

Convergence of Relative Entropy for Euler–Maruyama Scheme to Stochastic Differential Equations with Additive Noise

Yuan Yu 

School of Statistics and Mathematics, Shandong University of Finance and Economics, Jinan 250014, China; yuyuan_mail@126.com

Abstract: For a family of stochastic differential equations driven by additive Gaussian noise, we study the asymptotic behaviors of its corresponding Euler–Maruyama scheme by deriving its convergence rate in terms of relative entropy. Our results for the convergence rate in terms of relative entropy complement the conventional ones in the strong and weak sense and induce some other properties of the Euler–Maruyama scheme. For example, the convergence in terms of the total variation distance can be implied by Pinsker’s inequality directly. Moreover, when the drift is $\beta(0 < \beta < 1)$ -Hölder continuous in the spatial variable, the convergence rate in terms of the weighted variation distance is also established. Both of these convergence results do not seem to be directly obtained from any other convergence results of the Euler–Maruyama scheme. The main tool this paper relies on is the Girsanov transform.

Keywords: relative entropy; Euler–Maruyama scheme; Girsanov’s transform; Hölder continuity; weighted variation distance

1. Introduction

Consider the following d -dimensional stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \quad X_0 = x \in \mathbb{R}^d, \quad (1)$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^m$, and $(W_t)_{t \geq 0}$ is m -dimensional Brownian motion on some complete filtration probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$.

Usually, it can be proved that (1) has strong well-posedness under reasonable conditions, whereas the explicit representation of (1) is unknown. Instead, one may develop various numerical schemes to approximate (1); see [1] and references therein for more introductions. When the coefficients are regular, strong/weak convergence of numerical schemes for SDEs have been investigated considerably; see, for instance, monographs [1].

One of the most popular numerical schemes for SDEs is the Euler–Maruyama (EM) scheme, the introduction of which can be found in [1] and references therein. The Euler–Maruyama scheme is a numerical method commonly used for approximating the solutions of SDEs. SDEs are differential equations that involve both deterministic and stochastic (random) components. They are applied in various fields, including physics, finance, biology, and more. The Euler–Maruyama method is particularly useful for solving SDEs because it is a simple and computationally efficient approach. It is an extension of the Euler method, which is used for solving ordinary differential equations. The Euler–Maruyama method is adapted to handle the stochastic part of the equations. Besides its fundamental tools for the numerical solutions of SDEs, researchers also pay much attention to its convergence rate for SDEs. There are some related works in the literature.

For strong convergence of the EM scheme for SDEs, there are some basic results under irregular coefficients. Yan (2002) [2] uses Meyer–Tanaka’s formula and estimates local times to derive a strong convergence rate for EM schemes to one-dimensional SDEs, for which the drift is Lipschitz continuous and the diffusion is Hölder continuous. Gyöngy and



Citation: Yu, Y. Convergence of Relative Entropy for Euler–Maruyama Scheme to Stochastic Differential Equations with Additive Noise.

Entropy **2024**, *26*, 232. <https://doi.org/10.3390/e26030232>

Academic Editor: Philip Broadbridge

Received: 28 January 2024

Revised: 29 February 2024

Accepted: 2 March 2024

Published: 6 March 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Rásonyi (2011) [3] adopt a Yamada–Watanabe approximation approach to derive strong convergence for EM schemes of one-dimensional SDEs with Hölder-continuous diffusions, but the drifts cannot be Lipschitz continuous. Halidias and Kloeden (2008) [4] established strong convergence for an EM scheme of SDEs with monotone drift, which may be discontinuous. Leobacher and Szölgényi (2016) and Müller-Gronbach and Yaroslavtseva (2020) [5,6] investigated strong convergence of the EM scheme for one-dimensional SDEs with piecewise Lipschitz-continuous drifts.

Moreover, there are also some other authors who use some transformation tools to get strong convergence of an EM scheme for more complex cases. Leobacher and Szölgényi (2017) [7] studied an EM scheme for the multi-dimensional case, which was extended by Leobacher and Szölgényi (2018) [8] to the multi-dimensional and degenerate case. The proofs are based on a transformation that changes the piecewise Lipschitz-continuous drifts into globally Lipschitz-continuous ones. Besides the transformation mentioned above, the Zvonkin transform is an alternative tool to deal with the convergence for an EM scheme of SDEs with irregular coefficients. Bao, Huang and Yuan (2019) and Pamen and Taguchi (2017) [9,10] studied SDEs with Hölder- or Hölder–Dini-continuous drifts. Bao, Huang and Zhang (2022) [11] focused on the integrability condition; see the references in [9–11] for more results.

For weak convergence, one can refer to [12,13], wherein the drift satisfies the integrable condition and the main tool is the Girsanov transform.

We should remark that all of the references mentioned above study weak or strong convergence of the EM scheme. As far as we know, there are no results on the convergence in the sense of relative entropy.

