Applied Numerical Mathematics 184 (2023) 137–150

Contents lists available at ScienceDirect

Applied Numerical Mathematics

www.elsevier.com/locate/apnum

The backward Euler-Maruyama method for invariant measures of stochastic differential equations with super-linear coefficients ☆

Wei Liu [a,b], Xuerong Mao [c], Yue Wu [c,*]

[a] Department of Mathematics, Shanghai Normal University, Shanghai, 200234, PR China
[b] Lab for Educational Big Data and Policymaking, Shanghai Normal University, Shanghai, 200234, PR China
[c] Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond St, Glasgow, G1 1XH, UK

ARTICLE INFO

Article history:
Received 7 July 2022
Received in revised form 25 September 2022
Accepted 26 September 2022
Available online 29 September 2022

Keywords:
Stochastic differential equation
Stationary measure
Super-linear coefficients
Backward Euler-Maruyama method
Implicit method

ABSTRACT

The backward Euler-Maruyama (BEM) method is employed to approximate the invariant measure of stochastic differential equations, where both the drift and the diffusion coefficient are allowed to grow super-linearly. The existence and uniqueness of the invariant measure of the numerical solution generated by the BEM method are proved and the convergence of the numerical invariant measure to the underlying one is shown. Simulations are provided to illustrate the theoretical results and demonstrate the application of our results in the area of system control.

1. Introduction

Invariant measure is one of essential properties of stochastic differential equations (SDEs), when long time behaviours of SDEs are investigated, such as the persistence for biology and epidemic SDE models in [1,17]. However, the explicit forms of neither the true solutions nor the invariant measures to SDEs are easily found. Therefore, numerical methods become extremely important when SDE models are applied in practice.

For SDEs of the Itô form

$$\begin{cases} \mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, & \text{for } t > 0, \\ X_0 = x \in \mathbb{R}^d, \end{cases} \tag{1}$$

Yuan and Mao [30] studied the numerical invariant measure generated by the Euler-Maruyama (EM) method when both the coefficients $\mu(\cdot)$ and $\sigma(\cdot)$ obey the global Lipschitz condition. Under the same condition on the coefficients, Weng and Liu [25] investigated the numerical approximation to invariant measures of SDEs by the Milstein method. When some non-global Lipschitz terms appear in the coefficients, the backward Euler-Maruyama (BEM) method (also called the semi-implicit

☆ This work is supported by Shanghai Rising-Star Program (Grant No. 22QA1406900), Science and Technology Innovation Plan of Shanghai (Grant No. 20JC1414200), and the National Natural Science Foundation of China (Grant No. 11871343, 11971316, 12271368).

* Corresponding author.
E-mail addresses: weiliu@shnu.edu.cn, lwbvb@hotmail.com (W. Liu), x.mao@strath.ac.uk (X. Mao), yue.wu@strath.ac.uk (Y. Wu).

https://doi.org/10.1016/j.apnum.2022.09.017
0168-9274/

Euler method) and the truncated EM method were employed to handle the super-linearity. Liu and Mao [16] discussed the BEM method for numerically approximating the invariant measure when the one-sided Lipschitz condition was imposed on the draft coefficient $\mu(\cdot)$ but the global Lipschitz condition was still required for the diffusion coefficient $\sigma(\cdot)$. Jiang et al. [11] further studied the stochastic θ-method for this problem and discussed the effects of the different choices of θ on the requirements on the coefficients. When the constraints on the coefficients were further released, Li et al. [15] proposed the truncated EM method to approximate the invariant measure of the underlying SDEs. For SDEs with Markov switching, Bao et al. [3] investigated the numerical approximations to the invariant measures by the EM method and Li et al. [14] studied the similar problem with less constrains on the coefficients by the BEM method, in both of which the r-Wasserstein distance was employed for the discussion on the convergence of numerical invariant measures to underlying ones.

In this paper, we revisit the BEM method and study the numerical approximation to invariant measures of SDEs with both the drift and diffusion coefficients containing super-linear terms. Compared with the existing work by Liu and Mao [16], where only the drift coefficient was allowed to grow super-linearly, our work releases the condition on the diffusion coefficient such that the super-linear terms are also allowed. To achieve such a better result, a different technique is employed in this paper. Briefly speaking, instead of directly forming an iteration for the numerical solution of X_t, we construct an iteration for the numerical approximation of some linear combination of X_t and $\sigma(X_t)$. It should be mentioned that this technique is inspired by Andersson and Kruse [2]. Similar techniques were employed for the studies on the finite time convergence and the stability of the trivial solution of the BEM method in [5] and the stochastic θ method in [23], and for the study on the infinite time convergence of BEM method to the random periodic solution of the SDEs with additive noise in [26]. But, to our best knowledge, there is no existing work on the numerical approximation to invariant measures of SDEs with the super-linear drift and diffusion coefficients by using the BEM method.

Therefore, the result obtained in this paper can be regarded as an extension to [16] and a complement to the study on BEM method in the aspect of numerical invariant measures. In addition, our results can support the application of theorems on stabilisation of SDEs in the distribution sense that were recently developed in [13,29] (see Example 5.2 for the illustration). It should be mentioned that the numerical invariant measure of SDEs obtained in this paper could also assist in approximating the corresponding high-dimensional partial differential equation such as the high-dimensional Fokker–Planck equation. With the help of the neural network architecture, such an approach through a probabilistic representation to learn the solution of some partial differential equation could be quite efficient shown in [8].

Other approaches were also proposed and investigated for approximating invariant measures of SDEs. An incomplete list includes [6,20,22], among many others. The BEM method, as the simplest version of implicit methods, was widely studied for many different types of stochastic equations in [7,19,27,31,32]. We just mention some of them here and refer the readers to the reference therein for more works.

We end this introduction with some discussions on the competition between explicit and implicit methods. For stiff ordinary differential equations, implicit methods are preferred due to its good performance even on a time grid with a large step size as discussed in [24]. But for its stochastic counterpart, explicit methods are also popular as shown in [10,18]. Since many sample paths are usually needed to be simulated in practice, explicit methods have their advantages like simple algorithm structure, easy to implement and no need to solve nonlinear equation systems in each iteration, if simulations are conducted in some finite short intervals. For simulations of long time behaviours of SDEs, implicit methods that pose better stability properties allow large step-sizes and have low total computational costs. More interesting and detailed discussions on this topic can be found in, for example [9,21].

2. Mathematical preliminaries

Let $W:[0,\infty)\times\Omega\to\mathbb{R}^n$ be a standard Wiener process on the probability space $(\Omega,\mathcal{F},\mathbb{P})$, with the filtration defined by $\mathcal{F}_s^t:=\sigma\{W_u-W_v: s<v\le u<t\}$ and $\mathcal{F}^t=\mathcal{F}_0^t=\vee_{0\le s\le t}\mathcal{F}_s^t$. Throughout this paper, we shall use $|\cdot|$ for the Euclidean norm and $\langle\cdot,\cdot\rangle$ for the inner product in the Euclidean space. For a vector u, we define $\|u\|:=\sqrt{\mathbb{E}[|u|^2]}$ and $\|u\|_p:=\sqrt[p]{\mathbb{E}[|u|^p]}$. For a matrix B, $\|B\|_{\mathrm{HS}}$ means its Hilbert-Schmidt norm. In addition, we define $\|B\|:=\sqrt{\mathbb{E}[\|B\|_{\mathrm{HS}}^2]}$ and $\|B\|_p:=\sqrt[p]{\mathbb{E}[\|B\|_{\mathrm{HS}}^p]}$. Denote $a\vee b$ the larger one between scalars a and b, and $a\wedge b$ the smaller one. The family of all probability measures on $\mathbb{R}^d$ is denoted by $\mathcal{P}(\mathbb{R}^d)$. Let $\mathcal{B}(\mathbb{R}^d)$ denote the family of all Borel sets in $\mathbb{R}^d$.

Before we introduce the r-Wasserstein distance, we give a brief introduction to the coupling of probability measures on the same measurable space and refer the readers to Chapter 5 in [12] for detailed discussion on the coupling.

