\frac{\|x - p(x)\|}{c}\right) ,$$

with positive constant c and

$$\phi(u) = \begin{cases} (1+u)^3(1-u)^3 & |u| \leq 1 \\ 0 & |u| > 1. \end{cases}$$

Note that G'' is piecewise Lipschitz with exceptional set Θ.

If c is chosen sufficiently small, see [15, Lemma 3.18], G is invertible by [15, Theorem 3.14]. Furthermore, Itô's formula holds for G and G^{-1} by [15, Theorem 3.19].

With this we can define a process $Z = (Z_t)_{t \geq 0}$ by $Z_t = G(X_t)$, which solves the SDE

$$dZ_t = \tilde{\mu}(Z_t)dt + \tilde{\sigma}(Z_t)dW_t , \tag{6}$$

where

$$\tilde{\mu}(z) = G'(G^{-1}(z))\mu(G^{-1}(z)) + \frac{1}{2}\,\mathrm{tr}\left(\sigma(G^{-1}(z))^\top G''(G^{-1}(z))\sigma(G^{-1}(z))\right) ,$$

$$\tilde{\sigma}(z) = G'(G^{-1}(z))\sigma(G^{-1}(z)) .$$

From [15, Theorem 3.20] we know that $\tilde{\mu}$ and $\tilde{\sigma}$ are Lipschitz, and hence the solution to (6) can be approximated with strong order $1/2$ using the Euler–Maruyama scheme.

3 Main result

We are ready to formulate the main result.

Theorem 3.1 *Let Assumption* 2.1 *hold. Then the Euler–Maruyama method* (2) *converges to the solution of SDE* (1) *with strong order* $1/4 - \epsilon$ *for arbitrarily small* $\epsilon > 0$, *i.e. there exists a constant C such that for all* $\epsilon > 0$ *it holds that for sufficiently small step size* $\delta > 0$,

$$\mathbb{E}\left(\sup_{t\in[0,T]} \|X_t - X_t^\delta\|^2\right)^{1/2} \leq C\delta^{1/4-\epsilon}.$$

In preparation of the proof of the main result, we prove two lemmas.

Lemma 3.2 *Let Assumption* 2.1.1 *hold. Then there exists a constant C such that for sufficiently small step size* δ

$$\mathbb{E}\left(\int_0^T \|X_s^\delta - X_{\underline{s}}^\delta\|^2 ds\right) \leq C\delta.$$

Proof By the definition of the Euler–Maruyama method (2) we have

$$\mathbb{E}\left(\int_0^T \|X_{\underline{s}}^\delta - X_s^\delta\|^2 ds\right) = \sum_{j=0}^{T/\delta-1} \mathbb{E}\left(\int_{j\delta}^{(j+1)\delta} \|X_{j\delta}^\delta - X_s^\delta\|^2 ds\right)$$

$$\leq \frac{T}{\delta} \sup_{t\in\{j\delta:j=0,\dots,T/\delta-1\}} \mathbb{E}\left(\int_t^{t+\delta} \|X_t^\delta - X_s^\delta\|^2 ds\right)$$

$$= \frac{T}{\delta} \sup_{t\in\{j\delta:j=0,\dots,T/\delta-1\}} \mathbb{E}\left(\int_t^{t+\delta} \|X_t^\delta - X_t^\delta - \mu(X_t^\delta)\delta - \sigma(X_t^\delta)(W_s - W_t)\|^2 ds\right)$$

$$\leq \frac{2T}{\delta} \sup_{t\in\{j\delta:j=0,\dots,T/\delta-1\}} \mathbb{E}\left(\int_t^{t+\delta} \|\mu(X_t^\delta)\delta\|^2 ds + \int_t^{t+\delta} \|\sigma(X_t^\delta)(W_s - W_t)\|^2 ds\right)$$

$$\leq \frac{2T}{\delta}\left(\|\mu\|_\infty^2 \delta^3 + \|\sigma\|_\infty^2 \sup_{t\in\{j\delta:j=0,\dots,(T-\delta)/\delta\}} \int_t^{t+\delta} \mathbb{E}(\|W_s - W_t\|^2) ds\right)$$

$$= \frac{2T}{\delta}\left(\|\mu\|_\infty^2 \delta^3 + d\|\sigma\|_\infty^2 \sup_{t\in\{j\delta:j=0,\dots,(T-\delta)/\delta\}} \int_t^{t+\delta} (s-t) ds\right)$$

$$= \frac{2T}{\delta}\left(\|\mu\|_\infty^2 \delta^3 + \frac{d}{2}\|\sigma\|_\infty^2 \delta^2\right) \leq C\delta.$$