In this paper, we further characterize the asymptotic behaviors of an EM scheme for SDEs by studying the convergence rate in terms of the so-called relative entropy. Our main results show that the distribution of the EM iteration of SDEs driven by additive Gaussian noise converges to that of real stochastic process of SDEs in terms of relative entropy. And hence, we can get the asymptotic behaviors of its corresponding EM scheme.

Indeed, while relative entropy is commonly perceived as a measure of the dissimilarity between two probability distributions, it falls short of being considered a metric. This is primarily attributed to its asymmetry concerning the order of its parameters and its inability to satisfy the triangle inequality. Despite not meeting the criteria of a metric, relative entropy maintains strong connections with various other metrics. Some notable relationships include: total variation distance, Fisher information divergence, Wasserstein distance and so on; see, e.g., [14] and references therein. These relationships highlight the versatility of relative entropy and its role in connecting with various other measures of dissimilarity and divergence between probability distributions. While it may not possess all the properties of a metric, its specific characteristics make it a valuable tool in information theory and related fields.

Relative entropy's broad utilization across a spectrum of disciplines, spanning probability theory, statistics, statistical physics, machine learning, neural science, and information theory, can be attributed to its possession of numerous advantageous properties; see, e.g., [15] and references therein. These qualities make it a versatile and valuable tool in diverse applications and fields of study. In addition to conventional convergence analysis in both the strong and weak senses for EM schemes, the relative entropy convergence in our findings could unveil previously unexplored facets of SDEs. These discoveries hold the potential to reveal new properties of SDEs in uncharted research territories, presenting exciting prospects for further exploration and investigation.

As an example of the direct corollary of our main results, the convergence in total variation of an EM scheme can be implied by the well-known Pinsker's inequality. Pinsker's inequality states that the relative entropy between two probability measures provides an upper bound for their total variation distance.

Moreover, when the drift of SDE (4) is $\beta(0 < \beta < 1)$ -Hölder-continuous in the spatial variable, the convergence rate for the weighted variation distance will also be induced.

The paper is organized as follows: In Section 2, we review some related concepts and definitions of this paper. In Section 3, we state our assumptions and introduce the main results. In Section 4, we induce the proofs of all results. And conclusions and discussions are provided in Section 5.

2. Preliminaries

In this section, we first state some definitions and some related concepts for the main results of this paper.

2.1. Euler–Maruyama Scheme

The Euler–Maruyama scheme is a numerical method commonly used for approximating solutions to SDEs. SDEs involve both deterministic and stochastic components, making their solutions more challenging than those of ordinary differential equations. The basic idea is similar to the traditional Euler method for ordinary differential equations but adapted to handle the stochastic terms. The Euler–Maruyama scheme is one of the simplest time-discrete approximations of Itô's process, and it is sometimes called the Euler–Maruyama approximation or Euler approximation.

We consider an Itô's process satisfying the stochastic differential Equation (1):

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \quad X_0 = x \in \mathbb{R}^d,$$

on time interval $[0, T]$ with initial value x .

For a given discretization $0 = \tau_0 < \tau_1 < \dots < \tau_N = T$ of the time interval $[0, T]$, a discrete EM scheme satisfies the following iterative scheme:

$$Y_{n+1} = Y_n + b(Y_n)(\tau_{n+1} - \tau_n) + \sigma(Y_n)(W_{\tau_{n+1}} - W_{\tau_n}), \quad (2)$$

for $n = 0, 1, 2, \dots, N - 1$ with initial value $Y_0 = X_0$. We shall also write $\Delta_n = \tau_{n+1} - \tau_n$ for the n th time increment and call $\delta = \max_n \Delta_n$ the maximum time step.

In this paper, we shall consider equidistant discretization times $\tau_n = n\delta$ with $\delta = T/N$ for some integer N large enough so that $\delta \in (0, 1)$. When the diffusion coefficient is identically zero—that is, when $\sigma = 0$ —the stochastic iterative scheme reduces to the deterministic Euler scheme for the ordinary differential equation $dX_t = b(X_t)dt$. The main difference is that we need to generate the random increments $\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$ for $n = 0, 1, \dots, N - 1$ of the Wiener process $W = \{W_t, t \geq 0\}$. From the properties of Wiener processes, we know that these increments are independent Gaussian random variables with mean $E(\Delta W_n) = 0$ and variance $E((\Delta W_n)^2) = \Delta_n$. We can use a sequence of independent Gaussian pseudo-random numbers generated by one of the random number generators for the increments of the Wiener process.

The recursive structure of the discrete EM scheme, which evaluates approximate values for the Itô's process at the discretization instants only, is the key to its successful implementation for numerical approximation of SDEs. For a given time discretization, the discrete EM scheme determines values of the approximation process at the discretization times only. We also need the continuous EM scheme, i.e.,

$$dX_t^\delta = b(X_{t_\delta}^\delta)dt + \sigma(X_{t_\delta}^\delta)dW_t, \quad t \geq 0, \quad X_0^\delta = x \in \mathbb{R}^d,$$

where $t_\delta := \lfloor t/\delta \rfloor \delta$, and $\lfloor t/\delta \rfloor$ is the integer part of t/δ . More details and introductions can be found in [1] and references therein.