Let μ_1 and μ_2 be probability measures on the same measurable space $(S,\mathcal{S})$. A coupling of μ_1 and μ_2 is a probability measure ν on the product space $(S\times S,\mathcal{S}\times\mathcal{S})$ such that the marginals of ν coincide with μ_1 and μ_2, i.e. $\nu(A\times S)=\mu_1(A)$ and $\nu(S\times A)=\mu_2(A)$, for any $A\in\mathcal{S}$.

The r-Wasserstein distance between $\mu_1,\mu_2\in\mathcal{P}(\mathbb{R}^d)$ for any $r\in(0,1]$ is defined by

$$\mathbb{W}_r(\mu_1,\mu_2)=\inf_{\nu\in\mathcal{C}(\mu_1,\mu_2)}\int\limits_{\mathbb{R}^d\times\mathbb{R}^d}|Y_1-Y_2|^r\nu(\mathrm{d}Y_1,\mathrm{d}Y_2),$$

where $\mathcal{C}(\mu_1,\mu_2)$ denotes the set of all couplings of μ_1 and μ_2.

Given a stochastic process X_t on $(\Omega, \mathcal{F}, \mathbb{P})$, denote the transition probability kernel of X_t by $\mathbb{P}_t(x, B)$ for any $t > 0$ and any $B \in \mathcal{B}$. Sometimes we use δ_x to emphasise the initial value x and denote $\mathbb{P}_t(x, \cdot)$ by $\delta_x \mathbb{P}_t$. A probability measure $\pi(\cdot) \in \mathcal{P}(\mathbb{R}^d)$ is called an *invariant measure* of X_t, if

$$\pi(B) = \int_{\mathbb{R}^d} \mathbb{P}_t(x, B)\pi(\mathrm{d}x)$$

holds for any $t > 0$ and any $B \in \mathcal{B}(\mathbb{R}^d)$.

In this paper, we are interested in the stationary measure of the solution to the $\mathbb{R}^d$-valued SDE of the form

$$\begin{cases} \mathrm{d}X_t = \big[- AX_t + f(X_t)\big]\mathrm{d}t + g(X_t)\mathrm{d}W_t, & \text{for } t > 0, \\ X_0 = x \in \mathbb{R}^d. \end{cases} \tag{2}$$

We separate the drift coefficient into two parts with the emphasis on the negative linear term $-AX_t$, as it could be regarded as stabiliser term following [28]. We impose several assumptions on A, f and g as follows.

Assumption 2.1. The linear operator $A : \mathbb{R}^d \to \mathbb{R}^d$ is self-adjoint and positive definite.

Assumption 2.1 implies the existence of a positive, increasing sequence $(\lambda_i)_{i\in[d]} \subset \mathbb{R}$ such that $0 < \lambda_1 \leq \lambda_2 \leq \ldots \lambda_d$, and of an orthonormal basis $(e_i)_{i\in[d]}$ of $\mathbb{R}^d$ such that $Ae_i = \lambda_i e_i$ for every $i \in [d]$, where $[d] := \{1, \ldots, d\}$.

Assumption 2.2. The mappings $f : \mathbb{R}^d \to \mathbb{R}^d$ and $g : \mathbb{R}^d \to \mathbb{R}^{d\times n}$ are continuous. Moreover, there exists a constant $q \in [1, \infty)$ and a positive L such that

$$|f(u_1) - f(u_2)| \vee \|g(u_1) - g(u_2)\|_{\mathrm{HS}} \leqslant L(1 + |u_1|^{q-1} + |u_2|^{q-1})|u_1 - u_2|,$$

for $u_1, u_2 \in \mathbb{R}^d$.

It is straightforward to derive from Assumption 2.2 that

$$|f(u)| \vee \|g(u)\|_{\mathrm{HS}} \leqslant (2L + (|f(0)| \vee \|g(0)\|_{\mathrm{HS}}))(1 + |u|^q)$$

for $u \in \mathbb{R}^d$.

Assumption 2.3. There exist $c, c_1, c_2 \in (0, \infty)$ and $l_1 \geqslant 2, l_2 \geqslant 4q - 3$ such that

$$2\langle u_1 - u_2, f(u_1) - f(u_2)\rangle + l_1\|g(u_1) - g(u_2)\|^2_{\mathrm{HS}} \leqslant c|u_1 - u_2|^2$$

and

$$2\langle u, f(u)\rangle + l_2\|g(u)\|^2_{\mathrm{HS}} \leq c_1 + c_2|u|^2$$

for all $u, u_1, u_2 \in \mathbb{R}^d$.

It is well known that under these assumptions the *solution* $X_t : [0, \infty) \times \Omega \to \mathbb{R}^d$ to (2) is uniquely determined in [17]. To show the solution to SDE (2) is uniformly bounded in L^p sense, i.e.,

$$\sup_{t\geqslant 0} \|X_t\|^p_p < \infty,$$

and is Hölder-continuous in the temporal variable (see Proposition 3.1 and Proposition 3.2 in Section 3), which are two sufficient conditions to guarantee the existence and uniqueness of invariant measure of SDE (2), we need an additional assumption as imposed below.

Assumption 2.4. Recall that λ_1 is the smallest eigenvalue of A. The constants c and c_2 from Assumption 2.3 satisfy that

$$c \vee c_2 < \lambda_1.$$

Now, we give a brief revisit to the well-known BEM method.

Let us fix an equidistant partition $\mathcal{T}^h := \{jh,\ j \in \mathbb{N}\}$ with stepsize $h \in (0, 1)$. Note that $\mathcal{T}^h$ stretch along the positive real line because we are dealing with an infinite time horizon problem. Then to simulate the solution to (2) starting at 0, the *backward Euler-Maruyama method* on $\mathcal{T}^h$ is given by the recursion

$$\widehat{X}_{(j+1)h} = \widehat{X}_{jh} - Ah\widehat{X}_{(j+1)h} + hf\left(\widehat{X}_{(j+1)h}\right) + g(\widehat{X}_{jh})\Delta W_{jh} \tag{3}$$

for all $j \in \mathbb{N}$, where the initial value $\widehat{X}_0 = x$, and $\Delta W_{jh} := W_{(j+1)h} - W_{jh}$.

The implementation of (3) requires solving a nonlinear equation at each iteration. The well-posedness of the difference equation (3) is proved in the next lemma.

Lemma 2.1. *Let Assumptions 2.3 and 2.4 hold. Then the BEM method is well defined.*

Proof. For any $N \in \mathbb{N}$, rewrite the BEM method (3) into

$$\widehat{X}_{(N+1)h} + Ah\widehat{X}_{(N+1)h} - hf\left(\widehat{X}_{(N+1)h}\right) = \widehat{X}_{Nh} + g(\widehat{X}_{Nh})\Delta W_{Nh}.$$

Define $G(u) = u + Ahu - hf(u)$ for $u \in \mathbb{R}^d$. By Assumption 2.3, we have

$$2\langle u_1 - u_2, f(u_1) - f(u_2)\rangle \leqslant c|u_1 - u_2|^2$$

for all $u_1, u_2 \in \mathbb{R}^d$. Then, it is straightforward to see

$$\langle u_1 - u_2, G(u_1) - G(u_2)\rangle \geqslant \left(1 + \lambda_1 h - \frac{ch}{2}\right)|u_1 - u_2|^2.$$

Due to Assumption 2.4, $1 + \lambda_1 h - ch/2 > 0$ holds for all $h > 0$, which means that $G(\cdot)$ is monotonic. So $G(\cdot)$ has its inverse function $G^{-1}(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ such that for any $N \in \mathbb{N}$

$$\widehat{X}_{(N+1)h} = G^{-1}\left(\widehat{X}_{Nh} + g(\widehat{X}_{Nh})\Delta W_{Nh}\right).$$

That is to say, for any $N \in \mathbb{N}$ the unique $\widehat{X}_{(N+1)h}$ can always be found for the given $\widehat{X}_{Nh} + g(\widehat{X}_{Nh})\Delta W_{Nh}$, which completes the proof. □

To explore the invariant measure of the numerical solution, we introduce some more notations. For any $j \in \mathbb{N}$ and any $B \in \mathcal{B}(\mathbb{R}^d)$, let $\widehat{\mathbb{P}}_{jh}(x, B)$ be the transition probability kernel of $\widehat{X}_{jh}$. A probability measure $\widehat{\pi}(\cdot) \in \mathcal{P}(\mathbb{R}^d)$ is called an *invariant measure* of $\widehat{X}_{jh}$, if

$$\widehat{\pi}(B) = \int_{\mathbb{R}^d} \widehat{\mathbb{P}}_{jh}(x, B)\widehat{\pi}(\mathrm{d}x)$$

holds for any integer $j \in \mathbb{N}$ and any $B \in \mathcal{B}(\mathbb{R}^d)$.

