Numerical stationary distribution and its convergence for nonlinear stochastic differential equations

Wei Liu[a,*], Xuerong Mao[b]

[a] *Loughborough University, Department of Mathematical Sciences, Loughborough, Leicestershire, LE11 3TU, UK*
[b] *University of Strathclyde, Department of Mathematics and Statistics, Glasgow, G1 1XH, UK*

Abstract

To avoid finding the stationary distributions of stochastic differential equations by solving the nontrivial Kolmogorov-Fokker-Planck equations, the numerical stationary distributions are used as the approximations instead. This paper is devoted to approximate the stationary distribution of the underlying equation by the Backward Euler-Maruyama method. Currently existing results [21, 31, 33] are extended in this paper to cover larger range of nonlinear SDEs when the linear growth condition on the drift coefficient is violated.

Keywords: the backward Euler-Maruyama method, nonlinear SDEs, numerical stationary distribution, weak convergence

1. Introduction

Stochastic differential equations (SDEs) have been widely used in modelling uncertain phenomena in many areas [18, 25]. However, due to the difficulty to find general explicit solutions to non-linear SDEs, numerical approximations have been attracting a lot of attention in recent decades [16, 24]. One aspect of the numerical analyses for SDEs focuses on asymptotic properties of approximations, among which the asymptotic stability particularly has been interesting to many researchers. There are different types of stabilities, and the almost sure

*Corresponding author

Email addresses: `w.liu@lboro.ac.uk;lwbvb@hotmail.com` (Wei Liu), `x.mao@strath.ac.uk` (Xuerong Mao)

stability and the moment stability are the two that have been discussed a lot. We mention some of the works [2, 3, 4, 6, 13, 11, 19, 26, 30] and the references therein. Briefly, those two stabilities are defined by that for any given initial value the solution will decay to the trivial solution (in the sense of moment or almost surely) as time tends to infinity.

However, those stabilities mentioned above sometimes are too strong. In some cases, the solution will not decay to the trivial solution but oscillate as time advances. In this situation, the underlying solution may have a stationary distribution. Stationary distribution of SDEs has many modelling applications, for example in the dynamic of species population [22] and in epidemiology [5]. One way to find the stationary distribution is by solving the Kolmogorov-Fokker-Planck equation. But this is nontrivial. Another way is to approximate it using the stationary distribution obtained from some numerical solution. To follow this approach, one first needs to show the existence and uniqueness of the stationary distribution for the numerical solution. Then the numerical stationary distribution needs to be shown to converge to the underlying one.

The second author's series papers [21, 33, 31] are devoted to numerical stationary distributions of stochastic differential equations. In those series papers, the explicit Euler–Maruyama (EM) method was used due to the simple structure and moderate computational cost [8]. However, the explicit EM method has its own restriction, as mentioned in [14], it may not converge to the true solution of the super-linear-coefficient SDEs even in finite time. Therefore, both the drift coefficient and the diffusion coefficient were required to be global Lipschitz in the series papers. Those restrictions exclude many highly non-linear models, for example [1, 5, 7] and the references therein.

In this paper, we propose the Backward Euler-Maruyama (BEM) method as the approximation. The BEM method, which is a drift implicit scheme, has been broadly investigated and shown better at dealing with the highly non-linear SDEs in both finite time convergence problems and asymptotic problems. We mention some works [9, 10, 11, 12, 19, 26, 28] here and the references therein. In this paper, we are going to investigate the existence and uniqueness of the numerical stationary distribution of the BEM method and the convergence of it to the underlying stationary distribution. One of our key contributions is that we release the global Lipschitz condition on the drift coefficient by assuming the one-sided Lipschitz condition instead, but we still require the global Lipschitz condition on the diffusion coefficient. And this restriction is due to the techniques employed in the proofs in Section 3, in which the diffusion coefficient needs to be bounded by some linear term. We mention that some papers on the finite time convergence discussed certain type of SDE models with the non-global Lipschitz diffusion coefficient [28]. Therefore, one of the open problems is that can we use some other methods to approximate the stationary distributions of some classes of SDE models without the global Lipschitz on the diffusion coefficient?

This paper is constructed as follows. We first brief the method, definitions, conditions on the SDEs as well as other mathematical preliminaries in Section 2. Then, we propose the coefficients related sufficient conditions for the existence and uniqueness of the numerical stationary distribution in Section 3.1. Under the same conditions, the stationary distribution of the underlying solution is presented in Section 3.2. The convergence of the numerical stationary distribution is proved in Section 3.3. In Section 4, we demonstrate the theoretical results by some numerical simulations. We conclude this paper and discuss some future research in Section 5.

2. Mathematical Preliminaries

Throughout this paper, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions (that is, it is right continuous and increasing while $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets). Let $|\cdot|$ denote the Euclidean norm in $\mathbb{R}^d$. The transpose of a vector or matrix, M, is denoted by M^T and the trace norm of a matrix, M, is denoted by $|M| = \sqrt{\text{trace}(M^T M)}$. If M is a square matrix, the smallest and largest eigenvalues of M are denoted by $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$, respectively.

Let $f, g : \mathbb{R}^d \to \mathbb{R}^d$. To keep symbols simple, let $B(t)$ be a scalar Brownian motion. The results in this paper can be extended to the case of multi-dimensional Brownian motions. We consider the d-dimensional stochastic differential equation of Itô type

$$dx(t) = f(x(t))dt + g(x(t))dB(t) \tag{2.1}$$

with initial value $x(0) = x_0$.

We first assume that the drift coefficient satisfies the local Lipschitz condition and the diffusion coefficient satisfies the global Lipschitz condition.

Condition 2.1. *For any $h > 0$, there exists a constant $C_h > 0$ such that*

$$|f(x) - f(y)|^2 \leq C_h |x - y|^2,$$

for any $x, y \in \mathbb{R}^d$ with $\max(|x|, |y|) \leq h$.

Condition 2.2. *There exists a constant $\bar{K_2} > 0$ such that*

$$|g(x) - g(y)|^2 \leq \bar{K_2} |x - y|^2,$$

for any $x, y \in \mathbb{R}^d$.

We further impose the following condition on the drift coefficient.

Condition 2.3. *Assume there exist a symmetric positive-definite matrix $Q \in \mathbb{R}^{d \times d}$ and a constant $\bar{K_1} \in \mathbb{R}$ such that*

$$(x - y)^T Q(f(x) - f(y)) \leq \bar{K_1}(x - y)^T Q(x - y),$$

for any $x, y \in \mathbb{R}^d$.

From Condition 2.2 and 2.3, it is easy to see that for any $x \in \mathbb{R}^d$

$$x^T Q f(x) \leq K_1 x^T Q x + \alpha_1, \tag{2.2}$$

and

$$|g(x)|^2 \leq K_2 |x|^2 + \alpha_2, \tag{2.3}$$

with $K_2, \alpha_1, \alpha_2 > 0$ and $K_1 \in \mathbb{R}$.

2.1. The Backward Euler-Maruyama Method

The backward Euler-Maruyama method (BEM), also called the semi-implicit Euler method, to SDE (2.1) is defined by

$$X_{k+1} = X_k + f(X_{k+1})\Delta t + g(X_k)\Delta B_k, \quad X_0 = x(0) = x_0, \tag{2.4}$$

where $\Delta B_k = B(t_{k+1}) - B(t_k)$ is a Brownian motion increment and $t_k = k\Delta t$. We refer to [16, 24] for more details in numerical methods for SDEs.

Lemma 2.4. *Let Conditions 2.1, 2.2, 2.3 hold and $\Delta t < 0.5|\bar{K}_1|^{-1}$, the BEM solution (2.4) is well defined.*

Many papers have discussed the existence and uniqueness of the BEM solution (2.4), we therefore refer to [19, 20] for the proof of the lemma above. From now on, we always assume $\Delta t < 0.5|\bar{K}_1|^{-1}$.

It is useful to write (2.4) as

$$X_{k+1} - f(X_{k+1})\Delta t = X_k + g(X_k)\Delta B_k.$$

Define a function $G : \mathbb{R}^d \to \mathbb{R}^d$ by $G(x) = x - f(x)\Delta t$. Then G has its inverse function $G^{-1} : \mathbb{R}^d \to \mathbb{R}^d$. Moreover, the BEM (2.4) can be represented as

$$X_{k+1} = G^{-1}(X_k + g(X_k)\Delta B_k). \tag{2.5}$$

Lemma 2.5. *Let Conditions 2.1, 2.2 and 2.3 hold, then*

$$\mathbb{P}(X_{k+1} \in B \big| X_k = x) = \mathbb{P}(X_1 \in B \big| X_0 = x)$$

for any Borel set $B \subset \mathbb{R}^d$.

Proof. If $X_k = x$ and $X_0 = x$, by (2.4) we see

$$X_{k+1} - f(X_{k+1})\Delta t = x + g(x)\Delta B_k,$$

and

$$X_1 - f(X_1)\Delta t = x + g(x)\Delta B_0.$$

Because ΔB_k and ΔB_0 are identical in probability law, comparing the two equations above, we know that $X_{k+1} - f(X_{k+1})$ and $X_1 - f(X_1)\Delta t$ have the identical probability law. Then, due to Lemma 2.4, we have that X_{k+1} and X_1 are identical in probability law under $X_k = x$ and $X_0 = x$. Therefore, the assertion holds. ∎

To prove Theorem 2.7, we cite the following classical result (see, for example, Lemma 9.2 on page 87 of [18]).