2.2. Relative Entropy

Kullback and Leibler (1951) [16] firstly introduced the definition of relative entropy, which is also called Kullback–Leibler divergence (K-L divergence for short). The definition of relative entropy is as bellow.

Definition 1 (Relative entropy). Recall that the relative entropy of two probability measures ν and μ on \mathbb{R}^d is defined as

$$\text{Ent}(\nu|\mu) = \begin{cases} \int_{\mathbb{R}^d} \log\left(\frac{d\nu}{d\mu}\right) d\nu, & \nu \ll \mu; \\ \infty, & \text{otherwise,} \end{cases}$$

where $\frac{d\nu}{d\mu}$ is the Radon–Nikodym derivative of ν with respect to μ .

Relative entropy is a concept from information theory and probability theory that measures how one probability distribution diverges from another one. Roughly speaking, the relative entropy between two probability measures is a measure of the “distance” or difference measuring how “close” these two probability distributions are. In Chapter 4 of reference [17], the authors provide lots of properties of relative entropy (K-L divergence), establish the relationship between relative entropy, cross entropy and conventional differential entropy and give some examples of relative entropy calculations: for instance, exponential distributions, normal distributions and Poisson distributions.

2.3. Total Variation and Weighted Variation Distance

Definition 2 (Total variation). For two probability measures $\gamma, \tilde{\gamma}$ on \mathbb{R}^d , the total variation distance is formulated as

$$\|\gamma - \tilde{\gamma}\|_{var} = \sup_{\|f\|_{\infty} \leq 1} |\gamma(f) - \tilde{\gamma}(f)|.$$

Definition 3 (Pinsker’s inequality).

$$\|\gamma - \tilde{\gamma}\|_{var}^2 \leq 2\text{Ent}(\gamma|\tilde{\gamma}). \quad (3)$$

Remark 1. In view of (3), the convergence of the total variation distance can be implied by Pinsker’s inequality directly under the relative entropy convergence.

Definition 4 (Weighted variation distance). For any $k > 1$, the weighted variation distance for two probability measures $\gamma, \tilde{\gamma}$ on \mathbb{R}^d is formulated as

$$\|\gamma - \tilde{\gamma}\|_{k,var} = \sup_{|f| < 1 + |\cdot|^k} |\gamma(f) - \tilde{\gamma}(f)|.$$

Remark 2. The convergence of the weighted total variation distance cannot be implied by the relative entropy convergence, so we need to further investigate it.

2.4. Stochastic Differential Equation Description

Definition 5 (Stochastic differential equations with additive noise). In this paper, we consider the following SDE

$$dX_t = b(X_t)dt + dW_t, \quad t \geq 0, \quad X_0 = x \in \mathbb{R}^d, \quad (4)$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$, and $(W_t)_{t \geq 0}$ is d -dimensional Brownian motion on some complete filtration probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$.

Remark 3. For any $\delta \in (0, 1)$, the continuous Euler–Maruyama (EM) method for (4) is defined as

$$dX_t^\delta = b(X_{t_\delta}^\delta)dt + dW_t, \quad t \geq 0, \quad X_0^\delta = x \in \mathbb{R}^d, \quad (5)$$

with $t_\delta := \lfloor t/\delta \rfloor \delta$, where $\lfloor t/\delta \rfloor$ is the integer part of t/δ .

3. Assumptions and Main Results

3.1. Assumptions

Throughout the paper, we impose the following assumptions on the drift term b of the SDE (4).

(A) There exists a constant $\beta \in (0, 1]$ and $K > 0$ such that

$$|b(x) - b(y)| \leq K|x - y|^\beta, \quad x, y \in \mathbb{R}^d \tag{6}$$

Remark 4. By Zvonkin’s transform introduced in [18], under assumption (A), (4) has a unique strong solution $(X_t)_{t \geq 0}$ (see, for instance, [19]).

3.2. Main Results

Let \mathcal{L}_ξ denote the distribution of a random variable ξ . The main result is the following theorem.

Theorem 1. Assume the drift term of the SDE (4) satisfies assumption (A). Then there exists constants $C_{T,x,d}$ such that

$$\text{Ent}(\mathcal{L}_{X_t^\delta} | \mathcal{L}_{X_t}) \leq KC_{T,x,d}t\delta^\beta, \quad t \in [0, T]. \tag{7}$$

Consequently, we have

$$\lim_{\delta \rightarrow 0} \text{Ent}(\mathcal{L}_{X_t^\delta} | \mathcal{L}_{X_t}) = 0.$$

Remark 5. Theorem 1 gives the convergence rate of an EM scheme in the sense of relative entropy for SDEs (4) with additive noise, so its asymptotic behaviors can be established. The main tool of the proof relies on the Girsanov transform. The details of the proof can be found in Section 4.