We end up this section by pointing out **the crucial equality** for analysis of the backward Euler-Maruyama in our paper. For any $a, b \in \mathbb{R}^d$, the equality

$$|b|^2 - |a|^2 + |b - a|^2 = 2\langle b - a, b\rangle \tag{4}$$

holds.

3. Some properties of the underlying solution

In this section, we mainly explore properties of the solution to (2) for analysis later.

The first property we will show is the uniform boundedness for the p-th moment of the SDE solution.

Proposition 3.1. *Suppose that Assumptions 2.1 to 2.4 hold. Then, for any $p \in [2, l_2 + 1]$ the solution to (2) satisfies*

$$\sup_{t \geqslant 0} \|X_t\|_p^p < \infty. \tag{5}$$

Proof. Due to Assumption 2.4, we have $c_2 < 2\lambda_1$. Then, let ϵ be a sufficiently small positive number such that $p\lambda_1 - 0.5pc_2 - \epsilon > 0$. By the Itô formula,

$$\mathbb{E}[e^{\epsilon t}|X_t|^p] \leqslant |x|^p + \mathbb{E}\int_0^t e^{\epsilon s}\left[-(p\lambda_1 - \epsilon)|X_s|^p + 0.5p|X_s|^{p-2}\left(2\langle X_s, f(X_s)\rangle + (p-1)\|g(X_s)\|_{\mathrm{HS}}^2\right)\right]\mathrm{d}s.$$

As $p - 1 \leqslant l_2$, Assumption 2.3 indicates

$$\mathbb{E}[e^{\epsilon t}|X_t|^p] \le |x|^p + \mathbb{E}\int_0^t e^{\epsilon s}\Big[-(p\lambda_1-\epsilon)|X_s|^p + 0.5p|X_s|^{p-2}(c_1+c_2|X_s|^2)\Big]ds$$
$$= |x|^p + \mathbb{E}\int_0^t e^{\epsilon s}\Big[-(p\lambda_1-0.5pc_2-\epsilon)|X_s|^p + 0.5pc_1|X_s|^{p-2}\Big]ds$$

Since $p\lambda_1 - 0.5pc_2 - \epsilon > 0$, we know that the polynomial $-(p\lambda_1 - 0.5pc_2 - \epsilon)|X_s|^p + 0.5pc_1|X_s|^{p-2}$ is always bounded by a positive number almost surely for any $|X_s| \in \mathbb{R}$. Denote the upper bound by K. Hence

$$\mathbb{E}[e^{\epsilon t}|X_t|^p] \le |x|^p + \int_0^t e^{\epsilon s}K ds \le |x|^p + (K/\epsilon)e^{\epsilon t},$$

which implies

$$\mathbb{E}[|X_t|^p] \le |x|^p + (K/\epsilon), \quad \forall t \ge 0.$$

Therefore, the proof is completed. □

Following a similar argument as in Proposition 5.4 and 5.5 in [4], we can easily get the following bounds.

Proposition 3.2. *Suppose that Assumptions 2.1 to 2.4 hold, then there exists a positive constant $C_{q,A,f,g}$ which depends on q, d, A, f and g only, such that*

$$\|X_{t_1} - X_{t_2}\| \leqslant C_{q,A,f,g}\big(1 + \sup_{t\geqslant 0}\|X_t\|_{2q}^q\big)|t_2 - t_1|^{\frac{1}{2}}, \tag{6}$$

for all $t_1, t_2 \geqslant 0$. Moreover,

$$\int_{t_1}^{t_2} \big\|A\big(X_s - X_{t_3}\big) + f(X_s) - f(X_{t_3})\big\|ds \leqslant C_{q,A,f,g}\big(1 + \sup_{t\geqslant 0}\|X_t\|_{4q-2}^{2q-1}\big)|t_2 - t_1|^{\frac{3}{2}}, \tag{7}$$

for all $t_3 \in [t_1, t_2]$.

The next theorem states that the underlying solution admits a unique invariant measure. With the help of Propositions 3.1 and 3.2, the following theorem can be proved by following the same approach as the proof of Theorem 2.3 in [3] or Theorem 7.4 in [15]. So we omit the proof here.

Theorem 3.1. *Suppose that Assumptions 2.1 to 2.4 hold. Then the solution to (2) converges in the r-Wasserstein distance to a unique invariant measure $\pi \in \mathcal{P}(\mathbb{R}^d)$ with some exponential rate $\xi_2 > 0$ for any $r \in (0, 1]$, i.e. for any initial value x*

$$\mathbb{W}_r(\delta_x\mathbb{P}_t, \pi) \leqslant Ce^{-r\xi_2 t},$$

where C is a constant independent of t.

4. Main results

In this section we will prove that the BEM method (3) uniquely admits an invariant measure with the help of two lemmas, and show the order of convergence of the invariant measure of the BEM to the invariant measure of our target SDE (2). We present our three main theorems as follows. Proofs of them are postponed, after some more preparations being given.

The first main result in our paper states the existence and uniqueness of the invariant measure of the numerical solution generated by the BEM method.

Theorem 4.1. *Under Assumptions 2.1, 2.3 and 2.4, for any $h \in (0, 1)$ satisfying*

$$h < h^* := \frac{l_2 - 1}{2\lambda_1 - 2c_2} \wedge \frac{l_1 - 1}{2\lambda_1 - 2c},$$

the backward Euler-Maruyama method (3) converges in the r-Wasserstein distance to a unique invariant measure $\widehat{\pi} \in \mathcal{P}(\mathbb{R}^d)$ with some exponential rate $\xi_1 > 0$ on $\mathcal{T}^h$ for any $r \in (0, 1]$.

The next theorem states the strong convergence of the BEM method with the rate of 1/2. This result looks similar to that in [23] by setting $\theta = 1$ there. But, it should be noted that our assumptions are stronger than those in [23]. So the strong convergence is uniform in our case, i.e. the constant C in (8) is independent of t.

Theorem 4.2. *Under Assumption 2.1 to Assumption 2.4 and for h satisfying*

$$h < h^{**} := \frac{l_2 - 1}{2(\lambda_1 - c_2)} \wedge \frac{l_1 - 2}{4(\lambda_1 - c)},$$

there exists a constant C that depends on q, A, f, g and d such that the backward Euler-Maruyama method (3) approximates the true solution of (2) on $\mathcal{T}^h$ with

$$\sup_N \left\| X_{Nh} - \widehat{X}_{Nh} \right\| \leqslant C h^{1/2}. \tag{8}$$

The final main theorem states the convergence of the numerical invariant measure to the underlying one with the rate of 1/2.

Theorem 4.3. *Suppose that all the assumptions in Theorems 4.1 and 4.2 hold, then the numerical invariant measure $\widehat{\pi}$ converges to the underlying invariant measure π in the r-Wasserstein distance, that is for any $h \in (0, h^* \wedge h^{**})$*

$$\mathbb{W}_r(\widehat{\pi}, \pi) = \mathcal{O}\left(h^{r/2}\right)$$

holds for any $r \in (0, 1]$.

4.1. *Two properties of the numerical solution*

The next Lemma claims that there is a uniform bound for the second moment of the numerical solution under necessary assumptions.