Lemma 2.6. *Let $h(x,\omega)$ be a scalar bounded measurable random function of x, independent of $\mathcal{F}_s$. Let ζ be an $\mathcal{F}_s$-measurable random variable. Then*

$$\mathbb{E}(h(\zeta,\omega)\big|\mathcal{F}_s) = H(\zeta),$$

where $H(x) = \mathbb{E}h(x,\omega)$.

For any $x \in \mathbb{R}^d$ and any Borel set $B \subset \mathbb{R}^d$, define

$$\mathbb{P}(x,B) := \mathbb{P}(X_1 \in B\big|X_0 = x) \text{ and } \mathbb{P}_k(x,B) := \mathbb{P}(X_k \in B\big|X_0 = x).$$

Theorem 2.7. *The BEM solution (2.4) is a homogeneous Markov process with transition probability kernel $\mathbb{P}(x,B)$.*

Proof. The homogeneous property follows Lemma 2.5, so we only need to show the Markov property. Define

$$Y_{k+1}^x = G^{-1}(x + g(x)\Delta B_k),$$

for $x \in \mathbb{R}^d$ and $k \geq 0$. By (2.5) we know that $X_{k+1} = Y_{k+1}^{X_k}$. Let $\mathcal{G}_{t_{k+1}} = \sigma\{B(t_{k+1}) - B(t_k)\}$. Clearly, $\mathcal{G}_{t_{k+1}}$ is independent of $\mathcal{F}_{t_k}$. Moreover, Y_{k+1}^x depends completely on the increment $B(t_{k+1}) - B(t_k)$, so is $\mathcal{G}_{t_{k+1}}$-measurable. Hence, Y_{k+1}^x is independent of $\mathcal{F}_{t_k}$. Applying Lemma 2.6 with $h(x,\omega) = I_B(Y_{k+1}^x)$, we compute that

$$\mathbb{P}(X_{k+1} \in B\big|\mathcal{F}_{t_k}) = \mathbb{E}(I_B(X_{k+1})\big|\mathcal{F}_{t_k}) = \mathbb{E}\left(I_B(Y_{k+1}^{X_k})\big|\mathcal{F}_{t_k}\right) = \mathbb{E}\left(I_B(Y_{k+1}^x)\right)\big|_{x=X_k}$$
$$= \mathbb{P}(x,B)\big|_{x=X_k} = \mathbb{P}(X_k,B) = \mathbb{P}(X_{k+1} \in B\big|X_k).$$

The proof is complete. ∎

Therefore, we see that $\mathbb{P}(\cdot,\cdot)$ is the one-step transition probability and $\mathbb{P}_k(\cdot,\cdot)$ is the k-step transition probability, both of which are induced by the BEM solution.

We state a simple version of the discrete-type Gronwall inequality in the next Lemma (see, for example, [17]).

Lemma 2.8. *Let $\{u_n\}$ and $\{w_n\}$ be nonnegative sequences and α be a nonnegative constant. If*

$$u_n \le \alpha + \sum_{k=0}^{n-1} u_k w_k \quad \text{for } n \ge 0,$$

then

$$u_n \le \alpha \exp\left(\sum_{k=0}^{n-1} w_k\right).$$

2.2. Stationary Distributions

Denote the family of all probability measures on $\mathbb{R}^d$ by $\mathcal{P}(\mathbb{R}^d)$. Define by $\mathbb{L}$ the family of mappings $F : \mathbb{R}^d \to \mathbb{R}$ satisfying

$$|F(x) - F(y)| \le |x - y| \quad \text{and} \quad |F(x)| \le 1,$$

for any $x, y \in \mathbb{R}^d$. For $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}(\mathbb{R}^d)$, define metric $d_{\mathbb{L}}$ by

$$d_{\mathbb{L}}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{F \in \mathbb{L}} \left| \int_{\mathbb{R}^d} F(x)\mathbb{P}_1(dx) - \int_{\mathbb{R}^d} F(x)\mathbb{P}_2(dx) \right|.$$

The weak convergence of probability measures can be illustrated in terms of metric $d_{\mathbb{L}}$ [15]. That is, a sequence of probability measures $\{\mathbb{P}_k\}_{k\ge1}$ in $\mathcal{P}(\mathbb{R}^d)$ converge weakly to a probability measure $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d)$ if and only if

$$\lim_{k\to\infty} d_{\mathbb{L}}(\mathbb{P}_k, \mathbb{P}) = 0.$$

Then we define the stationary distribution for $\{X_k\}_{k\ge0}$ by using the concept of weak convergence.

Definition 2.9. *For any initial value $x \in \mathbb{R}^d$ and a given step size $\Delta t > 0$, $\{X_k\}_{k\ge0}$ is said to have a stationary distribution $\Pi_{\Delta t} \in \mathcal{P}(\mathbb{R}^d)$ if the k-step transition probability measure $\mathbb{P}_k(x, \cdot)$ converges weakly to $\Pi_{\Delta t}(\cdot)$ as $k \to \infty$ for every $x \in \mathbb{R}^d$, that is*

$$\lim_{k\to\infty} \left(\sup_{F\in\mathbb{L}} |\mathbb{E}(F(X_k)) - E_{\Pi_{\Delta t}}(F)| \right) = 0,$$

where

$$E_{\Pi_{\Delta t}}(F) = \int_{\mathbb{R}^d} F(y)\Pi_{\Delta t}(dy).$$

In [31], the authors provided the following three assumptions and proved that under those assumptions the Euler–Maruyama solution of the stochastic differential equation has a unique stationary distribution. We observe that the three assumptions are very general and actually can cover many other types of one-step numerical methods including the BEM method. This is because that, in their proofs (Theorem 3.1 in [31]), only the three assumptions were required but not the structure of the numerical method. Therefore, for any one-step numerical solution that is a homogeneous Markov process with a proper transition probability kernel and satisfies the three assumptions, Theorem 3.1 in [31] always holds. To keep the paper self contained, we state the assumptions and the theorem as follows.

Assumption 2.10. *For any $\varepsilon > 0$ and $x_0 \in \mathbb{R}^d$, there exists a constant $R = R(\varepsilon, x_0) > 0$ such that*

$$\mathbb{P}(|X_k^{x_0}| \geq R) < \varepsilon, \quad \textit{for any } k \geq 0.$$

Assumption 2.11. *For any $\varepsilon > 0$ and any compact subset K of $\mathbb{R}^d$, there exists a positive integer $k^* = k^*(\varepsilon, K)$ such that*

$$\mathbb{P}(|X_k^{x_0} - X_k^{y_0}| < \varepsilon) \geq 1 - \varepsilon, \quad \textit{for any } k \geq k^* \textit{ and any } (x_0, y_0) \in K \times K.$$

Assumption 2.12. *For any $\varepsilon > 0$, $n \geq 1$ and any compact subset K of $\mathbb{R}^d$, there exists a $R = R(\varepsilon, n, K) > 0$ such that*

$$\mathbb{P}\left(\sup_{0 \leq k \leq n} |X_k^{x_0}| \leq R\right) > 1 - \varepsilon, \quad \textit{for any } x_0 \in K.$$

Theorem 2.13. *Under Assumptions 2.10, 2.11 and 2.12, the BEM solution $\{X_k\}_{k \geq 0}$ has a unique stationary distribution $\Pi_{\Delta t}$.*

We refer the readers to Theorem 3.1 in [31] for the proof.

However, those three assumptions are not easy to check as they are not directly related to the drift and diffusion coefficients of the underlying SDEs. In the next section, we will provide some coefficient-related sufficient conditions for those assumptions. It should be noted that those sufficient conditions are method related, which makes them more constraint.

3. Main Results

This section is divided into three parts. In the first subsection, we propose three lemmas that are sufficient conditions for Assumption 2.10, 2.11 and 2.12. Then by Theorem 2.13, we see that the BEM solution has a unique stationary distribution. In the second subsection, we prove that given the same conditions in the three lemmas the underlying solution has a unique stationary distribution as well. The last subsection is devoted to the convergence of the numerical stationary distribution to the underlying stationary distribution.

3.1. Sufficient Conditions for the Numerical Stationary Distribution

Many works have discussed the second moment boundedness of the BEM solution in finite time, we only mention a few of them here [16, 20] and references therein. It should be emphasized that, comparing with techniques employed in Lemma 3.1, weaker conditions and more complicated techniques have already been developed in the existing literature. But those weaker conditions may not be sufficient for other lemmas in this paper. To keep the conditions consistent in this paper and to make the paper self-contained, we brief the following lemma. Without confusion, in some of the proofs we omit the superscript and simply denote $X_k^{x_0}$ by X_k.