Corollary 1. When assumption (A) is satisfied, we have

$$\|\mathcal{L}_{X_t^\delta} - \mathcal{L}_{X_t}\|_{var} \leq \sqrt{2KC_{T,x,d}t\delta^{\frac{\beta}{2}}}, \quad t \in [0, T].$$

Remark 6. Corollary 1 is the convergence of an EM scheme for the total variance distance. This can be implied by Pinker’s inequality (3) directly.

Theorem 2. If assumption (A) holds for some $0 < \beta < 1$, then for any $k \geq 1$ there exists a constant $c_{k,T,x,d}$ such that

$$\|\mathcal{L}_{X_t^\delta}(f) - \mathcal{L}_{X_t}(f)\|_{k,var} = \sup_{|f| < 1 + |\cdot|^k} |\mathcal{L}_{X_t^\delta}(f) - \mathcal{L}_{X_t}(f)| \leq c_{k,T,x,d}\sqrt{t}\delta^{\frac{\beta}{2}}, \quad t \in [0, T].$$

Remark 7. Theorem 2 is the convergence of an EM scheme for the weighted total variance distance. This convergence is not the direct application of the convergence in the relative entropy sense for the EM scheme of Theorem 1. And the details of the proof can be found in Section 4.

4. Proofs

4.1. Proof of Theorem 1

Before finishing the proof of Theorem 1, we prepare some auxiliary lemmas. The first lemma below plays a crucial role in the proof of Theorem 1.

Lemma 1. Assume (A). Then for any $k \geq 1$, there exists a constant $C_{T,d,k} > 0$ such that

$$\mathbb{E} \sup_{t \in [0, T]} |X_t^\delta|^k + \mathbb{E} \sup_{t \in [0, T]} |X_t|^k \leq C_{T,d,k}(1 + |x|^k), \quad t \in [0, T]. \tag{8}$$

Proof. Without loss of generality, we only prove the inequality for X_t^δ since it is similar for X_t .

For any $n \geq 1$, let $\zeta_n := \inf\{t \geq 0, |X_t^\delta| \geq n\}$. Firstly, we have

$$|X_{t \wedge \zeta_n}^\delta|^k \leq 2^{k-1} \left| \int_0^{t \wedge \zeta_n} b(X_{s_\delta}^\delta) ds \right|^k + 2^{k-1} |W_{t \wedge \zeta_n}|^k, \quad t \in [0, T]. \tag{9}$$

By **(A)**, it is easy to see that

$$|b(x)| \leq |b(x) - b(0)| + |b(0)| \leq K|x|^\beta + |b(0)| \leq K(1 + |x|) + |b(0)|, \quad x \in \mathbb{R}^d. \tag{10}$$

Combining this with (9), we can find a constant $c_0 > 0$ such that

$$\mathbb{E} \sup_{t \in [0, r]} |X_{t \wedge \zeta_n}^\delta|^k \leq c_0 T \mathbb{E} \int_0^r (1 + \sup_{t \in [0, s]} |X_{t \wedge \zeta_n}^\delta|^k) ds + 2^{k-1} \mathbb{E} \sup_{t \in [0, r]} |W_{t \wedge \zeta_n}|^k, \quad r \in [0, T]. \tag{11}$$

By the Burkholder–Davis–Gundy inequality, there exists a constant $c_1 > 0$ such that

$$\mathbb{E} \sup_{t \in [0, r]} |W_{t \wedge \zeta_n}|^k \leq c_1 (dr)^{\frac{k}{2}}, \quad r \in [0, T].$$

Putting this into (11) and applying Gronwall’s inequality, we find a constant $C_T > 0$ such that

$$\mathbb{E} \sup_{t \in [0, T]} |X_{t \wedge \zeta_n}^\delta|^k \leq C_{T,d,k} (1 + |x|^k).$$

Note that

$$\mathbb{P}(\zeta_n < T) \leq \mathbb{P}(|X_{T \wedge \zeta_n}^\delta| \geq n) \leq \frac{\mathbb{E} \sup_{t \in [0, T]} |X_{t \wedge \zeta_n}^\delta|^k}{n^k} \leq \frac{C_{T,d,k} (1 + |x|^k)}{n^k}.$$

This yields that \mathbb{P} -a.s. $\lim_{n \rightarrow \infty} \zeta_n = \infty$, which, combined with Fatou’s lemma, yields that

$$\mathbb{E} \sup_{t \in [0, T]} |X_t^\delta|^k \leq \liminf_{n \rightarrow \infty} \mathbb{E} \sup_{t \in [0, T]} |X_{t \wedge \zeta_n}^\delta|^k \leq C_{T,d,k} (1 + |x|^k).$$

So we complete the proof. \square

Lemma 2. Under **(A)**, there exists a constant $C_{T,x,d} > 0$ such that

$$\sup_{t \in [0, T]} \mathbb{E} |X_t^\delta - X_{t_\delta}^\delta|^4 \leq C_{T,x,d} \delta. \tag{12}$$

