Stability of regime-switching processes under perturbation of transition rate matrices*

Jinghai Shao† and Chenggui Yuan ‡

December 26, 2018

Abstract

This work is concerned with the stability of regime-switching processes under the perturbation of the transition rate matrices. From the viewpoint of application, two kinds of perturbations are studied: the size of the transition rate matrix is fixed, and only the values of entries are perturbed; the values of entries and the size of the transition matrix are all perturbed. Moreover, both regular and irregular coefficients of the underlying system are investigated, which clarifies the impact of the regularity of the coefficients on the stability of the underlying system.

AMS subject Classification (2010): 60J27, 60J60, 60A10

Keywords: Regime-switching diffusions, Stability, Wasserstein distance, Transition rate matrices

*Supported in part by NNSFs of China (Nos. 11771327, 11831014, 11431014)

†Center for Applied Mathematics, Tianjin University, Tianjin 300072, China. Email: shaojh@tju.edu.cn.

‡Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, UK. Email: C.Yuan@swansea.ac.uk

1 Introduction

Regime-switching models have emerged in many research fields such as biological, ecological, mathematical finance, economics and storage modeling. We refer the readers to [1, 3, 8, 10, 18, 19, 23, 29] and the monographs [15, 30] for the study on ergodicity, stochastic stability, numerical approximation of regime-switching diffusion processes with Markovian switching or state-dependent switching in a finite state space or in an infinite state space. These kinds of models contain two components (X_t, Λ_t). The first component (X_t) is used to describe the dynamical system under investigation and the second component (Λ_t) is used to describe the random change of the environment where the dynamical system lives in. Since the impact of the change of environment has been considered in these models, they can fit practice more precisely. Moreover, recent works have found more and more special characteristics of these models compared with those models without regime-switching. For instance, the invariant probability measures of Ornstein-Uhlenback processes and Cox-Ingersoll-Ross processes with regime-switching may be heavy tailed, whereas without regime-switching, their invariant probability measures must be light tailed; see, [2, 9] and [11].

The stability of regime-switching processes is of great interest and there is a great deal of literatures in this topic; see, for example, [3, 4, 14, 15, 29, 30] and references therein. All the aforementioned works focus on the stability of this system with respective to its equilibrium point or initial values. However, the stability of this system with respective to the perturbation of the transition rate matrix of (Λ_t) has not been studied before. This kind of stability plays a crucial role in the application of the regime-switching diffusion processes; for example, performing sensitivity analysis.

In application, the random switching of the environment is observed from empirical data. Then, the transition rate matrix $(q_{ij})_{i,j\in\mathcal{S}}$ is estimated by statistical method based on empirical data. Therefore, the error of estimation is crucial and cannot be removed. As a consequence, the impact of this error of estimation should be evaluated. For instance, as shown by Brown and Dybvig [5], based on the empirical data from US treasury yields, the poor empirical performance of the Cox-Ingersoll-Ross model without the regime-switching well suggests the existence of regime shifts. So, one may include the regime-switching of the financial market into the Cox-Ingersoll-Ross model. It is quite possible to consider that there are three different states in the financial market: bull market, bear market and a middle market. In this case, one uses a Markov chain (Λ_t) in a state space $\mathcal{S} = \{0, 1, 2\}$ to characterize the random change of the financial market. There is the error of estimation for $(q_{ij})_{i,j\in\mathcal{S}}$ of the transition rate matrix of (Λ_t). On the other hand, maybe other experts would like to separate the financial market into two different states: bull market and bear market. The effects of the option pricing by using models with two or three states could be quite different. Therefore, it is quite important to measure this difference.

For the regime-switching diffusions (X_t, Λ_t), (X_t) satisfies the following stochastic differential equation (SDE for short):

$$\mathrm{d}X_t = b(X_t, \Lambda_t)\mathrm{d}t + \sigma(X_t, \Lambda_t)\mathrm{d}W_t, \quad X_0 = x_0 \in \mathbb{R}^d,\ \Lambda_0 = i_0 \in \mathcal{S}, \tag{1.1}$$

where $b : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}^{d\times d}$, $\mathcal{S} = \{0, 1, \ldots, N\}$, $N < \infty$, and (W_t) is a d-dimensional Brownian motion. (Λ_t) is a continuous-time Markov chain on $\mathcal{S}$ with the transition rate matrix $Q = (q_{ij})_{i,j\in\mathcal{S}}$. Suppose that Q is conservative (i.e. $\sum_{j\in\mathcal{S}} q_{ij} = 0$ for every $i \in \mathcal{S}$) and totally stable (i.e. $q_i = -q_{ii} < +\infty$ for every $i \in \mathcal{S}$). Throughout this work, (Λ_t) and (W_t) are assumed to be mutually independent.

In this work we are concerned with the stability of the process (X_t) under perturbation of the transition rate matrix of (Λ_t). From the application point of view, there are mainly two types of perturbations of Q.