$\square$

For all $\delta, \varepsilon > 0$ and all $j = 0, \ldots, T/\delta - 1$, define

$$\Omega_{\delta,\varepsilon,j} := \left\{ \omega \in \Omega : \sup_{s \in [j\delta, (j+1)\delta]} \left\| X_s^\delta(\omega) - X_{\underline{s}}^\delta(\omega) \right\| \geq \varepsilon \right\}. \tag{7}$$

Lemma 3.3 *Let Assumption 2.1.1 hold. Then there exists a constant C such that for all $0 < \delta \leq 1$, all $\varepsilon > 0$, and all $j = 0, \ldots, T/\delta - 1$, it holds that $\mathbb{P}(\Omega_{\delta,\varepsilon,j}) \leq C \exp(-\varepsilon/\|\sigma\|_\infty \delta^{1/2})$.*

Proof

$$\mathbb{P}\left(\sup_{j\delta \leq s \leq (j+1)\delta} \|X_s^\delta - X_{\underline{s}}^\delta\| \geq \varepsilon \right) = \mathbb{P}\left(\sup_{j\delta \leq s \leq (j+1)\delta} \|\mu(X_{\underline{s}}^\delta)(s - \underline{s}) + \sigma(X_{\underline{s}}^\delta)(W_s - W_{\underline{s}})\| \geq \varepsilon \right)$$

$$\leq \mathbb{P}\left(\sup_{j\delta \leq s \leq (j+1)\delta} \left\{ \|\mu\|_\infty \delta + \|\sigma\|_\infty \|W_s - W_{\underline{s}}\| \right\} \geq \varepsilon \right) = \mathbb{P}\left(\sup_{j\delta \leq s \leq (j+1)\delta} \|W_s - W_{\underline{s}}\| \geq \frac{\varepsilon - \|\mu\|_\infty \delta}{\|\sigma\|_\infty} \right)$$

$$= \mathbb{P}\left(\sup_{0 \leq s \leq 1} \|W_s - W_0\| \geq \frac{\varepsilon - \|\mu\|_\infty \delta}{\|\sigma\|_\infty \delta^{1/2}} \right) = \mathbb{P}\left(\exp\left(\sup_{0 \leq s \leq 1} \|W_s\| \right) \geq \exp\left(\frac{\varepsilon - \|\mu\|_\infty \delta}{\|\sigma\|_\infty \delta^{1/2}} \right) \right)$$

$$\leq \mathbb{E}\left(\exp(\|W_1\|) \right) \exp\left(\frac{\|\mu\|_\infty \delta - \varepsilon}{\|\sigma\|_\infty \delta^{1/2}} \right) \leq C \exp\left(-\frac{\varepsilon}{\|\sigma\|_\infty \delta^{1/2}} \right),$$

where we applied Doob's submartingal inequality, and in the last step used that $\delta \leq 1$.
$\square$

Now, we are ready to prove our main result.

Proof of Theorem 3.1 Since G^{-1} is Lipschitz by the proof of [15, Theorem 3.20],

$$\mathbb{E}\left(\sup_{0 \leq t \leq T} \|X_t - X_t^\delta\|^2 \right)^{1/2} \leq L_{G^{-1}} \mathbb{E}\left(\sup_{0 \leq t \leq T} \|Z_t - G(X_t^\delta)\|^2 \right)^{1/2}, \tag{8}$$

with $Z = G(X)$ as in (6). Let Z^δ be the Euler–Maruyama approximation of Z. It holds that

$$\mathbb{E}\left(\sup_{0 \leq t \leq T} \|Z_t - G(X_t^\delta)\|^2 \right)^{1/2} \leq \mathbb{E}\left(\sup_{0 \leq t \leq T} \|Z_t - Z_t^\delta\|^2 \right)^{1/2}$$

$$+ \mathbb{E}\left(\sup_{0 \leq t \leq T} \|Z_t^\delta - G(X_t^\delta)\|^2 \right)^{1/2}. \tag{9}$$

For estimating the first term in (9), recall that by [15, Theorem 3.20], the transformed SDE (6) has Lipschitz coefficients. Since the Euler–Maruyama method converges with strong order $1/2$ for SDEs with Lipschitz coefficients (see [12, Theorem 10.2.2]), there exists a constant $C_1 > 0$ such that for sufficiently small $\delta > 0$,

$$\mathbb{E}\left(\sup_{0 \leq t \leq T} \|Z_t - Z_t^\delta\|^2 \right) \leq C_1 \delta. \tag{10}$$