Proof. Note that

$$X_t^\delta - X_{t_\delta}^\delta = \int_{t_\delta}^t b(X_{s_\delta}^\delta) ds + W_t - W_{t_\delta}$$

This together with (10), Lemma 1 and the fact $\mathbb{E} |W_t - W_{t_\delta}|^4 = (d^2 + 2d)(t - t_\delta)^2 \leq (d^2 + 2d)\delta^2$ implies that

$$\begin{aligned} \mathbb{E} |X_t^\delta - X_{t_\delta}^\delta|^4 &\leq 8\delta^3 \int_{t_\delta}^t \mathbb{E} |b(X_{s_\delta}^\delta)|^4 ds + 8(d^2 + 2d)\delta^2 \\ &\leq 8\delta^3 C_0 \int_{t_\delta}^t \mathbb{E} (1 + \sup_{t \in [0, T]} |X_t^\delta|^4) ds + 8(d^2 + 2d)\delta^2 \\ &\leq C_{T,x,d} \delta^2. \end{aligned}$$

Therefore, the proof is completed. \square

Lemma 3. Assume (A). Then there exist constants $\kappa_1(T), \kappa_2(T, x, d) > 0$ such that

$$\mathbb{E} \exp \left\{ \kappa_1(T) \sup_{t \in [0, r]} |X_t|^2 \right\} \leq \kappa_2(T, x, d), \quad r \in [0, T].$$

Proof. By Itô’s formula, we can find a constant $c_0 > 0$ such that

$$\begin{aligned} |X_t|^2 &\leq |x|^2 + \int_0^t \langle b(X_s), 2X_s \rangle ds + \int_0^t \langle 2X_s, dW_s \rangle + dt \\ &\leq |x|^2 + (c_0 + d)t + c_0 \int_0^t |X_s|^2 ds + \int_0^t \langle 2X_s, dW_s \rangle. \end{aligned}$$

Gronwall’s inequality yields that

$$\sup_{t \in [0, r]} |X_t|^2 \leq e^{c_0 r} \left[|x|^2 + (c_0 + d)r + \sup_{t \in [0, r]} \int_0^t \langle 2X_s, dW_s \rangle \right], \quad r \in [0, T]. \tag{13}$$

Note that for any $\varepsilon > 0$,

$$\begin{aligned} \mathbb{E} \exp \left\{ \varepsilon \sup_{t \in [0, r]} \int_0^t \langle 2X_s, dW_s \rangle \right\} &\leq e \mathbb{E} \exp \left\{ \varepsilon \int_0^r \langle 2X_s, dW_s \rangle \right\} \\ &\leq e \left\{ \mathbb{E} e^{8\varepsilon^2 \int_0^r |X_s|^2 ds} \right\}^{\frac{1}{2}} \\ &\leq e \left\{ \mathbb{E} e^{8\varepsilon^2 r \sup_{s \in [0, r]} |X_s|^2} \right\}^{\frac{1}{2}}, \quad r \in [0, T]. \end{aligned} \tag{14}$$

Taking $\varepsilon = \frac{1}{8T} e^{-c_0 T}$, we have $\varepsilon e^{-c_0 T} = 8T\varepsilon^2$. Combining this with (13) and (14), we derive

$$\begin{aligned} \mathbb{E} \exp \left\{ \varepsilon e^{-c_0 T} \sup_{t \in [0, T]} |X_t|^2 \right\} &\leq \exp \{ \varepsilon [|x|^2 + (c_0 + d)T] \} \mathbb{E} \exp \left\{ \varepsilon \sup_{t \in [0, T]} \int_0^t \langle 2X_s, dW_s \rangle \right\} \\ &\leq \exp \{ \varepsilon [|x|^2 + (c_0 + d)T] \} e \left\{ \mathbb{E} e^{8\varepsilon^2 T \sup_{s \in [0, T]} |X_s|^2} \right\}^{\frac{1}{2}}. \end{aligned}$$

By a stopping time technique, we may and do assume that

$$\mathbb{E} \exp \left\{ \varepsilon e^{-c_0 T} \sup_{t \in [0, T]} |X_t|^2 \right\} < \infty.$$

Then we get

$$\mathbb{E} \exp \left\{ \frac{1}{8T} e^{-2c_0 T} \sup_{t \in [0, T]} |X_t|^2 \right\} \leq \exp \left\{ 2 + \frac{1}{4T} e^{-c_0 T} [|x|^2 + (c_0 + d)T] \right\}.$$

Letting $\kappa_1(T) = \frac{1}{8T} e^{-2c_0 T}$ and $\kappa_2(T, x, d) = \exp \left\{ 2 + \frac{1}{4T} e^{-c_0 T} [|x|^2 + (c_0 + d)T] \right\}$, we complete the proof. \square