Lemma 3.1. *Given Conditions 2.1, 2.2 and 2.3, the second moment of the BEM solution (2.4) obeys*

$$\mathbb{E}\left(\sup_{0\le k\le n+1}|X_k|^2\right) \le q\left(|x_0|^2 + C_1(n+1)\left(2\alpha_1\Delta t + \alpha_2\Delta t + 2\sqrt{2\alpha_2\Delta t/\pi}\right)\right) \times \exp\left(q(n+1)C_1\left(1 + K_2\Delta t + 2(\sqrt{K_2}+\sqrt{\alpha_2})\sqrt{2\Delta t/\pi}\right)\right)$$

for any integer $n \ge 1$, where $C_1 = (1-2|K_1|\Delta t)^{-1}$ and $q = \lambda_{\max}(Q)/\lambda_{\min}(Q)$.

Proof. Fix any initial value $X(0) = x_0 \in \mathbb{R}^d$, from (2.4) we see that

$$X_{k+1}^T Q X_{k+1} = X_{k+1}^T Q(X_k + g(X_k)\Delta B_k) + X_{k+1}^T Q f(X_{k+1})\Delta t.$$

Since Q is a symmetric positive-definite matrix, by the Cholesky decomposition there exists a unique lower triangular matrix L such that $Q = LL^T$. Then by the elementary inequality and (2.2) we have

$$\begin{aligned}
X_{k+1}^T Q X_{k+1} &\le \frac{1}{2}|X_{k+1}^T L|^2 + \frac{1}{2}|L^T(X_k + g(X_k)\Delta B_k)|^2 + (K_1 X_{k+1}^T Q X_{k+1} + \alpha_1)\Delta t \\
&\le \frac{1}{2} X_{k+1}^T Q X_{k+1} + \frac{1}{2}[X_k^T Q X_k + g^T(X_k) Q g(X_k)|\Delta B_k|^2 + 2X_k^T Q g(X_k)\Delta B_k] \\
&\quad + (K_1 X_{k+1}^T Q X_{k+1} + \alpha_1)\Delta t.
\end{aligned}$$

This implies

$$X_{k+1}^T Q X_{k+1} \le C_1(X_k^T Q X_k + g^T(X_k) Q g(X_k)|\Delta B_k|^2 + 2X_k^T Q g(X_k)\Delta B_k) + 2C_1\alpha_1\Delta t,$$

where $C_1 = (1 - 2|K_1|\Delta t)^{-1}$. Taking sum on both sides gives

$$\begin{aligned}
X_{k+1}^T Q X_{k+1} &\le X_0^T Q X_0 + (C_1 - 1)\sum_{i=0}^{k}(X_i^T Q X_i) + 2\alpha_1 C_1(k+1)\Delta t \\
&\quad + C_1 \sum_{i=0}^{k}(2X_i^T Q g(X_i)\Delta B_i + g^T(X_i) Q g(X_i)|\Delta B_i|^2). \qquad (3.1)
\end{aligned}$$

It is not difficult to show that

$$\mathbb{E}\left(\sup_{0\le k\le n}\left(\sum_{i=0}^{k} g^T(X_i) Q g(X_i)|\Delta B_i|^2\right)\right) \le \Delta t \lambda_{\max}(Q)\sum_{i=0}^{n}\mathbb{E}(K_2|X_i|^2 + \alpha_2),$$

and

$$\begin{aligned}
&\mathbb{E}\left(\sup_{0\le k\le n}\left(\sum_{i=0}^{k} X_i^T Q g(X_i)\Delta B_i\right)\right) \le \lambda_{\max}(Q)\mathbb{E}\left(\sum_{i=0}^{n}|X_i||g(X_i)||\Delta B_i|\right) \\
&\le \lambda_{\max}(Q)(\sqrt{K_2} + \sqrt{\alpha_2})\sqrt{2\Delta t/\pi}\sum_{i=0}^{n}\mathbb{E}(|X_i|^2) + \lambda_{\max}(Q)\sqrt{2\alpha_2\Delta t/\pi}(n+1),
\end{aligned}$$

where $\mathbb{E}|\Delta B_i| = \sqrt{2\Delta t/\pi}$ is used. Therefore, taking supremum and expectation on both sides of (3.1) yields

$$\begin{aligned}
\mathbb{E}\left(\sup_{0\le k\le n+1}|X_k|^2\right) &\le \frac{\lambda_{\max}(Q)}{\lambda_{\min}(Q)}\Bigg(|x_0|^2 + C_1(n+1)\left(2\alpha_1\Delta t + \alpha_2\Delta t + 2\sqrt{2\alpha_2\Delta t/\pi}\right) \\
&\quad + C_1\left(1 + K_2\Delta t + 2(\sqrt{K_2} + \sqrt{\alpha_2})\sqrt{2\Delta t/\pi}\right)\sum_{i=0}^{n}\mathbb{E}\left(\sup_{0\le k\le i}|X_k|^2\right)\Bigg).
\end{aligned}$$

Then, using the discrete-type Gronwall inequality stated in Lemma 2.8 we see the assertion holds. ∎

From Lemma 3.1, by the Chebyshev inequality we can conclude that Assumption 2.12 holds under Conditions 2.1, 2.2 and 2.3.

Lemma 3.2. *Let (2.2) and (2.3) hold. If, for the same Q in (2.2), there exists a positive constant D such that for any $x \in \mathbb{R}^d$*

$$\frac{g^T(x)Qg(x)}{D + x^TQx} - \frac{2|x^TQg(x)|^2}{(D + x^TQx)^2} \leq K_3 + \frac{P_3(|x|)}{(D + x^TQx)^2} \tag{3.2}$$

where K_3 is a constant with $K_1 + 0.5K_3 < 0$ and $P_3(|x|)$ is a polynomial of $|x|$ with degree 3, then there exists a pair of constants $(p^*, \Delta t^*)$ with $p^* \in (0, 1)$ and $\Delta t^* \in (0, 0.5|K_1|^{-1})$ such that for any $p \in (0, p^*)$ and any $\Delta t \in (0, \Delta t^*)$ the BEM solution (2.4) has the property that for any $k \geq 1$

$$\mathbb{E}|X_k|^p \leq q(D^{p/2} + |X_0|^p - 2C_3'(p(K_1 + 0.5K_3))^{-1})$$

where $q = \lambda_{\max}(Q)/\lambda_{\min}(Q)$, and C_3' depends on K_1, α_1, D, Q and p.

Proof. Set $C_1 = (1 - 2K_1\Delta t)^{-1}$, from the proof of Lemma 3.1 we have that

$$\begin{aligned}
&DC_1 + X_{k+1}^TQX_{k+1} \\
&\leq DC_1 + C_1(X_k^TQX_k + 2X_k^TQg(X_k)\Delta B_k + g^T(X_k)Qg(X_k)|\Delta B_k|^2 + 2\alpha_1\Delta t) \\
&\leq C_1(D + X_k^TQX_k)(1 + \zeta_k),
\end{aligned}$$

where $\zeta_k = (D+X_k^TQX_k)^{-1}(2X_k^TQg(X_k)\Delta B_k+g^T(X_k)Qg(X_k)|\Delta B_k|^2+2\alpha_1\Delta t)$. Clearly $\zeta_k > -1$. For any $p \in (0, 1)$, thanks to the fundamental inequality that

$$(1 + u)^{p/2} \leq 1 + \frac{p}{2}u + \frac{p(p-2)}{8}u^2 + \frac{p(p-2)(p-4)}{2^3 \times 3!}u^3, \quad u > -1, \tag{3.3}$$

we see that

$$\begin{aligned}
&\mathbb{E}((D + X_{k+1}^TQX_{k+1})^{p/2}\big|\mathcal{F}_{k\Delta t}) \\
&\leq C_1^{p/2}(D + X_k^TQX_k)^{p/2}\mathbb{E}\left(1 + \frac{p}{2}\zeta_k + \frac{p(p-2)}{8}\zeta_k^2 + \frac{p(p-2)(p-4)}{2^3 \times 3!}\zeta_k^3|\mathcal{F}_{k\Delta t}\right).
\end{aligned} \tag{3.4}$$

Since ΔB_k is independent of $\mathcal{F}_{k\Delta t}$, we have that $\mathbb{E}(\Delta B_k \big| \mathcal{F}_{k\Delta t}) = \mathbb{E}(\Delta B_k) = 0$ and $\mathbb{E}(|\Delta B_k|^2 \big| \mathcal{F}_{k\Delta t}) = \mathbb{E}(|\Delta B_k|^2) = \Delta t$. Then

$$
\begin{aligned}
&\mathbb{E}(\zeta_k \big| \mathcal{F}_{k\Delta t}) \\
&= \mathbb{E}\left((D + X_k^T Q X_k)^{-1}(2X_k^T Q g(X_k)\Delta B_k + g^T(X_k)Qg(X_k)|\Delta B_k|^2 + 2\alpha_1 \Delta t)\big| \mathcal{F}_{k\Delta t}\right) \\
&= (D + X_k^T Q X_k)^{-1}(2X_k^T Q g(X_k)\mathbb{E}(\Delta B_k \big| \mathcal{F}_{k\Delta t}) + g^T(X_k)Qg(X_k)\mathbb{E}(|\Delta B_k|^2 \big| \mathcal{F}_{k\Delta t}) \\
&\quad + 2\alpha_1 \Delta t) \\
&= (D + X_k^T Q X_k)^{-1}(g^T(X_k)Qg(X_k)\Delta t + 2\alpha_1 \Delta t). \qquad (3.5)
\end{aligned}
$$