We now turn to the second term in (9), i.e. we estimate the difference between G applied to the Euler–Maruyama approximation of X and the Euler–Maruyama approximation of Z. Denote, for all $\tau \in [0, T]$,

$$u(\tau) := \mathbb{E} \left(\sup_{0 \leq t \leq \tau} \|G(X_t^\delta) - Z_t^\delta\|^2 \right).$$

With $\nu(x_1, x_2) = G'(x_1)\mu(x_2) + \frac{1}{2} \operatorname{tr}(\sigma(x_2)^\top G''(x_1)\sigma(x_2))$ we have by Itô's formula,

$$G(X_t^\delta) = G(X_0^\delta) + \int_0^t \nu(X_s^\delta, X_{\underline{s}}^\delta)ds + \int_0^t G'(X_s^\delta)\sigma(X_{\underline{s}}^\delta)dW_s,$$

so that

$$
\begin{aligned}
u(\tau) &= \mathbb{E} \left(\sup_{0 \leq t \leq \tau} \left\| \int_0^t \nu(X_s^\delta, X_{\underline{s}}^\delta)ds + \int_0^t G'(X_s^\delta)\sigma(X_{\underline{s}}^\delta)dW_s - \int_0^t \tilde{\mu}\left(Z_{\underline{s}}^\delta\right)ds \right. \right. \\
&\quad \left. \left. - \int_0^t \tilde{\sigma}\left(Z_{\underline{s}}^\delta\right)dW_s \right\|^2 \right) \\
&\leq \mathbb{E} \left(\sup_{0 \leq t \leq \tau} \left(4 \left\| \int_0^t \left(\nu(X_s^\delta, X_{\underline{s}}^\delta) - \nu(X_{\underline{s}}^\delta, X_{\underline{s}}^\delta) \right) ds \right\|^2 + 4 \left\| \int_0^t \left(G'(X_s^\delta)\sigma(X_{\underline{s}}^\delta) \right. \right. \right. \right. \\
&\quad \left. \left. - G'(X_{\underline{s}}^\delta)\sigma(X_{\underline{s}}^\delta) \right) dW_s \right\|^2 \\
&\quad \left. \left. + 4 \left\| \int_0^t \left(\tilde{\mu}(G(X_{\underline{s}}^\delta)) - \tilde{\mu}(Z_{\underline{s}}^\delta) \right) ds \right\|^2 + 4 \left\| \int_0^t \left(\tilde{\sigma}(G(X_{\underline{s}}^\delta)) - \tilde{\sigma}(Z_{\underline{s}}^\delta) \right) dW_s \right\|^2 \right) \right).
\end{aligned}
$$

Applying the Cauchy–Schwarz inequality to the Lebesgue integrals and the d-dimensional Burkholder–Davis–Gundy inequality [8, Lemma 3.7] to the Itô integrals, we obtain

$$
\begin{aligned}
u(\tau) &\leq 4T \,\mathbb{E} \left(\int_0^\tau \left\| \nu(X_s^\delta, X_{\underline{s}}^\delta) - \nu(X_{\underline{s}}^\delta, X_{\underline{s}}^\delta) \right\|^2 ds \right) \\
&\quad + 8d \,\mathbb{E} \left(\int_0^\tau \left\| G'(X_s^\delta)\sigma(X_{\underline{s}}^\delta) - G'(X_{\underline{s}}^\delta)\sigma(X_{\underline{s}}^\delta) \right\|^2 ds \right) \\
&\quad + 4T \,\mathbb{E} \left(\int_0^\tau \left\| \tilde{\mu}(G(X_{\underline{s}}^\delta)) - \tilde{\mu}(Z_{\underline{s}}^\delta) \right\|^2 ds \right) \\
&\quad + 8d \,\mathbb{E} \left(\int_0^\tau \left\| \tilde{\sigma}(G(X_{\underline{s}}^\delta)) - \tilde{\sigma}(Z_{\underline{s}}^\delta) \right\|^2 ds \right) \\
&=: 4T \, E_1 + 8d \, E_2 + 4T \, E_3 + 8d \, E_4.
\end{aligned}
\tag{11}
$$