Proof of Theorem 1. For any $n \geq 1$, let $\tau_n := \inf \{ t \geq 0, |X_t| \vee |X_t^\delta| \geq n \}$. By Lemma 1, it holds that \mathbb{P} -a.s. $\lim_{n \rightarrow \infty} \tau_n = \infty$. Recall that

$$dX_t = b(X_t)dt + dW_t, \quad t \in [0, T]. \tag{15}$$

Let

$$W_t^\delta = W_t - \int_0^t [b(X_{s_\delta}) - b(X_s)] ds, \quad t \in [0, T]. \tag{16}$$

Then (15) can be rewritten as

$$dX_t = b(X_{t_\delta}) dt + dW_t^\delta, \quad t \in [0, T].$$

Set

$$R_t = \exp \left\{ \int_0^t \langle [b(X_{s_\delta}) - b(X_s)], dW_s \rangle - \frac{1}{2} \int_0^t |b(X_{s_\delta}) - b(X_s)|^2 ds \right\}, \quad t \in [0, T].$$

Fix $t_0 \in (0, T]$. By **(A)** and Girsanov’s theorem, we conclude that $\{R_{t \wedge \tau_n}\}_{t \in [0, t_0]}$ is a martingale and W_t^δ is d -dimensional Brownian motion up to $t_0 \wedge \tau_n$ under probability measure $\mathbb{Q}^n = R_{t_0 \wedge \tau_n} \mathbb{P}$. This together with **(A)** and Lemma 2 implies that

$$\begin{aligned} & \mathbb{E}^{\mathbb{Q}^n} \int_0^{t_0 \wedge \tau_n} |b(X_{s_\delta}) - b(X_s)|^2 ds \\ &= \mathbb{E} \int_0^{t_0 \wedge \tau_n} |b(X_{s_\delta}^\delta) - b(X_s^\delta)|^2 ds \\ &\leq \mathbb{E} \int_0^{t_0} K^2 |X_{s_\delta}^\delta - X_s^\delta|^{2\beta} ds \\ &\leq \int_0^{t_0} K^2 \{\mathbb{E} |X_{s_\delta}^\delta - X_s^\delta|^2\}^\beta ds \leq K^2 C_{T,x,d} t_0 \delta^\beta. \end{aligned} \tag{17}$$

From this and (16), we derive that

$$\begin{aligned} \mathbb{E}[R_{t_0 \wedge \tau_n} \log R_{t_0 \wedge \tau_n}] &= \mathbb{E}^{\mathbb{Q}^n} [\log R_{t_0 \wedge \tau_n}] \\ &= \mathbb{E}^{\mathbb{Q}^n} \int_0^{t_0 \wedge \tau_n} \langle [b(X_{s_\delta}) - b(X_s)], dW_s \rangle - \frac{1}{2} \mathbb{E}^{\mathbb{Q}^n} \int_0^{t_0 \wedge \tau_n} |b(X_{s_\delta}) - b(X_s)|^2 ds \\ &= \mathbb{E}^{\mathbb{Q}^n} \int_0^{t_0 \wedge \tau_n} \langle [b(X_{s_\delta}) - b(X_s)], dW_s^\delta \rangle + \frac{1}{2} \mathbb{E}^{\mathbb{Q}^n} \int_0^{t_0 \wedge \tau_n} |b(X_{s_\delta}) - b(X_s)|^2 ds \\ &= \frac{1}{2} \mathbb{E}^{\mathbb{Q}^n} \int_0^{t_0 \wedge \tau_n} |b(X_{s_\delta}) - b(X_s)|^2 ds \\ &\leq \frac{1}{2} K^2 C_{T,x,d} t_0 \delta^\beta. \end{aligned}$$

This combined with the convergence theorem of martingales implies that $\{R_t\}_{t \in [0, t_0]}$ is a martingale, and it follows from Fatou’s lemma that

$$\mathbb{E}[R_{t_0} \log R_{t_0}] \leq \lim_{n \rightarrow \infty} \mathbb{E}[R_{t_0 \wedge \tau_n} \log R_{t_0 \wedge \tau_n}] \leq \frac{1}{2} K^2 C_{T,x,d} t_0 \delta^\beta.$$

Applying Girsanov’s theorem again, we conclude that $\{R_t\}_{t \in [0, t_0]}$ is a martingale and W_t^δ is d -dimensional Brownian motion up to t_0 under probability measure $\mathbb{Q} = R_{t_0} \mathbb{P}$, and hence, the distribution of $\{X_t\}_{t \in [0, t_0]}$ under \mathbb{Q} is equal to that of $\{X_t^\delta\}_{t \in [0, t_0]}$ under \mathbb{P} . As a result, it holds that

$$\mathbb{E} f(X_{t_0}^\delta) = \mathbb{E}^{\mathbb{Q}} f(X_{t_0}) = \mathbb{E}[R_{t_0} f(X_{t_0})], \quad |f| \leq 1 + |\cdot|^k, k \geq 1. \tag{18}$$

By Young’s inequality, we derive that

$$\text{Ent}(\mathcal{L}_{X_{t_0}^\delta} | \mathcal{L}_{X_{t_0}}) \leq C_{T,x,d} \delta^\beta, \quad t \in [0, T]. \tag{19}$$

The proof is completed. \square

4.2. Proof of Corollary 1

Corollary 1 is the direct result of Theorem 1 and Pinsker’s inequality (3).