First type of perturbation: The size of Q is fixed, however, each entry q_{ij} of Q may have small perturbation. Namely, there is another transition rate matrix $\widetilde{Q} = (\tilde{q}_{ij})_{i,j\in\mathcal{S}}$, and each entry $\tilde{q}_{ij}$ acts as an estimator of the element q_{ij} of Q. Without loss of generality, assume that $\widetilde{Q}$ is conservative and totally stable, then a unique transition function $\widetilde{P}_t$, $t \geq 0$ is determined (cf. e.g. [7, Corollary 3.12]). Let $(\tilde{\Lambda}_t)$ be a continuous-time Markov chain starting from i_0 corresponding to $\widetilde{Q}$. Then the distribution of $\tilde{\Lambda}_t$ is fixed, so, a new dynamical system $(\widetilde{X}_t)$ is induced from the process $(\tilde{\Lambda}_t)$, i.e.

$$\mathrm{d}\widetilde{X}_t = b(\widetilde{X}_t, \tilde{\Lambda}_t)\mathrm{d}t + \sigma(\widetilde{X}_t, \tilde{\Lambda}_t)\mathrm{d}W(t), \quad \widetilde{X}_0 = x_0 \in \mathbb{R}^d,\ \tilde{\Lambda}_0 = i_0 \in \mathcal{S}. \tag{1.2}$$

Under some suitable conditions of the coefficients $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$, SDEs (1.1) and (1.2) admit a unique solution (cf. e.g. [15]). Therefore, the distributions $\mathcal{L}(X_t)$ of X_t and $\mathcal{L}(\widetilde{X}_t)$ of $\widetilde{X}_t$ are determined in some sense by the transition rate matrix Q and $\widetilde{Q}$ respectively. The following basic and important question therefore arises:

- Can the difference between the distributions of X_t and $\widetilde{X}_t$ be estimated by the difference between Q and $\widetilde{Q}$?

Second type of perturbation: Both the entries of Q and the size of Q can be changed. In application, when facing the graphs drawn from experimental data, it is hard sometimes to determine the number of the regimes for the regime-switching processes. For example, if there are actually three regimes, the process stays for a very short period of time at one of them. From this kind of experimental data, it is very likely that a regime-switching model with only two regimes is detected. What is the impact caused by this incorrect choice of the number of states for the regime-switching processes?

Precisely, let $\widehat{Q}$ be a conservative transition rate matrix on $E := \mathcal{S}\backslash\{0, 1, \dots, m\}$ with $m < N$, which determines uniquely the semigroup $\hat{P}_t = \mathrm{e}^{t\widehat{Q}},\ t \geq 0$ on E. Let $(\hat{\Lambda}_t)$ be a continuous-time Markov chain on E corresponding to $(\hat{P}_t)$ or equivalently $\widehat{Q}$. Using the same coefficients $b(\cdot, \cdot)$, $\sigma(\cdot, \cdot)$ as those of SDE (1.1), we consider a new dynamical system $(\hat{X}_t)$ corresponding to $(\hat{\Lambda}_t)$ defined by:

$$\mathrm{d}\hat{X}_t = b(\hat{X}_t, \hat{\Lambda}_t)\mathrm{d}t + \sigma(\hat{X}_t, \hat{\Lambda}_t)\mathrm{d}W_t, \quad \hat{X}_0 = x_0 \in \mathbb{R}^d,\ \hat{\Lambda}_0 = i_0 \in E. \tag{1.3}$$

Under suitable conditions of b and σ, the solutions of (1.1) and (1.3) are uniquely determined (cf. [15]). This means that given $\widehat{Q}$ on E, the distribution of $\hat{X}_t$ is then determined. Denote $\mathcal{L}(X_t)$ and $\mathcal{L}(\hat{X}_t)$ the distributions of X_t and $\hat{X}_t$ respectively. We aim to measure the Wasserstein distance $W_2(\mathcal{L}(X_t), \mathcal{L}(\hat{X}_t))$ via the difference between the transition rate matrices $Q = (q_{ij})_{i,j\in\mathcal{S}}$ and $\widehat{Q} = (\hat{q}_{ij})_{i,j\in E}$. To achieve this, reformulate Q into the following form:

$$Q = \begin{pmatrix} Q_0 & A \\ B & Q_1 \end{pmatrix}, \tag{1.4}$$

where $Q_0 \in \mathbb{R}^{m\times m}$, $A \in \mathbb{R}^{m\times(N-m)}$, $B \in \mathbb{R}^{(N-m)\times m}$, and $Q_1 \in \mathbb{R}^{(N-m)\times(N-m)}$.

Our method in this paper establishes a connection between the stability of regime-switching processes with the perturbation theory of the continuous time Markov chains under the help of Skorokhod's representation theory for Markov chains. This result develops the classical perturbation theory (cf. e.g. [16, 17, 31]) focusing on the difference of fixed time t to that of a time interval $[0, t]$. The perturbation theory of continuous time Markov chain was applied to study the strong ergodicity of Markov chain (cf. [31] and references therein), and to perform sensitivity analysis (cf. [16, 17]). In this paper, we demonstrate its connection with the stability of regime-switching processes, allowing us to performing sensitivity analysis for regime-switching processes arising from applications. In addition, to clarify the impact of the regularity of the drifts of the underlying system on this stability issue, we consider the system with regular coefficients (i.e. satisfying one-sided Lipschitz condition) and irregular coefficients (i.e. satisfying integrability condition). To deal with the irregular case, we apply a technique based on the dimension-free Harnack inequality. The coefficients in the irregular case can be very singular; see example (1.15) below.