For estimating E_1 in (11), we will use that

$$\|\nu(x_1, x_2) - \nu(x_2, x_2)\|^2 \leq \begin{cases} \left(2L_{G'}^2 \|\mu\|_\infty^2 + \frac{1}{2} L_{G''}^2 \|\sigma\|_\infty^4\right) \|x_1 - x_2\|^2 & x_1 \notin \Theta^\varepsilon, \|x_1 - x_2\| < \varepsilon \\ 4\|\mu\|_\infty^2 \|G'\|_\infty^2 + \|\sigma\|_\infty^4 \|G''\|_\infty^2 & \text{otherwise}. \end{cases}$$

With this and the definition of $\Omega_{\delta,\varepsilon,j}^c$ in (7), we get

$$E_1 = \int_0^\tau \mathbb{E}\left(\left\|\nu(X_s^\delta, X_{\underline{s}}^\delta) - \nu(X_{\underline{s}}^\delta, X_{\underline{s}}^\delta)\right\|^2 \left(\mathbf{1}_{\{X_s^\delta \notin \Theta^\varepsilon\}} \mathbf{1}_{\Omega_{\delta,\varepsilon,\underline{s}/\delta}^c} + \mathbf{1}_{\{X_s^\delta \notin \Theta^\varepsilon\}} \mathbf{1}_{\Omega_{\delta,\varepsilon,\underline{s}/\delta}} + \mathbf{1}_{\{X_s^\delta \in \Theta^\varepsilon\}}\right)\right) ds$$

$$\leq \left(2L_{G'}^2 \|\mu\|_\infty^2 + \frac{1}{2} L_{G''}^2 \|\sigma\|_\infty^4\right) \varepsilon^2 T$$

$$+ \left(4\|\mu\|_\infty^2 \|G'\|_\infty^2 + \|\sigma\|^4 \|G''\|_\infty^2\right) \left(\int_0^T \mathbb{P}(\Omega_{\delta,\varepsilon,\underline{s}/\delta}) ds + \int_0^T \mathbb{P}(\{X_s^\delta \in \Theta^\varepsilon\}) ds\right).$$

By Lemma 3.3, $\int_0^T \mathbb{P}(\Omega_{\delta,\varepsilon,\underline{s}/\delta}) ds \leq C_2 \exp(-\varepsilon/\|\sigma\|_\infty \delta^{1/2})$, and by Theorem 2.7, $\int_0^T \mathbb{P}(\{X_s^\delta \in \Theta^\varepsilon\}) ds \leq C_3 \varepsilon$, for suitable constants C_2, C_3. In order to minimize the bound on E_1, we choose ε such that $\exp(-\varepsilon/\|\sigma\|_\infty \delta^{1/2}) + \varepsilon$ is minimized for δ sufficiently small, yielding $\varepsilon = -\|\sigma\|_\infty \delta^{1/2} \log(\|\sigma\|_\infty \delta^{1/2}) = \|\sigma\|_\infty \delta^{1/2-2\epsilon}(-\delta^{2\epsilon} \log(\|\sigma\|_\infty \delta^{1/2}))$ for arbitrarily small $\epsilon > 0$. Hence, with $C_4 = (2L_{G'}^2 \|\mu\|_\infty^2 + \frac{1}{2} L_{G''}^2 \|\sigma\|_\infty^4) T$, $C_5 = (4\|\mu\|_\infty^2 \|G'\|_\infty^2 + \|\sigma\|^4 \|G''\|_\infty^2) C_2$, $C_6 = (4\|\mu\|_\infty^2 \|G'\|_\infty^2 + \|\sigma\|^4 \|G''\|_\infty^2) C_3$, we get

$$E_1 \leq C_4 \varepsilon^2 + C_5 \exp\left(-\frac{\varepsilon}{\|\sigma\|_\infty \delta^{1/2}}\right) + C_6 \varepsilon$$

$$= C_4 \|\sigma\|_\infty^2 \delta^{1-4\epsilon}(-\delta^{2\epsilon} \log(\|\sigma\|_\infty \delta^{1/2}))^2$$

$$+ C_5 \|\sigma\|_\infty \delta^{1/2} + C_6 \|\sigma\|_\infty \delta^{1/2-2\epsilon}(-\delta^{2\epsilon} \log(\|\sigma\|_\infty \delta^{1/2})).$$