Using the facts that $\mathbb{E}(|\Delta B_k|^{2i}) = (2i-1)!!\Delta t^i$ and $\mathbb{E}((\Delta B_k)^{2i+1}) = 0$, similarly we get that

$$
\begin{aligned}
\mathbb{E}(\zeta_k^2 \big| \mathcal{F}_{k\Delta t}) &= (D + X_k^T Q X_k)^{-2}(4|X_k^T Q g(X_k)|^2 \Delta t + 3|g^T(X_k)Qg(X_k)|^2 \Delta t^2 + 4\alpha_1^2 \Delta t^2 \\
&\quad + 4\alpha_1 g^T(X_k)Qg(X_k)\Delta t^2) \\
&\geq (D + X_k^T Q X_k)^{-2}(4|X_k^T Q g(X_k)|^2 \Delta t), \qquad (3.6)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}(\zeta_k^3 \big| \mathcal{F}_{k\Delta t}) &= (D + X_k^T Q X_k)^{-3}(15|g^T(X_k)Qg(X_k)|^3 \Delta t^3 + 12\alpha_1^2 g^T(X_k)Qg(X_k)\Delta t^3 \\
&\quad + 8\alpha_1^3 \Delta t^3 + 24\alpha_1 |X_k^T Q g(X_k)|^2 \Delta t^2 + 18\alpha_1 |g^T(X_k)Qg(X_k)|^2 \Delta t^3 \\
&\quad + 36|X_k^T Q g(X_k)|^2 g^T(X_k)Qg(X_k)\Delta t^2) \\
&\leq C_2 \Delta t^2, \qquad (3.7)
\end{aligned}
$$

where C_2 is a constant dependent on K_2, α_1, α_2, $\lambda_{\max}(Q)$, $\lambda_{\min}(Q)$ and D.

Substituting (3.5), (3.6) and (3.7) back to (3.4) yields

$$
\begin{aligned}
&\mathbb{E}((D + X_{k+1}^T Q X_{k+1})^{p/2} \big| \mathcal{F}_{k\Delta t}) \\
&\leq C_1^{p/2}(D + X_k^T Q X_k)^{p/2}\mathbb{E}\Bigg(1 + \frac{p}{2}\left(\frac{g^T(X_k)Qg(X_k)}{D + X_k^T Q X_k} - \frac{2|X_k^T Q g(X_k)|^2}{(D + X_k^T Q X_k)^2}\right)\Delta t \\
&\quad + \frac{p^2}{2}\frac{|X_k^T Q g(X_k)|^2}{(D + X_k^T Q X_k)^2}\Delta t + \frac{p(p-2)(p-4)}{2^3 \times 3!}C_2 \Delta t^2\Bigg) + C_3 \Delta t
\end{aligned}
$$

where C_3 depends on K_1, α_1, D, $\lambda_{\min}(Q)$ and $\lambda_{\max}(Q)$. Considering the fraction

$$
\frac{(D + X_k^T Q X_k)^{p/2} P_3(|X_k|)}{(D + X_k^T Q X_k)^2},
$$

for $0 < p < 1$ the highest degree of $|X_k|$ in the numerator is $p+3$, which is smaller than the highest degree of $|X_k|$ in the denominator. Thus, for any $|X_k| \in \mathbb{R}$ there exists a positive constant upper bound for the fraction. By (3.2), we have

$$\begin{aligned} &\mathbb{E}((D + X_{k+1}^T Q X_{k+1})^{p/2} \big| \mathcal{F}_{k\Delta t}) \\ &\leq C_1^{p/2} (D + X_k^T Q X_k)^{p/2} (1 + \frac{p}{2} K_3 \Delta t + \frac{p^2}{2} K_2 q \Delta t + C_2' \Delta t^2) + C_3' \Delta t \end{aligned}$$

where C_2' depends on C_2 and p, and C_3' depends on C_3 and p. Taking expectation on both sides, we have

$$\begin{aligned} &\mathbb{E}((D + X_{k+1}^T Q X_{k+1})^{p/2}) \\ &\leq C_1^{p/2} (1 + \frac{p}{2} K_3 \Delta t + \frac{p^2}{2} K_2 q \Delta t + C_2' \Delta t^2) \mathbb{E}((D + X_k^T Q X_k)^{p/2}) + C_3' \Delta t. \end{aligned} \tag{3.8}$$

Set $\varepsilon = 0.5|K_1 + 0.5K_3|$, choose p^* sufficiently small such that $p^* K_2 q \leq 0.5\varepsilon$, then choose Δt^* sufficiently small such that for $p \in (0, p^*)$ and $\Delta t \in (0, \Delta t^*)$ we have

$$C_1 = (1 - 2K_1 \Delta t)^{-1} \geq 1 - pK_1 \Delta t - C_4 \Delta t^2 > 0, \tag{3.9}$$

where C_4 is a positive constant dependent on K_1 and p. By further reducing Δt^* such that for any $\Delta t \in (0, \Delta t^*)$

$$C_2' \Delta t < \frac{1}{8} p\varepsilon, \quad C_4 \Delta t < \frac{1}{4} \varepsilon, \quad |p(K_1 + \frac{1}{4}) \Delta t| < \frac{1}{2}.$$

Now using these three inequalities and (3.9), we derive from (3.8) that

$$\mathbb{E}((D + X_{k+1}^T Q X_{k+1})^{p/2}) \leq \frac{1 + 0.5p(K_3 + 0.5\varepsilon)\Delta t}{1 - p(K_1 + 0.25\varepsilon)\Delta t} \mathbb{E}((D + X_k^T Q X_k)^{p/2}) + C_3' \Delta t. \tag{3.10}$$

Considering the estimate that for any $\kappa \in [-0.5, 0.5]$

$$(1 - \kappa)^{-1} = 1 + \kappa + \kappa^2 \sum_{i=0}^{\infty} \kappa^i \leq 1 + \kappa + \kappa^2 \sum_{i=0}^{\infty} 0.5^i = 1 + \kappa + 2\kappa^2,$$

by further reducing Δt^* we see that for $\Delta t \in (0, \Delta t)$

$$4p(K_1 + \frac{1}{4}\varepsilon)^2 \Delta t + (K_3 + \frac{1}{2}\varepsilon)(p(K_1 + \frac{1}{4}\varepsilon)\Delta t + 2(p(K_1 + \frac{1}{4}\varepsilon)\Delta t)^2) < \varepsilon.$$

Then (3.10) indicates that

$$\begin{aligned}\mathbb{E}((D+X_{k+1}^TQX_{k+1})^{p/2}) &\le (1+0.5p(K_3+0.5\varepsilon)\Delta t)(1+p(K_1+0.25\varepsilon)\Delta t \\ &\quad +2(p(K_1+0.25\varepsilon)\Delta t)^2)\mathbb{E}((D+X_k^TQX_k)^{p/2})+C_3'\Delta t \\ &\le (1+p(K_1+0.5K_3+\varepsilon)\Delta t)\mathbb{E}((D+X_k^TQX_k)^{p/2})+C_3'\Delta t.\end{aligned}$$

By iteration, we obtain that

$$\begin{aligned}\mathbb{E}((D+X_{k+1}^TQX_{k+1})^{p/2}) &\le (1+p(K_1+0.5K_3+\varepsilon)\Delta t)^{k+1}(D+X_0^TQX_0)^{p/2} \\ &\quad +\frac{1-(1+p(K_1+0.5K_3+\varepsilon)\Delta t)^{k+1}}{1-(1+p(K_1+0.5K_3+\varepsilon)\Delta t)}C_3'\Delta t.\end{aligned}$$

Since $(1+p(K_1+0.5K_3+\varepsilon)\Delta t)\in(0,1)$ for any $p\in(0,p^*)$ and $\Delta t\in(0,\Delta t^*)$, we see that

$$\mathbb{E}((D+X_{k+1}^TQX_{k+1})^{p/2})\le(D+X_0^TQX_0)^{p/2}-2(p(K_1+0.5K_3))^{-1}C_3'.$$

Because Q is a symmetric positive-definite matrix, the assertion holds. ∎

From Lemma 3.2, we can conclude that Assumption 2.10 holds for sufficiently small Δt.

Now we are investigating the sufficient condition for Assumption 2.11. The techniques used in the proof of Lemma 3.3 are similar to those in Lemma 3.2.

Lemma 3.3. *Let Conditions 2.1, 2.2 and 2.3 hold. Assume that, for the same Q in (2.3),*

$$\frac{(g(x)-g(y))^TQ(g(x)-g(y))}{(x-y)^TQ(x-y)}-\frac{2|(x-y)^TQ(g(x)-g(y))|^2}{|(x-y)^TQ(x-y)|^2}\le K_4,\ \forall x,y\in\mathbb{R}^d \text{ with } x\neq y, \tag{3.11}$$

where K_4 is constant with $\bar{K}_1+0.5K_4<0$. Then for any two different initial values $x,y\in\mathbb{R}^d$, the BEM solution (2.4) has the property that for any $k\ge1$ there are sufficiently small Δt^* and p^* such that for any pair of Δt and p with $\Delta t\in(0,\Delta t^*)$ and $p\in(0,p^*)$

$$\mathbb{E}(|X_k^x-X_k^y|^p)\le q(1+0.5p(\bar{K}_1+0.5K_4)\Delta t)^k\mathbb{E}(|x-y|^p),$$

where $q=\lambda_{\max}(Q)/\lambda_{\min}(Q)$. Therefore, Assumption 2.11 follows.