Let us first consider the situation that the coefficients of (1.1) are regular. Assume the coefficients $b : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}^{d\times d}$ satisfy:

(H1) For each $i \in \mathcal{S}$ there exists a constant κ_i such that

$$2\langle x-y, b(x,i)-b(y,i)\rangle + 2\|\sigma(x,i)-\sigma(y,i)\|_{\mathrm{HS}}^2 \le \kappa_i|x-y|^2, \quad x,\, y \in \mathbb{R}^d.$$

(H2) There exists a constant K such that

$$|b(x,i)|^2 \le K(1+|x|^2), \quad \|\sigma(x,i)\|_{\mathrm{HS}}^2 \le K(1+|x|^2), \quad x \in \mathbb{R}^d,\ i \in \mathcal{S}.$$

In this case, we shall use the Wasserstein distance $W_2(\cdot,\cdot)$ to measure the difference between the distributions of X_t and $\widetilde{X}_t$, which is defined by

$$W_2(\nu_1,\nu_2)^2 = \inf_{\Pi\in\mathcal{C}(\nu_1,\nu_2)} \Big\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \Pi(\mathrm{d}x,\mathrm{d}y) \Big\}, \tag{1.5}$$

where $\mathcal{C}(\nu_1,\nu_2)$ denotes the set of all probability measures on $\mathbb{R}^d\times\mathbb{R}^d$ with marginals ν_1 and ν_2. To measure the difference between Q and $\widetilde{Q}$, we use the ℓ_1-norm $\|Q-\widetilde{Q}\|_{\ell_1}$ (i.e. the maximum absolute row sum norm) in this work, but other norm of matrix still works.

To state our results, we first introduce some notation. For an irreducible transition rate matrix Q on $\mathcal{S}$, its corresponding transition probability measure $P_t(i,\cdot)$ must be strongly ergodic (cf. e.g. [7, Theorems 4.43, 4.44]). Denote $\pi = (\pi_i)$ the invariant probability measure of Q. Define τ to be the largest positive constant such that

$$\sup_{i\in\mathcal{S}} \|P_t(i,\cdot)-\pi\|_{\mathrm{var}} = O(\mathrm{e}^{-\tau t}), \quad t>0, \tag{1.6}$$

where $\|\mu-\nu\|_{\mathrm{var}}$ stands for the total variation distance between two probability measures μ and ν, i.e. $\|\mu-\nu\|_{\mathrm{var}} = 2\sup\{|\mu(A)-\nu(A)|; A\in\mathscr{B}(S)\}$. Additionally, for $p>0$, let

$$Q_p = Q + p\,\mathrm{diag}(\kappa_0,\kappa_1,\dots,\kappa_N),$$

and

$$\eta_p = -\max\big\{\mathrm{Re}(\gamma);\, \gamma\in\mathrm{spec}(Q_p)\big\}, \tag{1.7}$$

where $\mathrm{diag}(\kappa_0, \kappa_1, \ldots, \kappa_N)$ denotes the diagonal matrix generated by the vector $(\kappa_0, \kappa_1, \ldots, \kappa_N)$, $\mathrm{spec}(Q_p)$ denotes the spectrum of the operator Q_p.

We are now in the position to state our main results of this work for SDEs with regular coefficients. The first result is about the estimate of the difference of distributions of the solutions of (1.1) and (1.2).

Theorem 1.1 *Let (X_t, Λ_t) and $(\widetilde{X}_t, \tilde{\Lambda}_t)$ be the solution of (1.1) and (1.2) respectively. Assume* (H1) *and* (H2) *hold. Then*

$$W_2(\mathcal{L}(X_t), \mathcal{L}(\widetilde{X}_t))^2 \le \big(4\varepsilon^{-1}+8\big) K C_2(p)^{\frac{1}{p}} \Big(N^2 t^2 \|Q - \widetilde{Q}\|_{\ell_1}\Big)^{\frac{1}{q}} \Psi(t, \varepsilon, \eta_p, K, p), \tag{1.8}$$

where $p > 1$, $q = p/(p-1)$, ε and $C_2(p)$ are positive constants, η_p is defined by (1.7), and

$$\Psi(t, \varepsilon, \eta_p, K, p) = \Big(\int_0^t \big[1 + (|x_0|^2 + 2Ks)\mathrm{e}^{(2K+1)s}\big]^p \mathrm{e}^{-(\eta_p - \varepsilon p)(t-s)} \mathrm{d}s\Big)^{\frac{1}{p}}. \tag{1.9}$$

If assume further that

$$|b(x,i)|^2 \le K, \quad \|\sigma(x,i)\|_{\mathrm{HS}}^2 \le K, \quad x \in \mathbb{R}^d,\ i \in \mathcal{S}, \tag{1.10}$$

then we have a simple estimate:

$$\begin{aligned} &W_2(\mathcal{L}(X_t), \mathcal{L}(\widetilde{X}_t))^2 \\ &\le (4\varepsilon^{-1}+8) K C_2(p)^{\frac{1}{p}} \big(N^2 t^2 \|Q - \widetilde{Q}\|_{\ell_1}\big)^{\frac{1}{q}} \Big(\frac{1 - \mathrm{e}^{-(\eta_p - \varepsilon p)t}}{\eta_p - \varepsilon p}\Big)^{\frac{1}{p}}. \end{aligned} \tag{1.11}$$

The second result is about the estimate of the difference of distributions of the solutions of (1.1) and (1.3).