Thus, with $C_7 = C_4 \|\sigma\|_\infty^2 + C_5 \|\sigma\|_\infty + C_6 \|\sigma\|_\infty$ and for arbitrarily small fixed $\epsilon > 0$, it holds that for sufficiently small δ

$$E_1 \leq C_7 \delta^{1/2-2\epsilon}. \tag{12}$$

For estimating E_2 in (11), we apply Lemma 3.2 to get

$$E_2 \leq L_{G'}^2 \|\sigma\|_\infty^2 \int_0^T \mathbb{E}\left(\left\|X_s^\delta - X_{\underline{s}}^\delta\right\|^2\right) ds \leq L_{G'}^2 \|\sigma\|_\infty^2 C_8 \delta. \tag{13}$$

For estimating E_3, E_4 in (11), we use that $\tilde{\mu}, \tilde{\sigma}$ are Lipschitz by [15, Theorem 3.20], to get

$$E_3 \leq L_{\tilde{\mu}}^2 \int_0^\tau \mathbb{E}\left(\left\|G(X_{\underline{s}}^\delta) - Z_{\underline{s}}^\delta\right\|^2\right) ds \leq L_{\tilde{\mu}}^2 \int_0^\tau u(s) ds, \tag{14}$$

$$E_4 \leq L_{\tilde{\sigma}}^2 \int_0^\tau \mathbb{E}\left(\left\|G(X_s^\delta) - Z_{\underline{s}}^\delta\right\|^2\right) ds \leq L_{\tilde{\sigma}}^2 \int_0^\tau u(s) ds. \tag{15}$$

Combining the estimates (12), (13), (14), (15) with (11), we get

$$0 \le u(\tau) \le C_9 \int_0^\tau u(s)ds + 4TC_7\delta^{1/2-2\epsilon} + 8dL_{G'}^2 \|\sigma\|_\infty^2 C_8\delta \,,$$

with $C_9 = 4TL_{\tilde{\mu}}^2 + 8dL_{\tilde{\sigma}}^2$. Using that $4TC_7\delta^{1/2-2\epsilon} + 8dL_{G'}^2 \|\sigma\|_\infty^2 C_8\delta \le C_{10}\delta^{1/2-2\epsilon}$ for $\delta \le 1$, and applying Gronwall's inequality yields for all $\tau \in [0, T]$,

$$u(\tau) \le C_{10} \exp(C_9\tau)\delta^{1/2-2\epsilon} \,. \tag{16}$$

Combining (10) and (16) with (9), and the result with (8), finally yields

$$\mathbb{E}\left(\sup_{0 \le t \le T} \|X_t - X_t^\delta\|^2 \right)^{1/2} \le L_{G^{-1}}\left(C_1\delta \right)^{1/2} + L_{G^{-1}}\left(C_{10} \exp(C_9T)\delta^{1/2-2\epsilon} \right)^{1/2} \le C\delta^{1/4-\epsilon} \,,$$

for a suitably chosen constant C, for arbitrarily small $\epsilon > 0$, and for δ sufficiently small.   □

4 Examples

We ran simulations for several examples—ones of theoretical interest as well as an example coming from applications.

When studying stochastic dynamical systems which include a noisy signal, then filtering this signal leads to a higher dimensional system with a degenerate diffusion coefficient. Stochastic control problems often lead to an optimal control policy which makes the drift of the system discontinuous. Examples are models with incomplete market information in mathematical finance where the rate with which cashflows are paid from a firm value process change systematically when the asset-liability ratio passes a certain threshold which then triggers a rating change.

The class of equations studied here appears frequently in several areas of applied mathematics and the natural sciences.

Step-function

In the first example the drift is the step function $\mu(x_1, x_2) = (3(\mathbf{1}_{\{x_1 \ge 0\}} - \mathbf{1}_{\{x_1 < 0\}}), 1)^\top$, and $\sigma \equiv \mathrm{id}_{\mathbb{R}^2}$. It can easily be checked that these coefficients satisfy Assumption 2.1. In particular, note that the non-parallelity condition is trivially satisfied, since σ is uniformly non-degenerate. Since μ does not satisfy a one-sided Lipschitz condition, our result is the first one that gives a strong convergence rate of the Euler–Maruyama method for this example.