Proof. From (2.4) we have

$$X_{k+1}^x-X_{k+1}^y=X_k^x-X_k^y+(f(X_{k+1}^x)-f(X_{k+1}^y))\Delta t+(g(X_k^x)-g(X_k^y))\Delta B_k.$$

Then, in the similar manner as the proof of Lemma 3.1, we see that

$$
\begin{aligned}
&(X^x_{k+1}-X^y_{k+1})^TQ(X^x_{k+1}-X^y_{k+1})\\
&\le (1-2\bar K_1\Delta t)^{-1}((X^x_k-X^y_k)^TQ(X^x_k-X^y_k)+2(X^x_k-X^y_k)^TQ(g(X^x_k)-g(X^y_k))\Delta B_k\\
&+(g(X^x_k)-g(X^y_k))^TQ(g(X^x_k)-g(X^y_k))|\Delta B_k|^2).
\end{aligned}
$$

Set

$$
\eta_k=\frac{2(X^x_k-X^y_k)^TQ(g(X^x_k)-g(X^y_k))\Delta B_k+(g(X^x_k)-g(X^y_k))^TQ(g(X^x_k)-g(X^y_k))|\Delta B_k|^2}{(X^x_k-X^y_k)^TQ(X^x_k-X^y_k)}
$$

we can have

$$
(X^x_{k+1}-X^y_{k+1})^TQ(X^x_{k+1}-X^y_{k+1})\le\frac{(X^x_k-X^y_k)^TQ(X^x_k-X^y_k)}{1-2\bar K_1\Delta t}(1+\eta_k).
$$

Taking conditional expectation on both sides and using the fundamental inequality (3.3), for any $p\in(0,1)$ we have that

$$
\begin{aligned}
&\mathbb{E}(|(X^x_{k+1}-X^y_{k+1})^TQ(X^x_{k+1}-X^y_{k+1})|^{p/2}\big|\mathcal{F}_{k\Delta t})\\
&\le\left|\frac{(X^x_k-X^y_k)^TQ(X^x_k-X^y_k)}{1-2\bar K_1\Delta t}\right|^{p/2}\mathbb{E}\left(1+\frac{p}{2}\eta_k+\frac{p(p-2)}{8}\eta_k^2+\frac{p(p-2)(p-4)}{2^3\times 3!}\eta_k^3|\mathcal{F}_{k\Delta t}\right).
\end{aligned}
\tag{3.12}
$$

It is not difficult to show that

$$
\mathbb{E}(\eta_k\big|\mathcal{F}_{k\Delta t})=\frac{(g(X^x_k)-g(X^y_k))^TQ(g(X^x_k)-g(X^y_k))}{(X^x_k-X^y_k)^TQ(X^x_k-X^y_k)}\Delta t,
$$

$$
\mathbb{E}(\eta_k^2\big|\mathcal{F}_{k\Delta t})\ge\frac{4|(X^x_k-X^y_k)^TQ(g(X^x_k)-g(X^y_k))|^2}{|(X^x_k-X^y_k)^TQ(X^x_k-X^y_k)|^2}\Delta t,
$$

and

$$
\mathbb{E}(\eta_k^3\big|\mathcal{F}_{k\Delta t})\le C_5\Delta t^2,
$$

where C_5 depends on K_2, $\lambda_{\min}(Q)$ and $\lambda_{\max}(Q)$. Together with (3.11) we derive from (3.12) that

$$
\begin{aligned}
&\mathbb{E}(|(X^x_{k+1}-X^y_{k+1})^TQ(X^x_{k+1}-X^y_{k+1})|^{p/2}\big|\mathcal{F}_{k\Delta t})\\
&\le\left|\frac{(X^x_k-X^y_k)^TQ(X^x_k-X^y_k)}{1-2\bar K_1\Delta t}\right|^{p/2}\left(1+\frac{p}{2}\frac{(g(X^x_k)-g(X^y_k))^TQ(g(X^x_k)-g(X^y_k))}{(X^x_k-X^y_k)^TQ(X^x_k-X^y_k)}\Delta t\right.\\
&\left.+\frac{p(p-2)}{8}\frac{4|(X^x_k-X^y_k)^TQ(g(X^x_k)-g(X^y_k))|^2}{|(X^x_k-X^y_k)^TQ(X^x_k-X^y_k)|^2}\Delta t+\frac{p(p-2)(p-4)}{2^3\times 3!}C_5\Delta t^2\right)\\
&\le\left|\frac{(X^x_k-X^y_k)^TQ(X^x_k-X^y_k)}{1-2\bar K_1\Delta t}\right|^{p/2}\left(1+\frac{p}{2}K_4\Delta t+\frac{p^2}{2}\bar K_2q\Delta t+\frac{p(p-2)(p-4)}{2^3\times 3!}C_5\Delta t^2\right)
\end{aligned}
$$

In the same way as in the proof of Lemma 3.2, we can choose sufficiently small Δt^* and p^* such that for any $p \in (0, p^*)$ and $\Delta t \in (0, \Delta t^*)$

$$\begin{aligned}&\mathbb{E}(|(X_{k+1}^x - X_{k+1}^y)^T Q (X_{k+1}^x - X_{k+1}^y)|^{p/2})\\ &\le (1 + 0.5p(\bar{K_1} + 0.5K_4)\Delta t)\mathbb{E}(|(X_k^x - X_k^y)^T Q (X_k^x - X_k^y)|^{p/2}).\end{aligned}$$

Therefore, by iteration and the fact that Q is a symmetric positive-definite matrix we show the assertion. ∎

Therefore, given the conditions in Lemma 3.1, 3.2 and 3.3, from Theorem 2.13 we conclude that there exists a unique stationary distribution for the BEM solution as time tends to infinity.

3.2. The Underlying Stationary Distribution

The existence and uniqueness of the stationary distribution for the underlying solution is discussed in this part under the same conditions as the previous subsection. We emphasize that Theorem 3.1 in [32] is key to this part.

Lemma 3.4. *Assume Conditions 2.1, 2.2 and 2.3 hold, the second moment of the solution of (2.1) satisfies*

$$\mathbb{E}\left(\sup_{0\le t\le T_1} |x(t)|^2\right) \le (1 + \mathbb{E}|x_0|^2)\exp(2T \times max(K_1\lambda_{\max}(Q) + K_2, \alpha_1 + \alpha_2)),$$

for any $T_1 > 0$.

We refer the readers to Theorem 2.4.1 in [18] for the proof.

Lemma 3.5. *Assume the conditions in Lemma 3.2 hold, there exists a constant $p^* \in (0, 1)$ such that for any $p \in (0, p^*)$*

$$\mathbb{E}|x(t)|^p \le q(c_1 t + \mathbb{E}|x_0|^p + D^{p/2})\exp\left(p\left[K_1 + \frac{1}{2}K_3 + \frac{p}{2}K_2 q\right]t\right) < \infty,$$

holds for any $t > 0$, where $q = \lambda_{\max}(Q)/\lambda_{\min}(Q)$ and c_1 is a positive constant dependent on p, K_1, K_2, α_1, α_2, D, $\lambda_{min}(Q)$ and $\lambda_{max}(Q)$.

Proof. For $p \in (0,1)$, from the Itô formula,

$$
\begin{aligned}
& d|x^T(t)Qx(t)+D|^{p/2} \\
&= [p|x^T(t)Qx(t)+D|^{p/2-1}(x^T(t)Qf(x(t))) + p(\frac{p}{2}-1)|x^T(t)Qx(t)+D|^{p/2-2}|x^T(t)Qg(x(t))|^2 \\
&+ \frac{p}{2}|x^T(t)Qx(t)+D|^{p/2-1}(g^T(x(t))Qg(x(t)))]dt \\
&+ p|x^T(t)Qx(t)+D|^{p/2-1}(x^T(t)Qg(x(t)))dB(t) \\
&= p|x^T(t)Qx(t)+D|^{p/2}\left[\frac{x^T(t)Qf(x(t))}{x^T(t)Qx(t)+D} + \frac{\frac{1}{2}g^T(x(t))Qg(x(t))}{x^T(t)Qx(t)+D} - \frac{|x^T(t)Qg(x(t))|^2}{|x^T(t)Qx(t)+D|^2}\right. \\
&+ \left.\frac{p}{2}\frac{|x^T(t)Qg(x(t))|^2}{|x^T(t)Qx(t)+D|^2}\right]dt + p|x^T(t)Qx(t)+D|^{p/2-1}(x^T(t)Qg(x(t)))dB(t).
\end{aligned}
$$