Lemma 4.1. *Under Assumptions 2.1, 2.3 and 2.4, for any $h \in (0, 1)$ satisfying*

$$h \leqslant \frac{l_2 - 1}{2(\lambda_1 - c_2)},$$

it holds for the BEM method (3) on $\mathcal{T}^h$ that

$$\|\widehat{X}_{Nh}\|^2 < |x|^2 + \|g(x)\|^2_{\mathrm{HS}} + \frac{c_1}{\lambda_1 - c_2} \tag{9}$$

for all $N \in \mathbb{N}$, where x is the initial data.

Proof. First note that from (4) for any $N \in \mathbb{N}$ we have that

$$|\widehat{X}_{Nh}|^2 - |\widehat{X}_{(N-1)h}|^2 + |\widehat{X}_{Nh} - \widehat{X}_{(N-1)h}|^2 = 2\langle \widehat{X}_{Nh} - \widehat{X}_{(N-1)h}, \widehat{X}_{Nh} \rangle. \tag{10}$$

From (3) we have that

$$2\langle \widehat{X}_{Nh} - \widehat{X}_{(N-1)h}, \widehat{X}_{Nh} \rangle = -2h\langle A\widehat{X}_{Nh}, \widehat{X}_{Nh} \rangle + 2h\langle f(\widehat{X}_{Nh}), \widehat{X}_{Nh} \rangle + 2\langle g\big(\widehat{X}_{(N-1)h}\big)\Delta W_{(N-1)h}, \widehat{X}_{Nh} \rangle. \tag{11}$$

Note that $\mathbb{E}\langle g\big(\widehat{X}_{(N-1)h}\big)\Delta W_{(N-1)h}, \widehat{X}_{(N-1)h} \rangle = 0$.

Taking the expectation of both sides of (11) and making use of Assumption 2.3 give

$$\begin{aligned}
&\|\widehat{X}_{Nh}\|^2 - \|\widehat{X}_{(N-1)h}\|^2 + \|\widehat{X}_{Nh} - \widehat{X}_{(N-1)h}\|^2 = 2\mathbb{E}\langle \widehat{X}_{Nh} - \widehat{X}_{(N-1)h}, \widehat{X}_{Nh} \rangle \\
&\leqslant -2h\mathbb{E}\langle (A - c_2 I)\widehat{X}_{Nh}, \widehat{X}_{Nh} \rangle - l_2 h \|g\big(\widehat{X}_{Nh}\big)\|^2 + 2hc_1 + h\|g\big(\widehat{X}_{(N-1)h}\big)\|^2 + \|\widehat{X}_{Nh} - \widehat{X}_{(N-1)h}\|^2.
\end{aligned}$$

Then cancelling the same term on both sides gives

$$(1 + 2h(\lambda_1 - c_2))\|\widehat{X}_{Nh}\|^2 + l_2 h \left\|g\big(\widehat{X}_{Nh}\big)\right\|^2 \leqslant 2hc_1 + h\left\|g\big(\widehat{X}_{(N-1)h}\big)\right\|^2 + \|\widehat{X}_{(N-1)h}\|^2.$$

Choose h such that $(1 + 2h(\lambda_1 - c_2)) \leqslant l_2$ and let $\alpha := \frac{c_1}{\lambda_1 - c_2}$. Rearranging the terms above gives

$$\big(1 + 2h(\lambda_1 - c_2)\big)\big(\|\widehat{X}_{Nh}\|^2 + h\left\|g\big(\widehat{X}_{Nh}\big)\right\|^2 - \alpha\big) \leqslant \|\widehat{X}_{(N-1)h}\|^2 + h\left\|g\big(\widehat{X}_{(N-1)h}\big)\right\|^2 - \alpha. \tag{12}$$

By iteration, this leads to

$$\|\widehat{X}_{Nh}\|^2 \leqslant \frac{1}{\left(1+2h(\lambda_1-c_2)\right)^N}\left(|x|^2+h\|g(x)\|_{\mathrm{HS}}^2-\alpha\right)+\alpha. \tag{13}$$

Because of Assumption 2.4, the term on the right hand side above can be bounded by $|x|^2+\|g(x)\|_{\mathrm{HS}}^2+\alpha$, which is independent of k and h. □

The next result shows two numerical solutions starting from different initial conditions can be arbitrarily close after sufficiently many iterations.

Lemma 4.2. *Under Assumptions 2.1, 2.3 and 2.4, and let $h \leqslant (l_1-1)/(2\lambda_1-2c)$, define $\widehat{X}_{Nh}$ and $\widehat{Y}_{Nh}$ solutions of the backward Euler-Maruyama scheme on $\mathcal{T}^h$ with different initial values $x, y \in \mathbb{R}^d$, respectively. Then*

$$\|\widehat{X}_{Nh}-\widehat{Y}_{Nh}\| \leqslant \sqrt{1+c}|x-y|e^{-\xi_1 Nh},$$

where $\xi_1=\frac{\lambda_1-c}{1+2(\lambda_1-c)}$.

Proof. Define $D_N := \widehat{X}_{Nh}-\widehat{Y}_{Nh}$. Let us use (4) again, which allows us to examine the following term:

$$\begin{aligned}
&2\mathbb{E}\langle D_N-D_{N-1}, D_N\rangle \\
&=-2h\mathbb{E}\langle AD_N, D_N\rangle+2h\mathbb{E}\langle f(\widehat{X}_{Nh})-f(\widehat{Y}_{Nh}), D_N\rangle \\
&\quad+2\mathbb{E}\langle (g(\widehat{X}_{(N-1)h})-g(\widehat{Y}_{(N-1)h}))\Delta W_{(N-1)h}, D_N\rangle \\
&\leqslant 2h\mathbb{E}\langle(-A+cI)D_N, D_N\rangle-l_1h\|g(\widehat{X}_{Nh})-g(\widehat{Y}_{Nh})\|^2 \\
&\quad+2\mathbb{E}\langle (g(\widehat{X}_{(N-1)h})-g(\widehat{Y}_{(N-1)h}))\Delta W_{(N-1)h}, D_N-D_{N-1}\rangle,
\end{aligned}$$

where we use Assumption 2.3 to deduce the last inequality and the last term is due to

$$\mathbb{E}\langle (g(\widehat{X}_{(N-1)h})-g(\widehat{Y}_{(N-1)h}))\Delta W_{(N-1)h}, D_{N-1}\rangle=0.$$

This leads to

$$(1+2h(\lambda_1-c))\|D_N\|^2+l_1h\left\|g(\widehat{X}_{Nh})-g(\widehat{Y}_{Nh})\right\|^2 \leqslant \|D_{N-1}\|^2+h\left\|g(\widehat{X}_{(N-1)h})-g(\widehat{Y}_{(N-1)h})\right\|^2.$$

Choose h such that $(1+2h(\lambda_1-c)) \leqslant l_1$, then by iteration we have

$$\begin{aligned}
\|D_N\|^2+h\left\|g(\widehat{X}_{Nh})-g(\widehat{Y}_{Nh})\right\|^2 &\leqslant \frac{1}{(1+2h(\lambda_1-c))^N}\left(\|D_0\|^2+h\left\|g(\widehat{X}_0)-g(\widehat{Y}_0)\right\|_{\mathrm{HS}}^2\right) \\
&\leqslant \frac{1+c}{(1+2h(\lambda_1-c))^N}|x-y|^2,
\end{aligned}$$

where we make use of Assumption 2.3 and $h \leqslant 1$ to deduce the last line. Since the fact that $a^N < e^{-(1-a)N}$ for any $a \in (0,1)$ and $\lambda_1 > c$ in Assumption 2.4, the assertion follows. □

4.2. The existence and uniqueness of the numerical invariant measure

Now, we are ready to give the proof of Theorem 4.1.