Discontinuity along the unit circle

In this example the drift has a discontinuity along the unit circle, and the diffusion coefficient is degenerate on the whole of $\mathbb{R}^2$:

$$\mu(x_1, x_2) = \begin{cases} (1, 1)^\top & x_1^2 + x_2^2 \geq 1 \\ (-x_1, x_2)^\top & x_1^2 + x_2^2 < 1, \end{cases} \qquad \sigma(x_1, x_2) = \frac{2}{1 + x_1^2 + x_2^2} \begin{pmatrix} x_1 & 0 \\ x_2 & 0 \end{pmatrix}.$$

Assumption 2.1 largely follows from Example 2.6. The non-parallelity condition is readily verified:

$$\left\| \frac{2}{(1 + x_1^2 + x_2^2)(x_1^2 + x_2^2)} \begin{pmatrix} x_1 & x_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\| = \frac{2\sqrt{x_1^2 + x_2^2}}{(1 + x_1^2 + x_2^2)(x_1^2 + x_2^2)} = 1$$

for all points (x_1, x_2) that lie on the unit circle, i.e. $x_1^2 + x_2^2 = 1$.

Dividend maximization under incomplete information

In insurance mathematics, a well-studied problem is the maximization of the expected discounted future dividend payments until the time of ruin of an insurance company, a value which serves as a risk measure. In [24] the problem is studied in a setup that allows for incomplete information about the market. This leads to a joint filtering and stochastic optimal control problem, and after solving the filtering problem, the driving dynamics are high dimensional and have a degenerate diffusion coefficient. This issue is described in more detail in [24]. Solving the stochastic optimal control problem in dimensions higher than three with the usual technique (solving an associated partial differential equation) becomes practically infeasible. Therefore, one has to resort to simulation. The SDE that has to be simulated has the coefficients

$$\mu(x_1, \ldots, x_d) = \begin{pmatrix} \vartheta_d + \sum_{i=1}^{d-1}(\vartheta_i - \vartheta_d)x_{i+1} - \bar{u}\mathbf{1}_{[f(x_2,\ldots,x_d),\infty)}(x_1) \\ q_{d1} + \sum_{j=1}^{d-1}(q_{j1} - q_{d1})x_{j+1} \\ \vdots \\ q_{d(d-1)} + \sum_{j=1}^{d-1}(q_{j(d-1)} - q_{d(d-1)})x_{j+1} \end{pmatrix}$$

$$\sigma(x_1, \ldots, x_d) = \begin{pmatrix} \beta & 0 \ldots 0 \\ x_2 \frac{\vartheta_1 - \vartheta_d - \sum_{j=1}^{d-1}(\vartheta_j - \vartheta_d)x_{j+1}}{\beta} & \vdots \quad \vdots \\ \vdots & \vdots \quad \vdots \\ x_d \frac{\vartheta_{d-1} - \vartheta_d - \sum_{j=1}^{d-1}(\vartheta_j - \vartheta_d)x_{j+1}}{\beta} & 0 \ldots 0 \end{pmatrix},$$

where $\bar{u}, \beta, (\vartheta_i)_{i=1}^d, (q_{ij})_{i,j=1}^d$ are known constants. The arguments $x_2, \ldots, x_d$ are elements of the simplex $\{(x_2, \ldots, x_d) \in [0, 1]^{d-1} : \sum_{j=1}^{d-1} x_{j+1} \leq 1\}$, and the
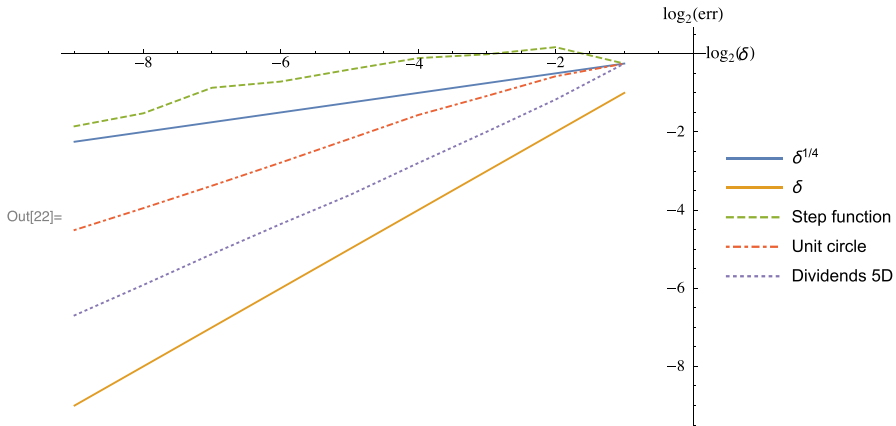
Fig. 1 Estimated L^2-errors

corresponding processes stay within this simplex almost surely, see [24]. The function f determines the hypersurface Θ along which the drift is discontinuous: $\Theta = \{(x_1, \ldots, x_d) : x_1 = f(x_2, \ldots, x_d)\}$. In our simulations we choose $d = 5$ and f affine linear, but note that we need not restrict ourselves to affine linear f.