Theorem 1.2 *Let (X_t, Λ_t) and $(\hat{X}_t, \hat{\Lambda}_t)$ be the solutions of (1.1) and (1.3) respectively. Suppose $\tilde{\Lambda}_0 = \Lambda_0 \in E$. Assume* (H1) *and* (H2) *hold. Then*

$$\begin{aligned} &W_2(\mathcal{L}(X_t), \mathcal{L}(\hat{X}_t))^2 \\ &\le \big(4\varepsilon^{-1}+8\big) K C_2(p)^{\frac{1}{p}} \big(Nt\big)^{\frac{2}{q}} \Big(\|B\|_{\ell_1} + \|Q_1 - \widehat{Q}\|_{\ell_1}\Big)^{\frac{1}{q}} \Psi(t, \varepsilon, \eta_p, K, p), \end{aligned} \tag{1.12}$$

where $p > 1$, $q = p/(p-1)$, ε and $C_2(p)$ are positive constants, η_p is defined by (1.7), and $\Psi(t, \varepsilon, \eta_p, K, p)$ is given by (1.9). Assume further that b and σ satisfy (1.10), then

$$\begin{aligned} &W_2(\mathcal{L}(X_t), \mathcal{L}(\hat{X}_t))^2 \\ &\le (4\varepsilon^{-1}+8) K C_2(p)^{\frac{1}{p}} \big(Nt\big)^{\frac{2}{q}} \Big(\|B\|_{\ell_1} + \|Q_1 - \widehat{Q}\|_{\ell_1}\Big)^{\frac{1}{q}} \Big(\frac{1 - \mathrm{e}^{-(\eta_p - \varepsilon p)t}}{\eta_p - \varepsilon p}\Big)^{\frac{1}{p}}. \end{aligned} \tag{1.13}$$

Next, we consider the stability of the dynamical system (X_t) under the perturbation of the transition rate matrix when the coefficients of the underlying SDE are irregular. Precisely, let

$$\mathrm{d}X_t = b(X_t, \Lambda_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad X_0 = x_0 \in \mathbb{R}^d,\ \Lambda_0 = i_0 \in \mathcal{S}, \tag{1.14}$$

where $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is still Lipschitz continuous, but b only satisfies some integrability condition. Here, (Λ_t) is also a continuous time Markov chain with a conservative and irreducible transition rate matrix $Q = (q_{ij})_{i,j\in\mathcal{S}}$. (Λ_t) is assumed to be independent of (W_t). A typical example of the irregular drift b concerned in this work is

$$b(x, i) = \beta_i \Big\{ \sum_{k=1}^{\infty} \log\Big(1 + \frac{1}{|x-k|^2}\Big)\Big\}^{\frac{1}{2}} - x, \tag{1.15}$$

where $\beta : \mathcal{S} \to \mathbb{R}_+$. This drift b is rather singular, whereas we can show that (X_t) is still stable in a suitable sense w.r.t. the perturbation of Q even in this situation. There are lots of researches on SDEs with irregular drifts in the form (1.15) or in $L^p([0,\infty); L^q(\mathbb{R}^d))$. We refer the readers to the recent works [28, 32] and references therein for more details on the motivations and applications.

Similar to (1.2) and (1.3), we consider the processes $(\widetilde{X}_t)$ and $(\hat{X}_t)$ corresponding to the perturbations $\widetilde{Q} = (\tilde{q}_{ij})_{i,j\in\mathcal{S}}$ and $\widehat{Q} = (\hat{q}_{ij})_{i,j\in E}$. Namely,

$$\mathrm{d}\widetilde{X}_t = b(\widetilde{X}_t, \tilde{\Lambda}_t)\mathrm{d}t + \sigma(\widetilde{X}_t)\mathrm{d}W_t,\ \widetilde{X}_0 = x_0,\ \tilde{\Lambda}_0 = i_0, \tag{1.16}$$

where $(\tilde{\Lambda}_t)$ is associated with $\widetilde{Q}$ and is independent of (W_t).

$$\mathrm{d}\hat{X}_t = b(\hat{X}_t, \hat{\Lambda}_t)\mathrm{d}t + \sigma(\hat{X}_t)\mathrm{d}W_t,\ \hat{X}_0 = x_0,\ \hat{\Lambda}_0 = i_0 \in E, \tag{1.17}$$

where $(\hat{\Lambda}_t)$ is associated with $\widehat{Q}$ on the state space E and is independent of (W_t). We shall measure the difference between the distribution $\mathcal{L}(X_t)$ and $\mathcal{L}(\widetilde{X}_t)$ by the Fortet-Mourier distance (also called bounded Lipschitz distance):

$$W_{bL}(\mu, \nu) = \sup\Big\{ \int_{\mathbb{R}^d} \phi\, \mathrm{d}\mu - \int_{\mathbb{R}^d} \phi\, \mathrm{d}\nu;\ \|\phi\|_{\mathrm{Lip}} + \|\phi\|_\infty \le 1 \Big\} \tag{1.18}$$

for two probability measures μ, ν on $\mathbb{R}^d$, $\|\phi\|_{\mathrm{Lip}} := \sup_{x,y,\in\mathbb{R}^d, x\neq y} \frac{|\phi(x)-\phi(y)|}{|x-y|}$. The Fortet-Mourier distance can also characterize the weak convergence of the probability measure space (cf. [26, Chapter 6]), and it is closely related to the L_1-Wasserstein distance via the Kantorovich-Rubinstein Theorem (cf. [25, Theorem 1.14]).