The proof of Theorem 4.1. Due to the Chebyshev inequality, for any initial value $x \in \mathbb{R}^d$ we obtain that $\left\{\delta_x\widehat{\mathbb{P}}_{jh}\right\}$ is tight, where δ_x is used to emphasise the initial value x, i.e. $\delta_x\widehat{\mathbb{P}}_{jh}=\widehat{\mathbb{P}}_{jh}(x,\cdot)$. Then, a subsequence that converges weakly to an invariant measure $\widehat{\pi} \in \mathcal{P}(\mathbb{R}^d)$ can be extracted. By the Hölder inequality and Lemma 4.2, we can see that for any $r \in (0,1]$

$$\mathbb{W}_r\left(\delta_x\widehat{\mathbb{P}}_{jh}, \delta_y\widehat{\mathbb{P}}_{jh}\right) \leqslant \|\widehat{X}_{jh}-\widehat{Y}_{jh}\|_r^r \leqslant \|\widehat{X}_{jh}-\widehat{Y}_{jh}\|^r \leqslant (1+c)^{r/2}|x-y|^r e^{-r\xi_1 jh}. \tag{14}$$

Then, thanks to Lemma 4.1 and the Kolmogorov-Chapman equation, for any $j, l > 0$ and $r \in (0,2]$ we have

$$
\begin{aligned}
\mathbb{W}_r\left(\delta_x\widehat{\mathbb{P}}_{jh}, \delta_x\widehat{\mathbb{P}}_{(j+l)h}\right) &= \mathbb{W}_r\left(\delta_x\widehat{\mathbb{P}}_{jh}, \delta_x\widehat{\mathbb{P}}_{jh}\widehat{\mathbb{P}}_{lh}\right) \\
&\leqslant \int_{\mathbb{R}^d} \mathbb{W}_r\left(\delta_x\widehat{\mathbb{P}}_{jh}, \delta_y\widehat{\mathbb{P}}_{jh}\right)\widehat{\mathbb{P}}_{lh}(x, \mathrm{d}y) \\
&\leqslant \int_{\mathbb{R}^d} (1+c)^{r/2}|x-y|^r e^{-r\xi_1 jh}\widehat{\mathbb{P}}_{lh}(x, \mathrm{d}y) \\
&\leqslant 2(1+c)^{r/2}\left(|x|^r + \|\widehat{X}_{lh}\|^r\right)e^{-r\xi_1 jh} \\
&\leqslant K_2(r)e^{-r\xi_1 jh},
\end{aligned}
\tag{15}
$$

where

$$
K_2(r) := 2(1+c)^{r/2}\left(|x|^r + \left(|x|^2 + \|g(x)\|_{\mathrm{HS}}^2 + \frac{c_1}{\lambda_1 - c_2}\right)^{r/2}\right).
$$

Now, letting $l \to \infty$ in (15), we have

$$
\mathbb{W}_r\left(\delta_x\widehat{\mathbb{P}}_{jh}, \widehat{\pi}\right) \leqslant K_2(r)e^{-r\xi_1 jh}.
$$

Moreover, we have

$$
\mathbb{W}_r\left(\delta_x\widehat{\mathbb{P}}_{jh}, \widehat{\pi}\right) \to 0, \quad \text{as } j \to \infty,
$$

which guarantees that $\widehat{\pi}$ is the unique invariant measure of $\left\{\delta_x\widehat{\mathbb{P}}_{jh}\right\}$. Now, assume that $\widehat{\pi}_1 \in \mathcal{P}(\mathbb{R}^d)$ is the invariant measure of $\widehat{X}_{jh}$ with the initial value x and $\widehat{\pi}_2 \in \mathcal{P}(\mathbb{R}^d)$ is the invariant measure of $\widehat{X}_{jh}$ with the initial value y, we can see

$$
\mathbb{W}_r(\widehat{\pi}_1, \widehat{\pi}_2) \leqslant \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{W}_r\left(\delta_x\widehat{\mathbb{P}}_{jh}, \delta_y\widehat{\mathbb{P}}_{jh}\right)\nu(\mathrm{d}x, \mathrm{d}y),
$$

for any $x, y \in \mathbb{R}^d$ with $x \neq y$. Therefore, by (14) the BEM method has a unique invariant measure. □

4.3. The uniform strong convergence of the BEM method

The proof of Theorem 4.2 is presented as follows.

The proof of Theorem 4.2. First note that

$$
\begin{aligned}
X_{Nh} &= X_{(N-1)h} - \int_{(N-1)h}^{Nh} AX_s\mathrm{d}s + \int_{(N-1)h}^{Nh} f(X_s)\mathrm{d}s + \int_{(N-1)h}^{Nh} g(X_s)\mathrm{d}W_s \\
&= X_{(N-1)h} - \int_{(N-1)h}^{Nh} A\big(X_s - X_{Nh}\big)\mathrm{d}s - hAX_{Nh} \\
&\quad + \int_{(N-1)h}^{Nh} (f(X_s) - f(X_{Nh}))\mathrm{d}s + hf(X_{Nh}) \\
&\quad + \int_{(N-1)h}^{Nh} \Big(g(X_s) - g(X_{(N-1)h})\Big)\mathrm{d}W_s + g\big(X_{(N-1)h}\big)\Delta W_{(N-1)h}.
\end{aligned}
\tag{16}
$$

Define $e_N := X_{Nh} - \widehat{X}_{Nh}$. Then

$$
\begin{aligned}
2\mathbb{E}\langle e_N - e_{N-1}, e_N\rangle = &-2h\mathbb{E}\langle Ae_N, e_N\rangle + 2h\mathbb{E}\langle f(X_{Nh}) - f(\widehat{X}_{Nh}), e_N\rangle \\
&+ 2\mathbb{E}\Big\langle -\int_{(N-1)h}^{Nh} A(X_s - X_{Nh})\mathrm{d}s, e_N\Big\rangle
\end{aligned}
$$

$$
\begin{aligned}
&+2\mathbb{E}\Big\langle \int_{(N-1)h}^{Nh} (f(X_s)-f(X_{Nh}))\mathrm{d}s, e_N\Big\rangle \\
&+2\mathbb{E}\Big\langle \int_{(N-1)h}^{Nh} (g(X_s)-g\big(X_{(N-1)h}\big))\mathrm{d}W_s, e_N\Big\rangle \\
&+2\mathbb{E}\langle (g(X_{(N-1)h})-g(\widehat{X}_{(N-1)h}))\Delta W_{(N-1)h}, e_N\rangle .
\end{aligned}
$$

Note that for $t\in[(N-1)h, Nh]$, $\int_{(N-1)h}^{t} e_{N-1}^T(g(X_s)-g\big(X_{(N-1)h}\big))\mathrm{d}W_s$ gives a martingale, where a^T represents the transpose of a vector or matrix a. To see it, define the stopping time $\tau_{N,K}:=\inf\{s:|X_s|>K+|X_{(N-1)h}|\}$. Note that $\{\tau_{N,K}\}_{K\in\mathbb{N}}$ is non-descreasing and $\lim_{K\to\infty}\tau_{N,K}=\infty$. Then one can check that $\int_{(N-1)h}^{t\wedge\tau_{N,K}} e_{N-1}^T(g(X_s)-g\big(X_{(N-1)h}\big))\mathrm{d}W_s$ is indeed a martingale. Then we have

$$
\mathbb{E}\Big\langle \int_{(N-1)h}^{Nh} \big(g(X_s)-g(X_{(N-1)h})\big)\mathrm{d}W_s, e_N\Big\rangle = \mathbb{E}\Big\langle \int_{(N-1)h}^{Nh} \big(g(X_s)-g(X_{(N-1)h})\big)\mathrm{d}W_s, e_N-e_{N-1}\Big\rangle.
$$

By Young's inequality

$$
2ab\leqslant \epsilon^2 a^2+\frac{b^2}{\epsilon^2},\quad \forall a,b>0,
$$

and Assumption 2.3, we are able to choose $\epsilon_0^2:=h(\lambda_1-c)/2$ such that

$$
\begin{aligned}
2\mathbb{E}\langle e_N-e_{N-1}, e_N\rangle \leqslant{}& 2h\mathbb{E}\langle(-A+cI)e_N, e_N\rangle - hl_1\|g(X_{Nh})-g(\widehat{X}_{Nh})\|^2 \\
&+2\epsilon_0^2\|e_N\|^2+\frac{1}{\epsilon_0^2}\Big\|-\int_{(N-1)h}^{Nh} A(X_s-X_{Nh})\mathrm{d}s\Big\|^2 \\
&+\frac{1}{\epsilon_0^2}\Big\|\int_{(N-1)h}^{Nh} (f(X_s)-f(X_{Nh}))\mathrm{d}s\Big\|^2 \\
&+2\Big\|\int_{(N-1)h}^{Nh} (g(X_s)-g(X_{(N-1)h}))\mathrm{d}W_s\Big\|^2 \\
&+2h\|g(X_{(N-1)h})-g(\widehat{X}_{(N-1)h})\|^2+\|e_N-e_{N-1}\|^2 .
\end{aligned}
$$