We need to check Assumption 2.1: Since $x_2, \ldots, x_d \in [0, 1]$, μ, σ are bounded, and all first order derivatives of the entries of σ are bounded. Hence, σ is Lipschitz. μ is piecewise Lipschitz, and since f is affine linear, $\Theta \in C^3$. Whether the non-parallelity condition holds depends on the choice of the parameters, but for ours the condition is satisfied. Assumption 2.1.5 can easily be checked. Note that the coefficients can be extended to the whole of $\mathbb{R}^d$ in a way that they still satisfy our assumptions.

Error estimate

The L^2-error is estimated by

$$\text{err}_k := \bar{e}\, \hat{E}\left(\left\| X_T^{(k)} - X_T^{(k-1)} \right\|^2 \right)^{1/2},$$

where $X_T^{(k)}$ is the numerical approximation of X_T with step size $\delta^{(k)}$, $\hat{E}$ is an estimator of the mean value using 2^{14} paths, and $\bar{e}$ is a normalizing constant so that $\text{err}_1 = \sqrt{1/4}$.

Figure 1 shows $\log_2$ of the estimated L^2-error of the Euler–Maruyama approximation of X_T plotted over $\log_2 \delta^{(k)}$ for the examples presented above. We observe that the theoretical convergence rate is approximately obtained for the example of a step-function and that the other examples converge at a faster rate. In particular, for the examples with degenerate diffusion coefficient, the convergence rate is not worse than for the other example. Even for the step-function example, for sufficiently small step-size the convergence rate seems to be higher than the theoretical one. Hence, it will be an interesting topic for future research to prove sharpness, or find a sharp bound.

Table 1 Runtimes in seconds using sequential computation and estimated errors for the Euler–Maruyama method (EM) and the transformation method (GM) with 512 time-steps and 1024 paths

	Computation time		Estimated error	
	EM	GM	EM	GM
Step function	10.84	86.92	0.1324	0.3362
Unit circle	14.39	5267.37	0.0195	0.0323
Dividends 5D	45.52	7398.97	0.0026	0.0032

Even though the proven rate for the Euler–Maruyama method is lower than for the transformation-based method from [15], the calculations are usually faster in practice using the first method, since the simulation of a single path is faster. Table 1 confirms this claim: we observe that computation times are higher by up to two orders of magnitude for the transformation method, while the estimated error is of comparable size.

For completeness, we remark that one can construct examples, where the transformation method is much faster while giving a smaller error. For example, start with prescribing the transform $G(x) = x + x|x|\phi(10x)$ and set $\mu(x) = \frac{1}{2}(G^{-1})''(G(x))$ and $\sigma(x) = (G^{-1})'(G(x))$. This leads to $\tilde{\mu}(z) = 0$ and $\tilde{\sigma}(z) = 1$. Hence, if we use the transformation method with the same G, then $Z^\delta = Z = W$ and the transformation method gives the estimate $G^{-1}(W)$, which is the exact solution.

5 Conclusion

In this paper we have for the first time proven strong convergence and also a positive strong convergence rate for an explicit method (the Euler–Maruyama method) for multidimensional SDEs with discontinuous drift that has a degenerate diffusion coefficient, or with a discontinuous drift that does not satisfy a one-sided Lipschitz condition, or both. The Euler–Maruyama method has the advantage that it does not need the exact form of the set of discontinuities of the drift as an input, and that in practice, computation of one path is fast in comparison to the second method in the literature that can deal with this class of SDEs. Our numerical experiments suggest that in addition to these advantages, it even seems that the Euler–Maruyama method converges at a higher than the theoretically obtained rate for many examples and it will be a topic of future research to prove sharpness, or find a sharp bound.