Under (2.2), (2.3) and (3.2) it implies

$$
\begin{aligned}
d|x^T(t)Qx(t)+D|^{p/2} \le{}& p|x^T(t)Qx(t)+D|^{p/2}\left[K_1 + \frac{1}{2}K_3 + \frac{p}{2}K_2q\right]dt + c_1dt \\
&+ p|x^T(t)Qx(t)+D|^{p/2-1}(x^T(t)Qg(x(t)))dB(t),
\end{aligned}
$$

where c_1 is a positive constant dependent on p, K_1, K_2, α_1, α_2, D, $\lambda_{min}(Q)$ and $\lambda_{max}(Q)$. Since $K_1+0.5K_3<0$, given $\varepsilon \in (0, |K_1+0.5K_3|)$ we may choose $p^* \in (0,1)$ so small that $0.5p^*K_2q<\varepsilon$, then for any $p \in (0,p^*)$ we see that

$$
\begin{aligned}
\mathbb{E}|x^T(t)Qx(t)+D|^{p/2} \le{}& p\left[K_1+\frac{1}{2}K_3+\frac{p}{2}K_2q\right]\int_0^t \mathbb{E}|x^T(s)Qx(s)+D|^{p/2}ds \\
&+ c_1t + \mathbb{E}|x_0^TQx_0+D|^{p/2}.
\end{aligned}
$$

By Gronwall's inequality, we see that

$$
\mathbb{E}|x^T(t)Qx(t)+D|^{p/2} \le (c_1t+\mathbb{E}|x_0^TQx_0+D|^{p/2})\exp\left(p\left[K_1+\frac{1}{2}K_3+\frac{p}{2}K_2q\right]t\right).
$$

Although the time variable, t, appears in both the coefficient of the exponentiation term and the exponent, the choice of the p and the fact that $K_1+0.5K_3<0$ guarantee that exponentiation term decreases as t increases. Thus, the term on the right hand side of the inequality above has an upper bound. ∎

Lemma 3.6. *Assume the conditions in Lemma 3.3 hold, for any two different initial values $x_0, y_0 \in \mathbb{R}^d$, there exists a constant $p^* \in (0,1)$ such that for any $p \in (0,p^*)$*

$$
\mathbb{E}|x^{x_0}(t)-x^{y_0}(t))|^p \le q\mathbb{E}|(x_0-y_0)|^p\exp(p(\bar{K}_1+0.5K_4+0.5p\bar{K}_2q)t),
$$

where $q=\lambda_{\max}(Q)/\lambda_{\min}(Q)$.

Proof. For $p \in (0,1)$, from the Itô formula,

$$
\begin{aligned}
& d|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2} \\
&= [p|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2-1}(x^{x_0}(t) - x^{y_0}(t))^T Q(f(x^{x_0}(t)) - f(x^{y_0}(t))) \\
&+ p(\frac{p}{2} - 1)|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2-2}|(x^{x_0}(t) - x^{y_0}(t))^T Q(g(x^{x_0}(t)) - g(x^{y_0}(t)))|^2 \\
&+ \frac{p}{2}|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2-1}(g(x^{x_0}(t)) - g(x^{y_0}(t)))^T Q(g(x^{x_0}(t)) - g(x^{y_0}(t)))]dt \\
&+ p|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2-1}(x^{x_0}(t) - x^{y_0}(t))^T Q(g(x^{x_0}(t)) - g(x^{y_0}(t)))dB(t) \\
&= p|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2}\bigg(\frac{(x^{x_0}(t) - x^{y_0}(t))^T Q(f(x^{x_0}(t)) - f(x^{y_0}(t)))}{(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))} \\
&+ \frac{(g(x^{x_0}(t)) - g(x^{y_0}(t)))^T Q(g(x^{x_0}(t)) - g(x^{y_0}(t)))}{2(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))} - \frac{|(x^{x_0}(t) - x^{y_0}(t))^T Q(g(x^{x_0}(t)) - g(x^{y_0}(t)))|^2}{|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^2} \\
&+ \frac{p}{2}\frac{|(x^{x_0}(t) - x^{y_0}(t))^T Q(g(x^{x_0}(t)) - g(x^{y_0}(t)))|^2}{|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^2}\bigg)dt \\
&+ p|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2-1}(x^{x_0}(t) - x^{y_0}(t))^T Q(g(x^{x_0}(t)) - g(x^{y_0}(t)))dB(t).
\end{aligned}
$$

Under Condition 2.2, 2.3 and (3.11) this implies

$$
\begin{aligned}
& d|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2} \\
&\le p|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2}\left(\bar{K}_1 + 0.5K_4 + 0.5p\bar{K}_2 q\right)dt \\
&+ p|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2-1}(x^{x_0}(t) - x^{y_0}(t))^T Q(g(x^{x_0}(t)) - g(x^{y_0}(t)))dB(t).
\end{aligned}
$$

Since $\bar{K}_1 + 0.5K_4 < 0$, given $\varepsilon \in (0, |\bar{K}_1 + 0.5K_4|)$ we may choose $p^* \in (0,1)$ so small that $0.5p\bar{K}_2 q < \varepsilon$, then for any $p \in (0, p^*)$ we have that

$$
\begin{aligned}
& \mathbb{E}|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2} \\
&\le \mathbb{E}|(x_0 - y_0)^T Q(x_0 - y_0)|^{p/2} \\
&+ p(\bar{K}_1 + 0.5K_4 + 0.5p\bar{K}_2 q)\int_0^t \mathbb{E}|(x^{x_0}(s) - x^{y_0}(s))^T Q(x^{x_0}(s) - x^{y_0}(s))|^{p/2}ds.
\end{aligned}
$$

Then Gronwall's inequality indicates that

$$
\begin{aligned}
& \mathbb{E}|(x^{x_0}(t) - x^{y_0}(t))^T Q(x^{x_0}(t) - x^{y_0}(t))|^{p/2} \\
&\le \mathbb{E}|(x_0 - y_0)^T Q(x_0 - y_0)|^{p/2}\exp(p(\bar{K}_1 + 0.5K_4 + 0.5p\bar{K}_2 q)t).
\end{aligned}
$$

As Q is a symmetric positive-definite matrix, the proof is complete. ∎

We conclude this part by the following theorem.

Theorem 3.7. *Given the conditions in Lemma 3.1, 3.2 and 3.3, the solution of (2.1) has a unique stationary distribution denoted by* $\pi(\cdot)$.

Having Lemma 3.4, 3.5 and 3.6, the proof of this theorem follows from Theorem 3.1 in [32].

3.3. The Convergence

Given Conditions 2.1, 2.2, 2.3 and those conditions assumed in Lemma 3.1, 3.2, 3.3, the convergence of the numerical stationary distribution to the underlying stationary distribution is discussed in this subsection.

Recall that the probability measure induced by the numerical solution, X_k, is denoted by $\mathbb{P}_k(\cdot,\cdot)$, similarly we denote the probability measure induced by the underlying solution,$x(t)$, by $\bar{\mathbb{P}}_t(\cdot,\cdot)$.

Lemma 3.8. *Let Conditions 2.1, 2.2, 2.3 hold and fix any initial value* $x_0 \in \mathbb{R}^d$. *Then, for any given* $T_1 > 0$ *and* $\varepsilon > 0$ *there exists a sufficiently small* $\Delta t^* > 0$ *such that*

$$d_{\mathbb{L}}(\bar{\mathbb{P}}_{k\Delta t}(x_0,\cdot), \mathbb{P}_k(x_0,\cdot)) < \varepsilon$$

provided that $\Delta t < \Delta t^*$ *and* $k\Delta t \leq T_1$.

The result can be derived from the fact that the BEM solution converges strongly to the underlying solution in finite time [10, 12, 16].

Now we are ready to show that the numerical stationary distribution converges to the underlying stationary distribution as time step diminishes.

Theorem 3.9. *Given Conditions 2.1, 2.2, 2.3, (3.2) and (3.11),*

$$\lim_{\Delta t \to 0} d_{\mathbb{L}}(\Pi_{\Delta t}(\cdot), \pi(\cdot)) = 0.$$

Proof. Fix any initial value $x_0 \in \mathbb{R}^d$ and set $\varepsilon > 0$ to be an arbitrary real number. According to Theorem 3.7, there exists a $\Theta^* > 0$ such that for any $t > \Theta^*$

$$d_{\mathbb{L}}(\bar{\mathbb{P}}_t(x_0,\cdot), \pi(\cdot)) < \varepsilon/3.$$

Similarly, by Theorem 2.13, there exists a pair of $\Delta t^{**} > 0$ and $\Theta^{**} > 0$ such that

$$d_{\mathbb{L}}(\mathbb{P}_k(x_0,\cdot), \Pi_{\Delta t}(\cdot)) < \varepsilon/3$$

for all $\Delta t < \Delta t^{**}$ and $k\Delta t > \Theta^{**}$. Let $\Theta = \max(\Theta^*, \Theta^{**})$, from Lemma 3.8 there exists a Δt^* such that for any $\Delta t < \Delta t^*$ and $k\Delta t < \Theta + 1$

$$d_{\mathbb{L}}(\bar{\mathbb{P}}_{k\Delta t}(x_0, \cdot), \mathbb{P}_k(x_0, \cdot)) < \varepsilon/3.$$

Therefore, for any $\Delta t < \min(\Delta t^*, \Delta t^{**})$, set $k = [\Theta/\Delta t] + 1/\Delta t$, we see the assertion holds by the triangle inequality. ∎

4. Examples

In this section, we illustrate the theoretical results by three examples. First, we consider a two-dimensional SDE with scalar Brownian motion.