By Proposition 3.2, we know there exists a constant C depending on q, A, f and g such that

$$
\begin{aligned}
&\Big\|-\int_{(N-1)h}^{Nh} A\big(X_s-X_{Nh}\big)\mathrm{d}s\Big\|^2+\Big\|\int_{(N-1)h}^{Nh} \Big(f(X_s)-f(X_{Nh})\Big)\mathrm{d}s\Big\|^2 \\
&\leqslant Ch^3\Big(1+\sup_{s\geqslant 0}\|X_s\|_{4q-2}^{2q-1}\Big):=\beta h^3 .
\end{aligned}
$$

Besides, by the Itô isometry and the Hölder continuity of X in temporal variable as shown in Proposition 3.2 (reusing C above),

$$
2\Big\|\int_{(N-1)h}^{Nh} (g(X_s)-g(X_{(N-1)h}))\mathrm{d}W_s\Big\|^2 \leqslant \frac{2cC\big(1+\sup_{s\geqslant 0}\|X_s\|_{2q}^{q}\big)}{l_1}h^2 := \widehat{c}h^2 .
$$

Note that β is bounded because of Proposition 3.2. Define $G_N=: g(X_{Nh})-g(\widehat{X}_{Nh})$. Then from (4) and the estimate above we have that

$$
\begin{aligned}
&\|e_N\|^2-\|e_{N-1}\|^2+hl_1\|G_N\|^2 = 2\mathbb{E}\langle e_N-e_{N-1}, e_N\rangle - \|e_N-e_{N-1}\|^2+hl_1\|G_N\|^2 \\
&\leqslant 2h\mathbb{E}\langle(-A+cI)e_N, e_N\rangle+2\epsilon_0^2\|e_N\|^2+\frac{\beta h^3}{\epsilon_0^2}+\widehat{c}h^2+2h\|G_{N-1}\|^2 .
\end{aligned}
$$

Define $\widehat{\alpha} := \frac{2\beta+c(\lambda_1-c)}{(\lambda_1-c)^2}h$. Since that $1+h(\lambda_1-c) \leqslant l_1/2$, then the inequality above can be rearranged to

$$\Big(1+h(\lambda_1-c)\Big)\big(\|e_N\|^2+2h\|G_N\|^2-\widehat{\alpha}\big) \leqslant \|e_{N-1}\|^2+2h\|G_{N-1}\|^2-\widehat{\alpha}.$$

By iteration we have

$$\|e_N\|^2+2h\|G_N\|^2 \leqslant \Big(1-\frac{1}{1+h(\lambda_1-c)^N}\Big)\frac{2\beta+c(\lambda_1-c)}{(\lambda_1-c)^2}h,$$

because $X_0=\widehat{X}_0$. Finally due to Assumption 2.4, we have $\|e_N\|^2 \leqslant \frac{2\beta+c(\lambda_1-c)}{(\lambda_1-c)^2}h$. Then the assertion follows. □

4.4. Convergence of the numerical invariant measure to the underlying counterpart

Now we are ready to show the last main theorem.

The proof of Theorem 4.3. It is clear to see that

$$\mathbb{W}_r\left(\widehat{\pi},\pi\right) \leqslant \mathbb{W}_r\left(\widehat{\pi},\delta_x\widehat{\mathbb{P}}_{jh}\right)+\mathbb{W}_r\left(\delta_x\mathbb{P}_{jh},\pi\right)+\mathbb{W}_r\left(\delta_x\mathbb{P}_{jh},\delta_x\widehat{\mathbb{P}}_{jh}\right).$$

Thanks to Theorems 4.1 and 3.1, the convergences of $\widehat{\mathbb{P}}_{jh}$ to $\widehat{\pi}$ and $\mathbb{P}_{jh}$ to π yield

$$\mathbb{W}_r\left(\widehat{\pi},\delta_x\widehat{\mathbb{P}}_{jh}\right) \leqslant K_2(r)e^{-r\xi_1 jh} \quad \text{and} \quad \mathbb{W}_r\left(\delta_x\mathbb{P}_{jh},\pi\right) \leqslant Ce^{-r\xi_2 jh},$$

where C is a genetic constant in this proof that may be different from line to line. Then applying Theorem 4.2 gives the final assertion. □

5. Numerical examples

In this section, two numerical examples are presented. Example 5.1 is used to illustrate that the BEM method admits a unique invariant measure, which then converges to the underlying one. In Example 5.2, we discuss the application of our numerical method in the stabilisation of SDEs in the distribution sense.

Example 5.1. Consider a scalar mean-reverting type model with super-linear coefficients

$$dX_t=\left(b-\alpha X_t-\beta X_t^3\right)\mathrm{d}t+\sigma X_t^2\mathrm{d}W_t, \quad X_0=x.$$

By setting $b=1$, $\alpha=1$, $\beta=2$ and $\sigma=1$, it is not hard to see that all the assumptions are satisfied. Therefore, according to our theorems there exists a unique invariant measure for the BEM method. One thousand sample paths are simulated with $X_0=5$ and $h=0.01$, which are then used to construct empirical density functions at different time points. It is clear to see from the left plot in Fig. 1 that the shapes of empirical density functions at $t=0.1$, $t=0.3$ and $t=0.5$ are quite different but the ones at $t=4$ and $t=10$ are much more similar, which indicates the existence of the invariant measure. From the right plot in Fig. 1, we can see the empirical density functions at the same time point $t=90$ but with different initial values $-5, 5, 15$ are quite close to each other, which indicates uniqueness of the invariant measure. To measure the difference between empirical density functions at consecutive time points $t=ih$ and $t=(i+1)h$ for $i=0,1,\ldots$, the Kolmogorov–Smirnov (K-S) test is employed to test a sequence of hypotheses that

$$H_0: \quad \text{Two samples at } t=ih \text{ and } t=(i+1)h \text{ are from the same distribution,}$$

$$H_1: \quad \text{Two samples at } t=ih \text{ and } t=(i+1)h \text{ are from different distributions,}$$

for $i=0,1,\ldots 200$. It can be observed from the upper plot in Fig. 2 that as time gets large the differences between empirical density functions at consecutive time points vanish, which indicates the existence of the invariant measure for the numerical solution. The lower plot in Fig. 2 also confirms this conclusion as the p values are quite close to 1 as time advances.

Now we turn to our second example, which could be regarded as an illustration of the application of our results in the system control problem. To make it clear, we brief the problem as follows.

In the very recent works [13,29], the authors discussed the design of some controllers to stabilise some SDEs that originally are not stable in distribution. To be more precise, for some unstable SDE (i.e. not stable in the distribution sense)

$$\begin{cases}\mathrm{d}X_t=f(X_t)\mathrm{d}t+g(X_t)\mathrm{d}W_t, & \text{for } t>0,\\ X_0=x\in\mathbb{R}^d,\end{cases}$$