To provide a suitable integrability condition on the drift b, we need to introduce an auxiliary function V and its associated probability measure μ_0. Let $V \in C^2(\mathbb{R}^d)$, define

$$Z_0(x) = -\sum_{i,j=1}^{d} \big(a_{ij}(x)\partial_j V(x)\big)e_i, \tag{1.19}$$

where $(a_{ij}(x)) = \sigma(x)\sigma^*(x)$, σ^* denotes the transpose of σ given in (1.14), $\{e_i\}_{i=1}^d$ is the canonical orthonormal basis of $\mathbb{R}^d$ and ∂_j is the directional derivative along e_j. Let

$$\mu_0(\mathrm{d}x) = \mathrm{e}^{-V(x)}\mathrm{d}x. \tag{1.20}$$

Assume that V satisfies:

(A) there exists a $K_0 > 0$ such that $|Z_0(x) - Z_0(y)| \le K_0|x-y|$ for all $x, y \in \mathbb{R}^d$, and $\mu_0(\mathbb{R}^d) = 1$.

Let

$$Z(x,i) = b(x,i) - Z_0(x), \quad x \in \mathbb{R}^d,\ i \in \mathcal{S}. \tag{1.21}$$

For the example b in (1.15), we can take $V(x) = x^2/2 + \log\sqrt{2\pi}$, then $Z_0(x) = -x$ and $\mu_0(\mathrm{d}x) = \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x$. Also, the integrability condition (1.22) below can be verified by direct calculation for this example. In this part, for $f \in \mathscr{B}(\mathbb{R}^d)$, $\mu_0(f)$ denotes $\int_{\mathbb{R}^d} f(x)\mu_0(\mathrm{d}x)$.

Theorem 1.3 *Let (X_t, Λ_t) be a solution of (1.14) and $(\widetilde{X}_t, \tilde{\Lambda}_t)$ a solution of (1.16). Suppose $V \in C^2(\mathbb{R}^d)$ satisfying condition (A). Let $T > 0$ be fixed. Assume that there exists a constant $\eta > 2Td$ such that*

$$\max_{i\in\mathcal{S}} \mu_0\Big(\mathrm{e}^{\eta|\sigma^{-1}(\cdot)Z(\cdot,i)|^2}\Big) < \infty. \tag{1.22}$$

Then

$$W_{bL}(\mathcal{L}(X_t), \mathcal{L}(\widetilde{X}_t)) \le C\max\Big\{\|Q-\widetilde{Q}\|_{\ell_1}^{\frac{1}{2q_0}}, \|Q-\widetilde{Q}\|_{\ell_1}^{\frac{1}{2q_0\gamma}}\Big\}, \quad t\in[0,T], \tag{1.23}$$

for some constant C depending on $N, T, x_0, \tau_1, K_0, \gamma, p_0$ and $\max_{i\in\mathcal{S}}\mu_0\Big(\mathrm{e}^{\eta|\sigma^{-1}(\cdot)Z(\cdot,i)|^2}\Big)$, where $p_0 > 1$ is a constant satisfying $2p_0^2Td < \eta$, $q_0 = p_0/(p_0-1)$ and $\gamma > 1$ is a constant.

Theorem 1.4 *Let (X_t, Λ_t) be a solution of (1.14) and $(\hat{X}_t, \hat{\Lambda}_t)$ a solution of (1.17). Suppose $V \in C^2(\mathbb{R}^d)$ satisfying condition (A). Let $T > 0$ be fixed. Assume there exists a constant $\eta > 2Td$ such that (1.22) holds. Suppose (1.4) holds. Then*

$$\begin{aligned} &W_{bL}(\mathcal{L}(X_t), \mathcal{L}(\hat{X}_t)) \\ &\le C \max\Big\{ \big(\|B\|_{\ell_1} + \|Q_1 - \widehat{Q}\|_{\ell_1}\big)^{\frac{1}{2q_0}}, \big(\|B\|_{\ell_1} + \|Q_1 - \widehat{Q}\|_{\ell_1}\big)^{\frac{1}{2q_0\gamma}} \Big\}, \quad t \in [0, T], \end{aligned} \tag{1.24}$$

for some constant C depending on $N, T, x_0, \tau_1, K_0, \gamma, p_0$ and $\max_{i\in\mathcal{S}} \mu_0\Big(\mathrm{e}^{\eta|\sigma^{-1}(\cdot)Z(\cdot,i)|^2}\Big)$, where $p_0 > 1$ is a constant satisfying $2p_0^2 Td < \eta$, $q_0 = p_0/(p_0 - 1)$ and $\gamma > 1$ is a constant.