Example 4.1.

$$\begin{aligned} dx(t) = &(\mathrm{diag}(x_1(t), x_2(t))b + \mathrm{diag}(x_1(t), x_2(t))A\mathrm{diag}(x_1(t), x_2(t))x(t) + c_1)\, dt \\ &+ (\mathrm{diag}(x_1(t), x_2(t))\sigma + c_2)\, dB(t), \end{aligned} \tag{4.1}$$

where $x(t) = (x_1(t), x_2(t))^T$, $\mathrm{diag}(x_1(t), x_2(t))$ *denotes a diagonal matrix with non-zero entries* $x_1(t)$ *and* $x_2(t)$ *on the diagonal,* $b = (1,1)^T$, $A = (a_{ij})_{i,j=1,2}$ *with* $a_{1,1} = -1, a_{1,2} = -0.7, a_{2,1} = -1.2, a_{2,2} = -2$, $c_1 = (0.5, 0.7)^T$, $c_2 = (3.5, 4)^T$ *and* $\sigma = (3.5, 4)^T$.

Choosing Q to be an identity matrix, it is clear that the drift and diffusion coefficients of (4.1) satisfy Conditions 2.1, 2.2, 2.3 and (2.2) with $\bar{K}_1 = 1$ and $K_1 = 1.7$, which indicate that Lemma 3.1 holds. To check conditions for Lemma 3.2, we see that

$$\frac{(3.5x_1 + 0.3)^2 + (4x_2 + 0.2)^2}{D + (x_1^2 + x_2^2)} - \frac{2|3.5x_1^2 + 0.3x_1 + 4x_2^2 + 0.2x_2|^2}{(D + (x_1^2 + x_2^2))^2}.$$

Set $D = 0.04/25$, we can derive that (3.2) is satisfied with $K_3 = -7$ and $K_1 + 0.5K_3 < 0$, then Lemma 3.2 holds. Finally, we have that (3.11) is satisfied with $K_4 = -7$ and $\bar{K}_1 + 0.5K_4 < 0$, that is Lemma 3.3 holds.

We simulate 1000 paths, each of which has 10000 iterations. In Figure 1, we plot one path of the BEM solution for $x_1(t)$ and $x_2(t)$. Intuitively, some stationary behaviour displays.

We further plot the empirical cumulative distribution function (ECDF) of the last iterations of the 1000 paths and the ECDF of last 1000 iterations of one

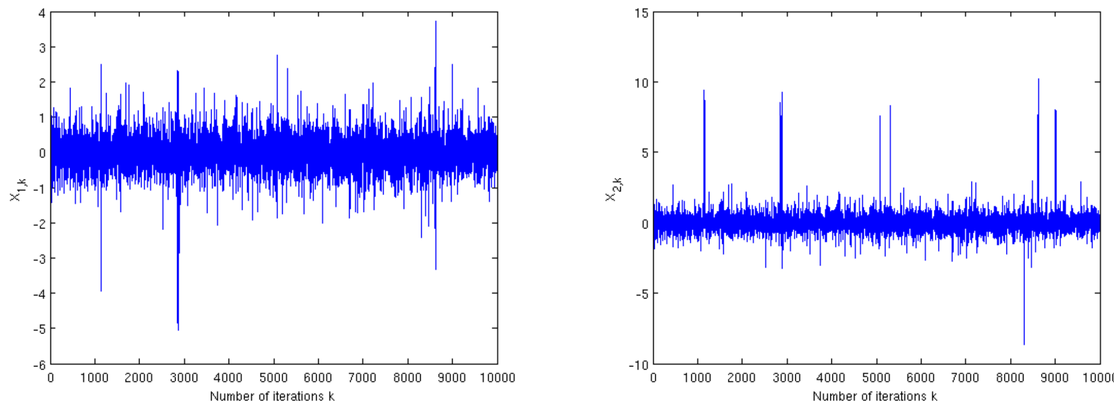

Figure 1: Left: the BEM solution to $x_1(t)$; Right: the BEM solution to $x_2(t)$.

path in Figure 2. It can be seen that the shapes and the intervals of the ECDFs are similar. To measure the similarity quantitatively, we use the Kolmogorov-Smirnov test (K-S test) [23] to test the alternative hypothesis that the last iterations of the 1000 paths and last 1000 iterations of one path are from different distributions against the null hypothesis that they are from the same distribution for both $x_1(t)$ and $x_2(t)$. With 5% significance level, the K-S test indicates that we cannot reject the null hypothesis. This example illustrates the existence of the stationary distribution as the time variable becomes large. Moreover, it may indicate that instead of simulating many paths to construct the stationary distribution, one could just use the last few iterations of one path to approximate the stationary distribution.

To compare the numerical stationary distribution with the theoretical one, we next consider a nonlinear scalar SDE, whose stationary distribution can be explicitly derived from the Kolmogorov-Fokker-Planck equation.

Example 4.2.

$$dx(t) = -0.5(x + x^3)dt + dB(t).$$

It is straightforward to see that $\bar{K}_1 = K_1 = -0.5$ and $K_3 = K_4 = 0$, hence all the conditions required in Section 2 and 3 are satisfied. The corresponding Kolmogorov-Fokker-Planck equation for the theoretical probability density

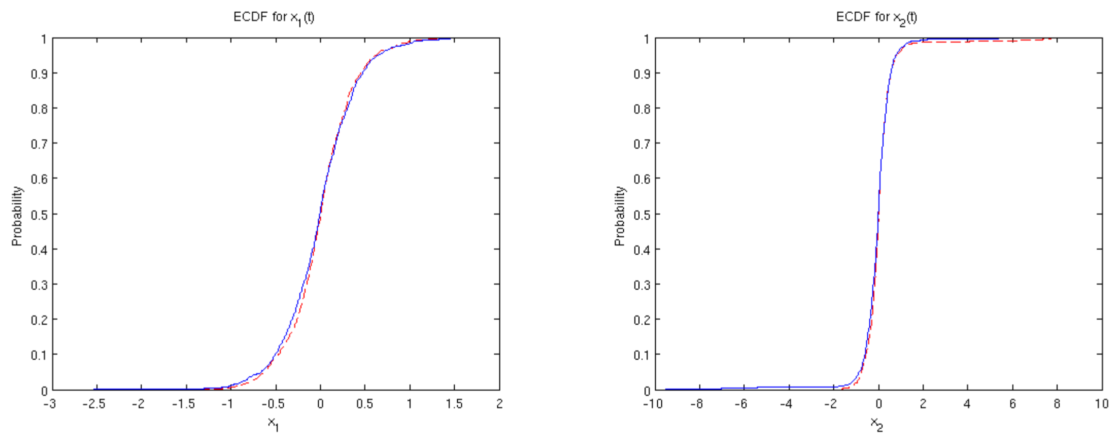

Figure 2: Left: ECDFs for x_1; Right: ECDFs for x_2. The red dashed line is last 1000 iterations of one path; The blue solid line is last iterations of the 1000 paths.

function of the stationary distribution $p(x)$ is

$$0.5\frac{d^2p(x)}{dx^2} - \frac{d}{dx}(-0.5(x+x^3)p(x)) = 0.$$

And the exact solution is known to be [27]

$$p(x) = \frac{1}{I_{\frac{1}{4}}(\frac{1}{8}) + I_{-\frac{1}{4}}(\frac{1}{8})}\exp(\frac{1}{8} - \frac{1}{2}x^2 - \frac{1}{4}x^4),$$

where $I_\nu(x)$ is a modified Bessel function of the first kind. We simulate one path with 100000 iterations and plot the ECDF of last 20000 iterations in red dashed line in Figure 3. The theoretical cumulative distribution function is plotted on the same figure in blue solid line. The similarity of those two distribution is clear seen, which indicates that the numerical stationary distribution is a good approximation to the theoretical one. The mean and variance of the numerical stationary distribution are 0 and 0.453, respectively, which are close to the theoretical counterparts 0 and 0.466.

This example also demonstrates that the numerical method for stochastic differential equations can serve as an alternative way to approximate deterministic differential equations.

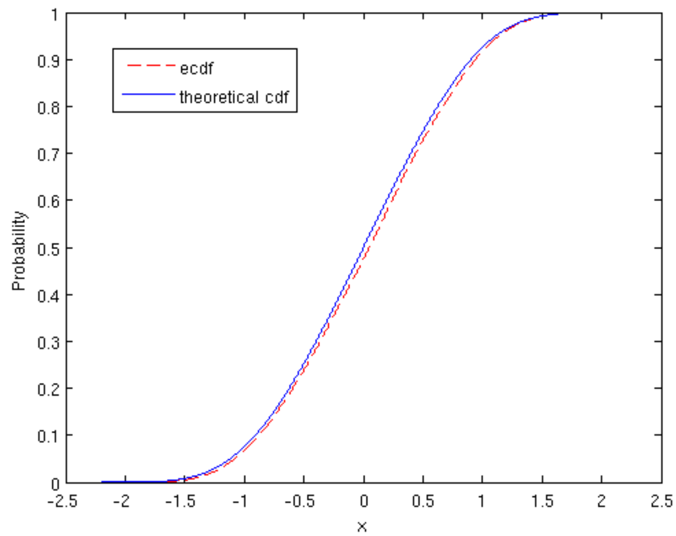

Figure 3: Comparison of the ECDF with the theoretical cumulative distribution function

At last, we consider a linear scalar equation, the Langevin equation [29]. The comparison of the BEM method in this paper with the EM method studied in [33] demonstrates that the BEM has less constraint on the step size.