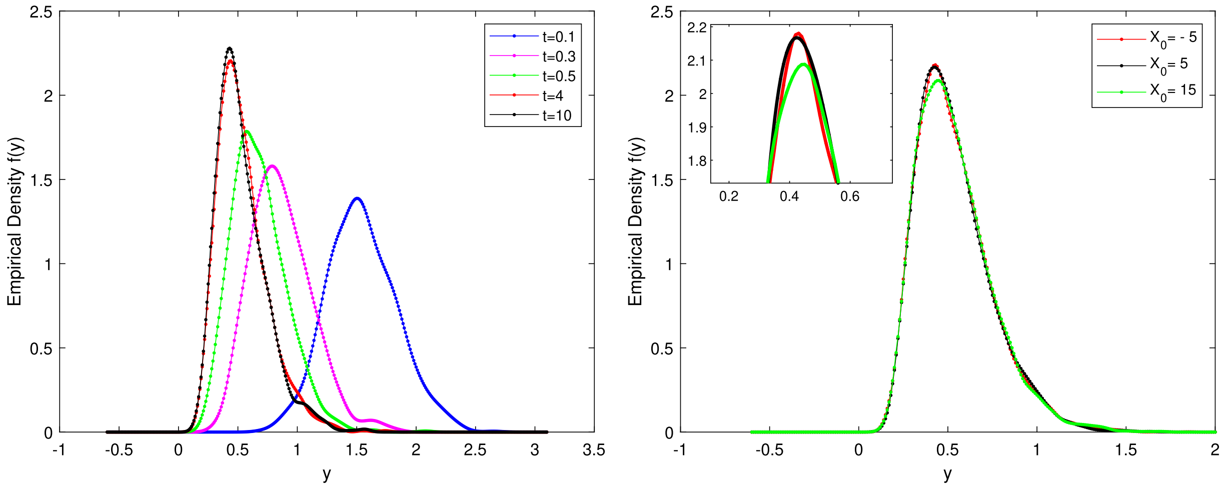

Fig. 1. Left: Empirical density functions at different time points. Right: Empirical density functions at t=90 with different initial values. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

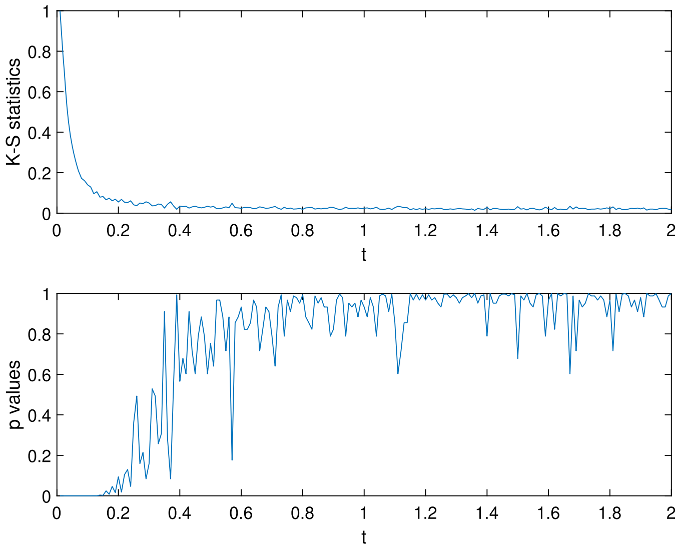

Fig. 2. K-S tests for samples at consecutive time points for Example 5.1.

the authors in those two works used some past state $X_{t-\tau}$, where the small enough constant $\tau > 0$ represents the time delay, to design a controller $AX_{t-\tau}$ such that the controlled system

$$\begin{cases} dX_t = \left[f(X_t) - AX_{t-\tau}\right]dt + g(X_t)dW_t, & \text{for } t > 0, \\ X_0 = x \in \mathbb{R}^d \end{cases} \tag{17}$$

is stable in distribution. In their works, the authors proposed the method to design the controller and proved theoretically that the controlled system is indeed stable in distribution. But in practice, numerical methods are always required for the applications of those theorems, as the explicit forms of the true solutions of stochastic systems can hardly be found, not to mention the explicit forms of the invariant distributions. Therefore, trusted numerical methods are essential for demonstrating those theorems in [13,29] and displaying the shapes of the invariant distributions. By saying trusted numerical methods, we mean those methods that have been proved to be able to approximate the underlying true invariant distributions. And this is what we proved in this paper for the BEM method.

It is clear that if the $AX_{t-\tau}$ is replaced by AX_t in the controlled system (17), then it looks exactly like the SDE (2) studied in this paper. Since our results obtained in this paper do not include delay terms in the equations, we use AX_t as the controller in our Example 5.2. In future, we are going to work out the numerical invariant measures for some stochastic delay differential equations.

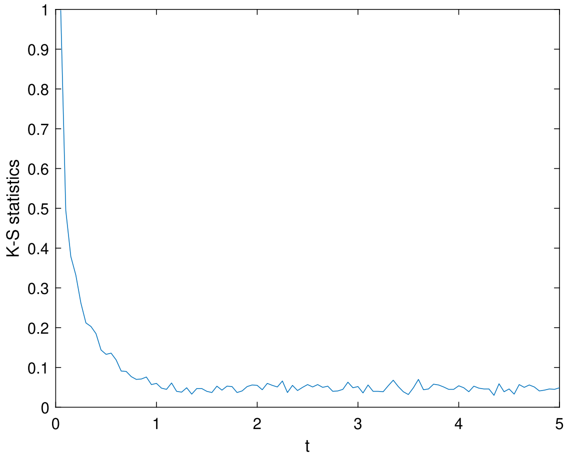

Fig. 3. K-S tests for samples at consecutive time points for Example 5.2.

Example 5.2. Consider a two dimensional SDE

$$\begin{cases} \mathrm{d}X_{t,1} = \big[10 + 2X_{t,1} - X_{t,2}\big]\mathrm{d}t + \big[0.5 + 0.1X_{t,2}\big]\mathrm{d}W_{t,1}, \\ \mathrm{d}X_{t,2} = \big[5 + X_{t,1} + 3X_{t,2} - X_{t,2}^3\big]\mathrm{d}t + \big[0.3 + 0.1\big(X_{t,1} + X_{t,2}^2\big)\big]\mathrm{d}W_{t,2}, \\ X_0 = (5, 5) \end{cases}$$

which is unstable in distribution for any initial data. According to theorems in [13,29], one can design a controller

$$A = \begin{pmatrix} -5 & 0 \\ -2 & -4 \end{pmatrix}$$

such that the controlled system

$$\begin{cases} \mathrm{d}X_{t,1} = \big[10 - 3X_{t,1} - X_{t,2}\big]\mathrm{d}t + \big[0.5 + 0.1X_{t,2}\big]\mathrm{d}W_{t,1} \\ \mathrm{d}X_{t,2} = \big[5 - X_{t,1} - 3X_{t,2} - X_{t,2}^3\big]\mathrm{d}t + \big[0.3 + 0.1\big(X_{t,1} + X_{t,2}^2\big)\big]\mathrm{d}W_{t,2} \end{cases} \tag{18}$$

is stable in the distribution sense. But, in practice one may further ask the question: what does the unique distribution look like?

To answer the question, one may turn to our results in this paper. Since it is not hard to check that coefficients of (18) satisfy the requirements, we can regard the numerical invariant distribution generated by the BEM method as a trusted approximate to the underlying one. 1000 sample paths generated by the BEM method with the step size of 0.05 are simulated. Similar to Example 5.1, the K-S test is applied to illustrate that the distributions generated by the BEM method indeed tend to a unique one as the time advances. The asymptotic behaviour of the K-S statistics in Fig. 3 confirms it. More importantly, Fig. 3 also indicates that one does not have to simulate sample paths for long time to see the invariant distribution, as the differences between empirical distributions decay to zero in a quite fast way. Therefore, to see the shape of the unique distribution of (18), it is sufficient to use the empirical distribution of the numerical solutions generated by the BEM method at relatively small time point. Fig. 4 displays the empirical density function of the solution $(X_{t,1}, X_{t,2})$ at $t = 4$, which could be used to answer the question raised in Example 5.2. In practice, one can further use some non-parametric and parametric approaches to find out what the distribution is and the estimated values of parameters of it.