2 Proofs of main results

2.1 SDEs with regular coefficients

Let us first introduce the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ used throughout this work. Let

$$\Omega_1 = \big\{\omega \big| \, \omega : [0, \infty) \to \mathbb{R}^d \text{ continuous}, \ \omega_0 = 0\big\},$$

which is endowed with the local uniform convergence topology and the Wiener measure $\mathbb{P}_1$ so that its coordinate process $W(t, \omega) = \omega(t)$, $t \ge 0$, is a d-dimensional Brownian motion. Put

$$\Omega_2 = \big\{\omega \big| \, \omega : [0, \infty) \to \mathcal{S} \text{ right continuous with left limits}\big\},$$

endowed with the Skorokhod topology and a probability measure $\mathbb{P}_2$. The Markov chains (Λ_t) and $(\tilde{\Lambda}_t)$ are all constructed in the space $(\Omega_2, \mathscr{B}(\Omega_2), \mathbb{P}_2)$. Set

$$(\Omega, \mathscr{F}, \mathbb{P}) = (\Omega_1 \times \Omega_2, \mathscr{B}(\Omega_1) \times \mathscr{B}(\Omega_2), \mathbb{P}_1 \times \mathbb{P}_2).$$

Thus under $\mathbb{P} = \mathbb{P}_1 \times \mathbb{P}_2$, (Λ_t) and $(\tilde{\Lambda}_t)$ are independent of the Brownian motion (W_t). Denote by $\mathbb{E}_{\mathbb{P}_1}$ taking the expectation with respect to the probability measure $\mathbb{P}_1$, and similarly $\mathbb{E}_{\mathbb{P}_2}$.

Next, we construct a coupling process $(\Lambda_t, \tilde{\Lambda}_t)$ such that (Λ_t) and $(\tilde{\Lambda}_t)$ are continuous-time Markov chains with transition rate matrix Q and $\widetilde{Q}$ respectively. Denote $H = \max_{i\in\mathcal{S}}\{q_i, \tilde{q}_i\}$ and $M = N(N-1)H$. Let ξ_k, $k = 1, 2, \ldots$, be random variables supported on $[0, M]$ satisfying $\mathbb{P}_2(\xi_k \in \mathrm{d}x) = \mathbf{m}(\mathrm{d}x)/M$ where $\mathbf{m}(\mathrm{d}x)$ stands for the Lebesgue measure on $[0, M]$. Let τ_k, $k = 1, 2, \ldots$, be nonnegative random variables such that $\mathbb{P}_2(\tau_k > t) = \exp(-Mt)$, $t \geq 0$. Suppose that $\{\xi_k\}$ and $\{\tau_k\}$ are mutually independent. Let

$$\zeta_1 = \tau_1, \zeta_2 = \tau_1 + \tau_2, \ldots, \zeta_k = \tau_1 + \tau_2 + \ldots + \tau_k, \quad k \geq 1,$$

and

$$\mathcal{D}_{p_1} = \{\zeta_1, \zeta_2, \ldots, \zeta_k, \ldots\}.$$

After constructing such random variables, define

$$p_1(\zeta_k) = \xi_k, \quad k \geq 1,$$

and further define the Poisson random measure

$$N_1((0, t] \times U) = \#\big\{s \in \mathcal{D}_{p_1}; s \leq t,\, p_1(s) \in U\big\},\ t > 0,\, U \in \mathscr{B}(\mathbb{R}).$$

Construct two families of left-closed, right-open intervals $\{\Gamma_{ij}\}_{i,j\in\mathcal{S}}$ and $\{\widetilde{\Gamma}_{ij}\}_{i,j\in\mathcal{S}}$ on the half line in the following manner:

$$\begin{aligned}
&\Gamma_{12} = [0, q_{12}), && \widetilde{\Gamma}_{12} = [0, \tilde{q}_{12}),\\
&\Gamma_{13} = [q_{12}, q_{12} + q_{13}), && \widetilde{\Gamma}_{13} = [\tilde{q}_{12}, \tilde{q}_{12} + \tilde{q}_{13}),\\
&\quad\cdots\cdots\\
&\Gamma_{21} = [q_1, q_1 + q_{21}), && \widetilde{\Gamma}_{21} = [\tilde{q}_1, \tilde{q}_1 + \tilde{q}_{21}),
\end{aligned}$$

and so on. For convenience of notation, put $\Gamma_{ii} = \widetilde{\Gamma}_{ii} = \emptyset$, and $\Gamma_{ij} = \emptyset$ if $q_{ij} = 0$; $\widetilde{\Gamma}_{ij} = \emptyset$ if $\tilde{q}_{ij} = 0$. Define functions h, $\tilde{h} : \mathcal{S} \times \mathbb{R} \to \mathbb{R}$ by

$$\begin{aligned}
h(i, z) &= \sum_{\ell\in\mathcal{S}} (\ell - i)\mathbf{1}_{\Gamma_{i\ell}}(z),\\
\tilde{h}(i, z) &= \sum_{\ell\in\mathcal{S}} (\ell - i)\mathbf{1}_{\widetilde{\Gamma}_{i\ell}}(z).
\end{aligned}$$