Example 4.3. *We write the Itô type equation of the Langevin equation as*

$$dx(t) = -\alpha x(t)dt + \sigma dB(t) \quad \text{on } t \geq 0, \tag{4.2}$$

where $\alpha > 0$ *and* $\sigma \in \mathbb{R}$.

From (2.4), given the initial value $X_0 = x(0) \in \mathbb{R}$ we have

$$X_{k+1} = X_k + -\alpha X_{k+1}\Delta t + \sigma \Delta B_k.$$

This gives that X_{k+1} is normally distributed with mean

$$\mathbb{E}(X_{k+1}) = (1 + \alpha\Delta t)^{-(k+1)} x(0)$$

and variance

$$\begin{aligned}
Var(X_{k+1}) &= (1+\alpha\Delta t)^{-2}Var(X_k) + \sigma^2(1+\alpha\Delta t)^{-2}\Delta t \\
&= \sigma^2\Delta t[(1+\alpha\Delta t)^{-2} + (1+\alpha\Delta t)^{-4} + ... + (1+\alpha\Delta t)^{-2(k+1)}] \\
&= \sigma^2\Delta t\frac{1-(1+\alpha\Delta t)^{-(k+1)}}{(1+\alpha\Delta t)^2-1} \\
&= \frac{\sigma^2}{2\alpha+\alpha^2\Delta t}.
\end{aligned}$$

So the distribution of the BEM solution approaches the normal distribution $N(0, \sigma^2/(2\alpha + \alpha^2\Delta t))$ as $k \to \infty$ for any $\Delta t > 0$. Recall, from Example 3.5.1 in [18], that the underlying solution of (4.2) approaches its stationary distribution $N(0, \sigma^2/(2\alpha))$ as $t \to \infty$, then it is interesting to observe that $N(0, \sigma^2/(2\alpha + \alpha^2\Delta t))$ will further converge to stationary distribution of the true solution as $\Delta t \to 0$.

5. Conclusions and Future Research

This paper extends the results in second author's series paper [21, 33, 31]. By using the Backward Euler-Maruyama method, the linear growth condition on the drift coefficient is replaced by the one-sided Lipschitz condition and the stationary distribution of many more SDEs can be approximated by the numerical stationary distribution. However, it should be mentioned that, compared to the three assumptions in Section 2, those sufficient conditions in Section 3.1 are still stronger. And this is because those assumptions are in probability, while those sufficient conditions are in terms of moment. Therefore, it is interesting to construct some coefficient-related sufficient conditions which are in probability. And this may be achieved by using different Lyapunov functions other than the one $V(x) = (D + x^TQx)^{p/2}$ employed in this paper.

Another interesting future work is to investigate the convergence rates of the distributions of different types of numerical methods. Furthermore, it may be interesting to conduct some numerical analyses about approximating deterministic differential equations by the numerical stationary distributions of SDEs.

Acknowledgements

The authors would like to thank the anonymous referee and the editor for their useful comments and remarks which improved the paper significantly. The authors also would like to thank the EPSRC (grant EP/J017418/1), the Royal Society of London, the Royal Society of Edinburgh, the National Natural Science Foundation of China (grants 11071037, 71073023) and Chinese Hai Wai Ming Shi Program for their financial support.

References

[1] E. Allen. *Modeling with Itô stochastic differential equations*, volume 22 of *Mathematical Modelling: Theory and Applications*. Springer, Dordrecht, 2007.

[2] G. Berkolaiko, E. Buckwar, C. Kelly, and A. Rodkina. Almost sure asymptotic stability analysis of the θ-Maruyama method applied to a test system with stabilising and destabilising stochastic perturbations. *LMS J. Comput. Math.*, 15:71–83, 2012.

[3] E. Buckwar and T. Sickenberger. A structural analysis of asymptotic mean-square stability for multi-dimensional linear stochastic differential systems. *Appl. Numer. Math.*, 62(7):842–859, 2012.

[4] E. Buckwar, M. Riedler and P. E. Kloeden The numerical stability of stochastic ordinary differential equations with additive noise. *Stoch. Dyn.*, 11(2-3):265–281, 2011.

[5] A. Gray, D. Greenhalgh, L. Hu, X. Mao, and J. Pan. A stochastic differential equation SIS epidemic model. *SIAM J. Appl. Math.*, 71(3):876–902, 2011.

[6] D. J. Higham. Mean-square and asymptotic stability of the stochastic theta method. *SIAM J. Numer. Anal.*, 38(3):753–769 (electronic), 2000.

[7] D. J. Higham. Modeling and simulating chemical reactions. *SIAM Rev.*, 50(2):347–368, 2008.

[8] D. J. Higham. Stochastic ordinary differential equations in applied and computational mathematics. *IMA J. Appl. Math.*, 76(3):449–474, 2011.

[9] D. J. Higham and P. E. Kloeden. Strong convergence rates for backward Euler on a class of nonlinear jump-diffusion problems. *J. Comput. Appl. Math.*, 205(2):949–956, 2007.

[10] D. J. Higham, X. Mao, and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, 40(3):1041–1063 (electronic), 2002.

[11] D. J. Higham, X. Mao, and C. Yuan. Almost sure and moment exponential stability in the numerical simulation of stochastic differential equations. *SIAM J. Numer. Anal.*, 45(2):592–609 (electronic), 2007.

[12] Y. Hu. Semi-implicit Euler-Maruyama scheme for stiff stochastic equations. In *Stochastic analysis and related topics, V (Silivri, 1994)*, volume 38 of *Progr. Probab.*, pages 183–202. Birkhäuser Boston, Boston, MA, 1996.

[13] C. Huang. Exponential mean square stability of numerical methods for systems of stochastic differential equations. *J. Comput. Appl. Math.*, 236(16):4016–4026, 2012.

[14] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 467(2130):1563–1576, 2011.

[15] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, 1981.

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

[17] X. Mao. *Stability of stochastic differential equations with respect to semimartingales*, volume 251 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow, 1991.

[18] X. Mao. *Stochastic differential equations and applications*. Horwood Publishing Limited, Chichester, second edition, 2008.

[19] X. Mao, Y. Shen, and A. Gray. Almost sure exponential stability of backward Euler-Maruyama discretizations for hybrid stochastic differential equations. *J. Comput. Appl. Math.*, 235(5):1213–1226, 2011.

[20] X. Mao and L. Szpruch. Strong convergence rates for backward Euler-Maruyama method for non-linear dissipative-type stochastic differential equations with super-linear diffusion coefficients. *Stochastics*, 85(1):144–171, 2013.

[21] X. Mao, C. Yuan, and G. Yin. Numerical method for stationary distribution of stochastic differential equations with Markovian switching. *J. Comput. Appl. Math.*, 174(1):1–27, 2005.

[22] X. Mao. Stationary distribution of stochastic population systems. *Systems Control Lett.*, 60(6):398–405, 2011.

[23] F. J Massey Jr. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

[24] G. N. Milstein. *Numerical integration of stochastic differential equations*, volume 313 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995. Translated and revised from the 1988 Russian original.

[25] B. Oksendal. *Stochastic differential equations.* Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.

[26] H. Schurz. *Stability, stationarity, and boundedness of some implicit numerical methods for stochastic differential equations and applications.* Logos Verlag Berlin, Berlin, 1997.

[27] T. T. Soong. *Random differential equations in science and engineering.* Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1973. Mathematics in Science and Engineering, Vol. 103.

[28] L. Szpruch, X. Mao, D. J. Higham, and J. Pan. Numerical simulation of a strongly nonlinear Ait-Sahalia-type interest rate model. *BIT*, 51(2):405–425, 2011.

[29] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the Brownian motion. *Physical review*, 36(5):823, 1930.

[30] F. Wu, X. Mao, and L. Szpruch. Almost sure exponential stability of numerical solutions for stochastic delay differential equations. *Numer. Math.*, 115(4):681–697, 2010.

[31] C. Yuan and X. Mao. Stationary distributions of Euler-Maruyama-type stochastic difference equations with Markovian switching and their convergence. *J. Difference Equ. Appl.*, 11(1):29–48, 2005.

[32] C. Yuan and X. Mao. Asymptotic stability in distribution of stochastic differential equations with Markovian switching. *Stochastic Process. Appl.*, 103(2):277–291, 2003.

[33] C. Yuan and X. Mao. Stability in distribution of numerical solutions for stochastic differential equations. *Stochastic Anal. Appl.*, 22(5):1133–1150, 2004.