4.3. Proof of Theorem 2

Proof. Firstly, we have

$$\begin{aligned} \mathbb{E}[R_{t_0} - 1]^2 &= \mathbb{E}[R_{t_0}^2] - 1 \\ &= \mathbb{E}\left\{ \exp\left\{ 2 \int_0^{t_0} \langle [b(X_{s_\delta}) - b(X_s)], dW_s \rangle - 4 \int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right\} \right. \\ &\quad \left. \times \exp\left\{ 3 \int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right\} \right\} - 1 \\ &\leq \left\{ \mathbb{E} \exp\left\{ 6 \int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right\} \right\}^{\frac{1}{2}} - 1 \\ &\leq \mathbb{E} \exp\left\{ 6 \int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right\} - 1 \\ &\leq \mathbb{E} \left[\exp\left\{ 6 \int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right\} 6 \int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right] \\ &\leq 6 \left\{ \mathbb{E} \exp\left\{ 12 \int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right\} \right\}^{\frac{1}{2}} \left[\mathbb{E} \left(\int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right)^2 \right]^{\frac{1}{2}} \\ &\leq 6 \left\{ \mathbb{E} \exp\left\{ 48K^2 t_0 \sup_{s \in [0, t_0]} |X_s|^{2\beta} ds \right\} \right\}^{\frac{1}{2}} \left[\mathbb{E} \left(\int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right)^2 \right]^{\frac{1}{2}} \end{aligned}$$

By Lemma 3 and for $\beta < 1$, we derive

$$6 \left\{ \mathbb{E} \exp\left\{ 48K^2 t_0 \sup_{s \in [0, t_0]} |X_s|^{2\beta} ds \right\} \right\}^{\frac{1}{2}} < C(T, x, d)$$

for some constant $C(T, x, d) > 0$. Moreover, Hölder’s inequality and Lemma 2 imply that

$$\left[\mathbb{E} \left(\int_0^{t_0} |b(X_{s_\delta}) - b(X_s)|^2 ds \right)^2 \right]^{\frac{1}{2}} \leq \sqrt{t_0} \left[\int_0^{t_0} \mathbb{E} |b(X_{s_\delta}) - b(X_s)|^4 ds \right]^{\frac{1}{2}} \leq c_1 t_0 \delta^\beta.$$

So we conclude that

$$\mathbb{E}[R_{t_0} - 1]^2 \leq c t_0 \delta^\beta$$

for some constant $c > 0$. From this together with (18) and Lemma 1, we derive

$$\begin{aligned} &\|\mathcal{L}_{X_t^\delta}(f) - \mathcal{L}_{X_t}(f)\|_{k, var} \\ &= \sup_{|f| < 1 + |\cdot|^k} |\mathcal{L}_{X_t^\delta}(f) - \mathcal{L}_{X_t}(f)| \\ &= \sup_{|f| < 1 + |\cdot|^k} |\mathbb{E}[(R_t - 1)f(X_t)]| \\ &\leq [\mathbb{E}(R_t - 1)^2]^{\frac{1}{2}} [\mathbb{E}(1 + |X_t|^k)^2]^{\frac{1}{2}} \\ &\leq c_{k, T, x, d} \sqrt{t} \delta^{\frac{\beta}{2}}. \end{aligned}$$

The proof is completed. \square

5. Conclusions and Discussion

In this paper, we studied the convergence rate of the relative entropy for a Euler–Maruyama scheme of stochastic differential equations driven by additive Gaussian noise, and we obtained its asymptotic behaviors. Our results for the convergence rate in terms of relative entropy complement the conventional ones in the strong and weak sense and induce some other properties of the Euler–Maruyama scheme. The convergence in terms of total variation distance can be implied by Pinsker’s inequality directly. And the convergence rate in terms of the weighted variation distance is also established.

Finally, we discuss some more complicated processes.

(1) SDEs with multiplicative noise: The SDE becomes

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x \quad (20)$$

The EM scheme satisfies

$$dX_t^\delta = b(X_{t_\delta}^\delta)dt + \sigma(X_{t_\delta}^\delta)dW_t, \quad X_0^\delta = x. \quad (21)$$

When σ is invertible, we can still rewrite (20) as

$$dX_t = b(X_{t_\delta})dt + \sigma(X_t)d\tilde{W}_t \quad (22)$$

with

$$d\tilde{W}_t = dW_t - \sigma^{-1}(X_t)(b(X_{t_\delta}) - b(X_t))dt.$$

Different from the additive noise case, (21) and (22) solve different SDEs and, hence, the Girsanov transform is unavailable. We need to develop new approaches to deal with the multiplicative noise case.

(2) Geometric Brownian motion (GBM): In reference [20], the authors investigated time-averaging and nonergodicity for GBM in the presence of drift and with resetting. Although GBM in the presence of drift has explicit representation and follows log-normal distribution, it solves an SDE with linear and multiplicative noise:

$$dX_t = \mu X_t dt + \sigma X_t dW_t.$$

The difficulty appearing in the multiplicative noise case will still exist.