Then, according to [24, Chapter II] or [30], the solution of the SDE

$$\mathrm{d}\Lambda_t = \int_{[0,M]} h(\Lambda_{t-}, z) N_1(\mathrm{d}t, \mathrm{d}z), \quad \Lambda_0 = i_0, \tag{2.1}$$

is a continuous-time Markov chain with transition rate matrix $Q = (q_{ij})$. Similarly, the solution of the SDE

$$\mathrm{d}\tilde{\Lambda}_t = \int_{[0,M]} \tilde{h}(\tilde{\Lambda}_{t-}, z) N_1(\mathrm{d}t, \mathrm{d}z), \quad \tilde{\Lambda}_0 = i_0, \tag{2.2}$$

is a continuous-time Markov chain with transition rate matrix $\widetilde{Q} = (\tilde{q}_{ij})$. Therefore, through the SDEs (2.1) and (2.2), we construct the desired coupling process $(\Lambda_t, \tilde{\Lambda}_t)$. Furthermore, consider the following SDEs:

$$\mathrm{d}X_t = b(X_t, \Lambda_t)\mathrm{d}t + \sigma(X_t, \Lambda_t)\mathrm{d}W_t, \quad X_0 = x_0,\ \Lambda_0 = i_0, \tag{2.3}$$

$$\mathrm{d}\widetilde{X}_t = b(\widetilde{X}_t, \tilde{\Lambda}_t)\mathrm{d}t + \sigma(\widetilde{X}_t, \tilde{\Lambda}_t)\mathrm{d}W_t, \quad \widetilde{X}_0 = x_0,\ \tilde{\Lambda}_0 = i_0. \tag{2.4}$$

Then, the system (X_t, Λ_t) given by (2.3) and (2.1) has the same distribution as the system given in (1.1). Similarly, $(\widetilde{X}_t, \tilde{\Lambda}_t)$ given by (2.4) and (2.2) has the same distribution as the system given in (1.2). Under the help of the constructed systems (X_t, Λ_t) and $(\widetilde{X}_t, \tilde{\Lambda}_t)$ in this section, we can provide the proof of Theorem 1.1.

Lemma 2.1 *Let (X_t, Λ_t), $(\widetilde{X}_t, \tilde{\Lambda}_t)$ be the solution of (2.3) and (2.4) respectively with $X_0 = \widetilde{X}_0 = x_0 \in \mathbb{R}^d$. Assume* (H2) *holds. Then, for $\mathbb{P}_2$-almost surely $\omega_2 \in \Omega_2$,*

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}_1}[|X_t|^2](\omega_2) \le (|x_0|^2 + 2Kt)\mathrm{e}^{(2K+1)t}, \\ &\mathbb{E}_{\mathbb{P}_1}[|\widetilde{X}_t|^2](\omega_2) \le (|x_0|^2 + 2Kt)\mathrm{e}^{(2K+1)t}, \quad t > 0. \end{aligned} \tag{2.5}$$

Proof. By Itô's formula and (H2),

$$\begin{aligned} \mathrm{d}|X_t|^2 &= \big[2\langle X_t, b(X_t, \Lambda_t)\rangle + \|\sigma(X_t, \Lambda_t)\|_{\mathrm{HS}}^2\big]\mathrm{d}t + 2\langle X_t, \sigma(X_t, \Lambda_t)\mathrm{d}W_t\rangle \\ &\le \big[|X_t|^2 + 2K(1 + |X_t|^2)\big]\mathrm{d}t + 2\langle X_t, \sigma(X_t, \Lambda_t)\mathrm{d}W_t\rangle. \end{aligned}$$

Taking the expectation w.r.t. $\mathbb{P}_1$ and using Gronwall's inequality, we obtain

$$\mathbb{E}_{\mathbb{P}_1}[|X_t|^2](\omega_2) \le (|x_0|^2 + 2Kt)\mathrm{e}^{(2K+1)t}, \quad \mathbb{P}_2\text{-a.s. } \omega_2.$$

Similarly, the estimate on $\mathbb{E}_{\mathbb{P}_1}[|\widetilde{X}_t|^2](\omega_2)$ holds. □

Lemma 2.2 *For the processes (Λ_t) and $(\tilde{\Lambda}_t)$ given in (2.1) and (2.2) respectively, it holds*

$$\int_0^t \mathbb{P}(\Lambda_s \neq \tilde{\Lambda}_s)\mathrm{d}s \le N^2 t^2 \|Q - \widetilde{Q}\|_{\ell_1}. \tag{2.6}$$

Proof. Let $\Gamma_{ij}\Delta\widetilde{\Gamma}_{ij}=\big(\Gamma_{ij}\backslash\widetilde{\Gamma}_{ij}\big)\bigcup\big(\widetilde{\Gamma}_{ij}\backslash\Gamma_{ij}\big)$. By virtue of the construction of Γ_{ij} and $\widetilde{\Gamma}_{ij}$, we have

$$\begin{aligned}\mathbf{m}(\Gamma_{ij}\Delta\widetilde{\Gamma}_{ij})&\le\Big|\sum_{k=1}^{i-1}q_k+\sum_{k=1,k\neq i}^{j-1}q_{ik}-\sum_{k=1}^{i-1}\tilde{q}_k-\sum_{k=1,k\neq i}^{j-1}\tilde{q}_{ik}\Big|\\&\quad+\Big|\sum_{k=1}^{i-1}q_k+\sum_{k=1,k\neq i}^{j}q_{ik}-\sum_{k=1}^{i-1}\tilde{q}_k-\sum_{k=1,k\neq i}^{j}\tilde{q}_{ik}\Big|\\&\le 2(i-1)\|Q-\widetilde{Q}\|_{\ell_1}+\|Q-\widetilde{Q}\|_{\ell_1}\\&\le 2N\|Q-\widetilde{Q}\|_{\ell_1}.\end{aligned}$$

See also [21] for more details on previous calculation.