To end up this section, we give a short informal discussion on the potential application of our results in numerical approximates to stationary Fokker-Planck equations. It is well known that if there exists a unique invariant measure π for the SDE

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t,$$

then the true π can be found by solving the following partial differential equation (PDE)

$$\frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2\pi) - \frac{\partial}{\partial x}(\mu\pi) = 0 \quad \text{with the condition} \quad \int_{x\in\mathbb{R}^d} \pi(x)\mathrm{d}x = 1. \tag{19}$$

The numerical invariant measure $\widehat{\pi}$ obtained in this paper can be regarded as a good estimator for the solution of (19), as the convergence of $\widehat{\pi}$ to π actually has been proved in this paper. For example, Fig. 4 indeed display the solution to

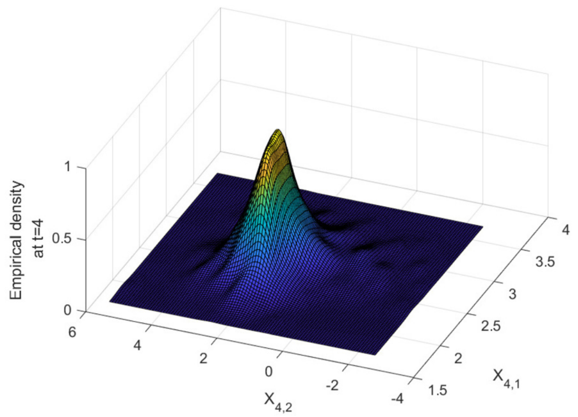

Fig. 4. Empirical density function at $t = 4$.

the stationary Fokker-Planck equation that is corresponding to the SDE (18). Fokker-Planck equations and their stationary forms are of importance on their own rights in various problems arising in chemical reactions, statistical physics, and fluid mechanics, however, their practical use is hindered by the curse of dimensionality. Based on the success of [8], it is expected that under some smart design of neural network architecture through the probabilistic representation and the numerical simulation, one may establish an effective stochastic framework for the PDE (19), which could avoid the curse of dimensionality.

6. Conclusion and future research

In this paper, we revisited the classical BEM method and showed the existence and uniqueness of its invariant measure when both the drift and the diffusion coefficients are allowed to contain some super-linear terms. In addition, the convergence of the numerical invariant measure to its underlying counterpart was also proved. Numerical simulations were provided to demonstration our theorems and their potential applications in system controls.

As we mentioned occasionally in this paper, there are many works that have not been done in this area. One definitely interesting work is to extend the results in this paper to stochastic delay differential equations, for which the concept of invariant measure is quite different from the case of SDEs. Another question that is worth to be considered is the numerical invariant measure of hybrid SDEs with super-linear drift and diffusion coefficients, in which the switches among different modes would play important roles in the stability in distribution of the whole system.

Declaration of competing interest

The Authors declare that there is no conflict of interest.

Acknowledgements

The authors would like to thank the editor and the reviewers for the very careful and valuable comments and suggestions, which helped to improve this work a lot.

References

[1] E. Allen, Modeling with Itô Stochastic Differential Equations, vol. 22, Springer Science & Business Media, 2007.
[2] A. Andersson, R. Kruse, Mean-square convergence of the BDF2-Maruyama and backward Euler schemes for SDE satisfying a global monotonicity condition, BIT Numer. Math. 57 (2017) 21–53.
[3] J. Bao, J. Shao, C. Yuan, Approximation of invariant measures for regime-switching diffusions, Potential Anal. 44 (2016) 707–727.
[4] W.J. Beyn, E. Isaak, R. Kruse, Stochastic C-stability and B-consistency of explicit and implicit Euler-type schemes, J. Sci. Comput. 67 (2016) 955–987.
[5] Z. Chen, S. Gan, Convergence and stability of the backward Euler method for jump–diffusion SDEs with super-linearly growing diffusion and jump coefficients, J. Comput. Appl. Math. 363 (2020) 350–369.
[6] W. Fang, M.B. Giles, Adaptive Euler–Maruyama method for SDEs with non-globally Lipschitz drift, in: International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Springer, 2016, pp. 217–234.
[7] J. Gao, H. Liang, S. Ma, Strong convergence of the semi-implicit Euler method for nonlinear stochastic Volterra integral equations with constant delay, Appl. Math. Comput. 348 (2019) 385–398.
[8] J. Han, A. Jentzen, W. E, Solving high-dimensional partial differential equations using deep learning, Proc. Natl. Acad. Sci. USA 115 (2018) 8505–8510.
[9] D.J. Higham, Stochastic ordinary differential equations in applied and computational mathematics, IMA J. Appl. Math. 76 (2011) 449–474.
[10] M. Hutzenthaler, A. Jentzen, P.E. Kloeden, Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients, Ann. Appl. Probab. 22 (2012) 1611–1641.
[11] Y. Jiang, L. Weng, W. Liu, Stationary distribution of the stochastic theta method for nonlinear stochastic differential equations, Numer. Algorithms 83 (2020) 1531–1553.

[12] D.A. Levin, Y. Peres, Markov Chains and Mixing Times, American Mathematical Society, Providence, RI, 2017.
[13] X. Li, W. Liu, Q. Luo, X. Mao, Stabilisation in distribution of hybrid stochastic differential equations by feedback control based on discrete-time state observations, Automatica 140 (2022) 110210.
[14] X. Li, Q. Ma, H. Yang, Y. Chenggui, The numerical invariant measure of stochastic differential equations with Markovian switching, SIAM J. Numer. Anal. 56 (2018) 1435–1455.
[15] X. Li, X. Mao, G. Yin, Explicit numerical approximations for stochastic differential equations in finite and infinite horizons: truncation methods, convergence in p th moment and stability, IMA J. Numer. Anal. 39 (2019) 847–892.
[16] W. Liu, X. Mao, Numerical stationary distribution and its convergence for nonlinear stochastic differential equations, J. Comput. Appl. Math. 276 (2015) 16–29.
[17] X. Mao, Stochastic Differential Equations and Applications, Elsevier, 2007.
[18] X. Mao, The truncated Euler–Maruyama method for stochastic differential equations, J. Comput. Appl. Math. 290 (2015) 370–384.
[19] X. Mao, L. Szpruch, Strong convergence rates for backward Euler–Maruyama method for non-linear dissipative-type stochastic differential equations with super-linear diffusion coefficients, Stoch. Int. J. Probab. Stoch. Process. 85 (2013) 144–171.
[20] J.C. Mattingly, A.M. Stuart, D.J. Higham, Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise, Stoch. Process. Appl. 101 (2002) 185–232.
[21] G.N. Milstein, M.V. Tretyakov, Stochastic Numerics for Mathematical Physics, vol. 456, Springer, 2004.
[22] D. Talay, Second-order discretization schemes of stochastic differential systems for the computation of the invariant law, Stoch. Int. J. Probab. Stoch. Process. 29 (1990) 13–36.
[23] X. Wang, J. Wu, B. Dong, Mean-square convergence rates of stochastic theta methods for SDEs under a coupled monotonicity condition, BIT Numer. Math. 60 (2020) 759–790.
[24] G. Wanner, E. Hairer, Solving Ordinary Differential Equations II, vol. 375, Springer Berlin Heidelberg New York, 1996.
[25] L. Weng, W. Liu, Invariant measures of the Milstein method for stochastic differential equations with commutative noise, Appl. Math. Comput. 358 (2019) 169–176.
[26] Y. Wu, Backward Euler–Maruyama method for the random periodic solution of a stochastic differential equation with a monotone drift, J. Theor. Probab. (2022) 1–18.
[27] Y. Xiao, M. Song, M. Liu, Convergence and stability of the semi-implicit Euler method with variable step size for a linear stochastic pantograph differential equation, Int. J. Numer. Anal. Model. 8 (2011) 214–225.
[28] A. Yevik, H. Zhao, Numerical approximations to the stationary solutions of stochastic differential equations, SIAM J. Numer. Anal. 49 (2011) 1397–1416.
[29] S. You, L. Hu, J. Lu, X. Mao, Stabilization in distribution by delay feedback control for hybrid stochastic differential equations, IEEE Trans. Autom. Control 67 (2021) 971–977.
[30] C. Yuan, X. Mao, Stability in distribution of numerical solutions for stochastic differential equations, Stoch. Anal. Appl. 22 (2004) 1133–1150.
[31] C. Zhang, Y. Xie, Backward Euler–Maruyama method applied to nonlinear hybrid stochastic differential equations with time-variable delay, Sci. China Math. 62 (2019) 597–616.
[32] S. Zhou, Strong convergence and stability of backward Euler–Maruyama scheme for highly nonlinear hybrid stochastic differential delay equation, Calcolo 52 (2015) 445–473.