(3) Anomalous-diffusion processes: In reference [21], the authors studied the nonergodicity, non-Gaussianity and aging of scaled fractional Brownian motion. We believe that our present method for estimating the relative entropy between EM scheme and the true solution is also available in SDEs driven by additive fractional Brownian motion since the Girsanov transform can also be used, and the Girsanov transform is relatively simple for the case of $H < \frac{1}{2}$.

We will leave these processes mentioned above for a future study.

Funding: This research was funded by the Natural Science Foundation of the Shandong Province of China (ZR202111290596).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The author thanks the Editor, the Associate Editor and the referees for their time and comments, which have greatly improved the paper. We also thank Xing Huang from Tianjin University for his discussion and suggestions for this article.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Kloden, P.E.; Platen, E. *Numerical Solutions of Stochastic Differential Equations*; Springer: Berlin/Heidelberg, Germany, 1992.
2. Yan, L. The Euler scheme with irregular coefficients. *Ann. Probab.* **2002**, *30*, 1172–1194. [[CrossRef](#)]
3. Gyöngy, I.; Rásonyi, M. A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients. *Stoch. Process. Appl.* **2011**, *121*, 2189–2200. [[CrossRef](#)]
4. 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.* **2008**, *48*, 51–59. [[CrossRef](#)]
5. Leobacher, G.; Szölgényi, M. A numerical method for SDEs with discontinuous drift. *BIT Numer. Math.* **2016**, *56*, 151–162. [[CrossRef](#)]
6. Müller-Gronbach, T.; Yaroslavtseva, L. On the performance of the Euler-Maruyama scheme for SDEs with discontinuous drift coefficient. *Ann. L'Institut Henri Poincaré Probab. Stat.* **2020**, *56*, 1162–1178. [[CrossRef](#)]
7. Leobacher, G.; Szölgényi, M. A strong order 1/2 method for multidimensional SDEs with discontinuous drift. *Ann. Appl. Probab.* **2017**, *27*, 2383–2418. [[CrossRef](#)]
8. Leobacher, G.; Szölgényi, M. Convergence of the Euler-Maruyama method for multidimensional SDEs with discontinuous drift and degenerate diffusion coefficient. *Numer. Math.* **2018**, *138*, 219–239. [[CrossRef](#)] [[PubMed](#)]
9. Bao, J.; Huang, X.; Yuan, C. Convergence rate of Euler–Maruyama Scheme for SDEs with Hölder–Dini continuous drifts. *J. Theor. Probab.* **2019**, *32*, 848–871. [[CrossRef](#)]
10. Pamen, O.M.; Taguchi, D. Strong rate of convergence for the Euler–Maruyama approximation of SDEs with Hölder continuous drift coefficient. *Stoch. Process. Appl.* **2017**, *127*, 2542–2559. [[CrossRef](#)]
11. Bao, J.; Huang, X.; Zhang, S. Convergence rate of EM algorithm for SDEs under integrability condition. *arXiv* **2022**, arXiv:2009.04781.
12. Shao, J. Weak convergence of Euler-Maruyama’s approximation for SDEs under integrability condition. *arXiv* **2018**, arXiv:1808.07250.
13. Suo, Y.; Yuan, C.; Zhang, S.-Q. Weak convergence of Euler scheme for SDEs with singular drift. *arXiv* **2020**, arXiv:2005.04631.
14. Villani, C. *Optimal Transport: Old and New*; Springer: Berlin/Heidelberg, Germany, 2009.
15. Sommaruga, G. *Formal Theories of Information: From Shannon to Semantic Information Theory and General Concepts of Information*; Springer: Berlin/Heidelberg, Germany, 2009.
16. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
17. Calin, O.; Udriște, C. *Geometric Modeling in Probability and Statistics*; Springer: Berlin/Heidelberg, Germany, 2014.
18. Zvonkin, A.K. A transformation of the phase space of a diffusion process that removes the drift. *Math. USSR Sb.* **1974**, *93*, 129–149. [[CrossRef](#)]
19. Flandoli, M.; Gubinelli, M.; Priola, E. Flow of diffeomorphisms for SDEs with unbounded Hölder continuous drift. *Bull. Sci. Math.* **2010**, *134*, 405–422. [[CrossRef](#)]
20. Vinod, D.; Cherstvy, A.G.; Metzler, R.; Sokolov, I.M. Time-averaging and nonergodicity of reset geometric Brownian motion with drift. *Phys. Rev. E* **2022**, *106*, 034137. [[CrossRef](#)] [[PubMed](#)]
21. Liang, Y.; Wang, W.; Metzler, R.; Cherstvy, A.G. Anomalous diffusion, nonergodicity, non-Gaussianity, and aging of fractional Brownian motion with nonlinear clocks. *Phys. Rev. E* **2023**, *108*, 034113. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.