References

1. Étoré, P., Martinez, M.: Exact simulation for solutions of one-dimensional stochastic differential equations involving a local time at zero of the unknown process. Monte Carlo Methods Appl. **19**(1), 41–71 (2013)
2. Étoré, P., Martinez, M.: Exact simulation for solutions of one-dimensional stochastic differential equations with discontinuous drift. ESAIM Probab. Stat. **18**, 686–702 (2014)
3. Foote, R.L.: Regularity of the distance function. Proc. Am. Math. Soc. **92**(1), 153–155 (1984)
4. Gyöngy, I.: A note on Euler's approximation. Potential Anal. **8**, 205–216 (1998)
5. Gyöngy, I., Rásonyi, M.: A note on Euler's approximation for SDEs with Hölder continuous diffusion coefficients. Stoch. Process. Appl. **121**(10), 2189–2200 (2011)
6. Hairer, M., Hutzenthaler, M., Jentzen, A.: Loss of regularity for Kolmogorov equations. Ann. Probab. **43**(2), 468–527 (2015)
7. Halidias, N., Kloeden, P.E.: A note on the Euler–Maruyama scheme for stochastic differential equations with a discontinuous monotone drift coefficient. BIT Numer. Math. **48**(1), 51–59 (2008)
8. Hutzenthaler, M., Jentzen, A., Kloeden, P.E.: Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. Ann. Appl. Probab. **22**(4), 1611–1641 (2012)
9. Itô, K.: On stochastic differential equations. Mem. Am. Math. Soc. **4**, 1–57 (1951)
10. Jentzen, A., Müller-Gronbach, T., Yaroslavtseva, L.: On stochastic differential equations with arbitrary slow convergence rates for strong approximation. Commun. Math. Sci. **14**(6), 1477–1500 (2016)
11. Karatzas, I., Shreve, S.E.: Brownian motion and stochastic calculus. Graduate texts in mathematics, 2nd edn. Springer, New York (1991)
12. Kloeden, P.E., Platen, E.: Numerical solutions of stochastic differential equations. Stochastic modelling and applied probability. Springer, Berlin (1992)
13. Kohatsu-Higa, A., Lejay, A., Yasuda, K.: Weak approximation errors for stochastic differential equations with non-regular drift. Preprint, Inria, hal-00840211 (2013)
14. Leobacher, G., Szölgyenyi, M.: A numerical method for SDEs with discontinuous drift. BIT Numer. Math. **56**(1), 151–162 (2016)
15. Leobacher, G., Szölgyenyi, M.: (2016) A strong order 1/2 method for multidimensional SDEs with discontinuous drift. Ann. Appl. Probab. (2016). arXiv:1512.02807
16. Leobacher, G., Szölgyenyi, M., Thonhauser, S.: On the existence of solutions of a class of SDEs with discontinuous drift and singular diffusion. Electron. Commun. Probab. **20**(6), 1–14 (2015)
17. Lima, E.L.: The Jordan–Brouwer separation theorem for smooth hypersurfaces. Am. Math. Mon. **95**(1), 39–42 (1988)
18. Müller-Gronbach, T., Yaroslavtseva, L.: On hard quadrature problems for marginal distributions of SDEs with bounded smooth coefficients (2016). arXiv:1603.08686
19. Ngo, H.L., Taguchi, D.: On the Euler–Maruyama approximation for one-dimensional stochastic differential equations with irregular coefficients. (2016). arXiv:1509.06532
20. Ngo, H.L., Taguchi, D.: Strong convergence for the Euler–Maruyama approximation of stochastic differential equations with discontinuous coefficients. (2016). arXiv:1604.01174
21. Ngo, H.L., Taguchi, D.: Strong rate of convergence for the Euler–Maruyama approximation of stochastic differential equations with irregular coefficients. Math. Comput. **85**(300), 1793–1819 (2016)
22. Sabanis, S.: A note on tamed Euler approximations. Electron. Commun. Probab. **18**(47), 1–10 (2013)
23. Shardin, A.A., Szölgyenyi, M.: Optimal control of an energy storage facility under a changing economic environment and partial information. Int. J. Theor. Appl. Finance **19**(4), 1–27 (2016)
24. Szölgyenyi, M.: Dividend maximization in a hidden Markov switching model. Stat. Risk Model. **32**(3–4), 143–158 (2016)

25. Veretennikov, A.Y.U.: On strong solutions and explicit formulas for solutions of stochastic integral equations. Math. USSR Sb. **39**(3), 387–403 (1981)
26. Veretennikov, A.Y.U.: On the criteria for existence of a strong solution of a stochastic equation. Theory Probab. Appl. **27**(3), 441–449 (1982)
27. Veretennikov, A.Y.U.: On stochastic equations with degenerate diffusion with respect to some of the variables. Math. USSR Izvestiya **22**(1), 173–180 (1984)
28. Yaroslavtseva, L.: On non-polynomial lower error bounds for adaptive strong approximation of SDEs. (2016). arXiv:1609.08073
29. Zvonkin, A.K.: A transformation of the phase space of a diffusion process that removes the drift. Math. USSR Sb. **22**(129), 129–149 (1974)