For $\delta\in(0,1)$ and $s>0$, let $s_\delta=[\frac{s}{\delta}]$, the integer part of s/δ. Let $N(t)=N_1((0,t]\times\mathbb{R})$. For every $t\in(0,\delta]$, since $\Lambda_0=\tilde{\Lambda}_0=i_0$, we have

$$\begin{aligned}\mathbb{P}(\Lambda_t\neq\tilde{\Lambda}_t)&=\mathbb{P}(\Lambda_t\neq\tilde{\Lambda}_t,N(t)\ge 1))\\&=\mathbb{P}(\Lambda_t\neq\tilde{\Lambda}_t,N(t)=1)+\mathbb{P}(\Lambda_t\neq\tilde{\Lambda}_t,N(t)\ge 2).\end{aligned}$$

There is a constant $C>0$ such that

$$\mathbb{P}(N(t)\ge 2)\le\mathbb{P}(N(\delta)\ge 2)=1-\mathrm{e}^{-M\delta}-M\delta\mathrm{e}^{-M\delta}\le C\delta^2.\tag{2.7}$$

On the other hand,

$$\begin{aligned}\mathbb{P}(\Lambda_t\neq\tilde{\Lambda}_t,N(t)=1)&=\int_0^t\mathbb{P}(\Lambda_t\neq\tilde{\Lambda}_t,\tau_1\in\mathrm{d}s,\,\tau_2>t-s)\\&=\int_0^t\mathbb{P}\Big(\xi_1\notin\bigcup_{j\in\mathcal{S}}\big(\Gamma_{i_0j}\bigcap\widetilde{\Gamma}_{i_0j}\big),\tau_1\in\mathrm{d}s\Big)\mathrm{e}^{-M(t-s)}\\&\le 2N^2t\mathrm{e}^{-Mt}\|Q-\widetilde{Q}\|_{\ell_1}.\end{aligned}$$

Hence,

$$\mathbb{P}(\Lambda_t\neq\tilde{\Lambda}_t)\le C\delta^2+2N^2\delta\|Q-\widetilde{Q}\|_{\ell_1},\quad 0<t\le\delta.\tag{2.8}$$

Note that the estimate is independent of the common initial value of (Λ_t) and $(\tilde{\Lambda}_t)$.

To proceed,

$$\mathbb{P}(\Lambda_{2\delta}\neq\tilde{\Lambda}_{2\delta})=\mathbb{P}(\Lambda_{2\delta}\neq\tilde{\Lambda}_{2\delta},\Lambda_\delta=\tilde{\Lambda}_\delta)+\mathbb{P}(\Lambda_{2\delta}\neq\tilde{\Lambda}_{2\delta},\Lambda_\delta\neq\tilde{\Lambda}_\delta)$$

$$\le \mathbb{P}(\Lambda_{2\delta} \neq \tilde{\Lambda}_{2\delta} \big| \Lambda_\delta = \tilde{\Lambda}_\delta) + \mathbb{P}(\Lambda_\delta \neq \tilde{\Lambda}_\delta).$$

By the time-homogeneity of $(\Lambda_t, \tilde{\Lambda}_t)$ and the estimate (2.8), it follows that

$$\mathbb{P}(\Lambda_{2\delta} \neq \tilde{\Lambda}_{2\delta}) \le 2C\delta^2 + 4N^2\delta \|Q - \widetilde{Q}\|_{\ell_1}.$$

Deduce inductively to yield that, for each $k \ge 2$,

$$\mathbb{P}(\Lambda_{k\delta} \neq \tilde{\Lambda}_{k\delta}) \le kC\delta^2 + 2kN^2\delta \|Q - \widetilde{Q}\|_{\ell_1}. \tag{2.9}$$

By virtue of (2.8) and (2.9), we have that for $t > 0$,

$$\begin{aligned}
\int_0^t \mathbb{P}(\Lambda_s \neq \tilde{\Lambda}_s)\mathrm{d}s &= \int_0^t \mathbb{P}(\Lambda_s \neq \tilde{\Lambda}_s, \Lambda_{s_\delta} = \tilde{\Lambda}_{s_\delta})\mathrm{d}s + \int_0^t \mathbb{P}(\Lambda_s \neq \tilde{\Lambda}_s, \Lambda_{s_\delta} \neq \tilde{\Lambda}_{s_\delta})\mathrm{d}s \\
&\le \int_0^t \mathbb{P}(\Lambda_s \neq \tilde{\Lambda}_s \big| \Lambda_{s_\delta} = \tilde{\Lambda}_{s_\delta})\mathbb{P}(\Lambda_{s_\delta} = \tilde{\Lambda}_{s_\delta})\mathrm{d}s + \int_0^t \mathbb{P}(\Lambda_{s_\delta} \neq \tilde{\Lambda}_{s_\delta})\mathrm{d}s \\
&\le \int_0^t \mathbb{P}(\Lambda_s \neq \tilde{\Lambda}_s \big| \Lambda_{s_\delta} = \tilde{\Lambda}_{s_\delta})\mathrm{d}s + \sum_{k=1}^K \mathbb{P}(\Lambda_{k\delta} \neq \tilde{\Lambda}_{k\delta})\delta \\
&\le C\delta^2 t + 2N^2\delta t \|Q - \widetilde{Q}\|_{\ell_1} + \frac{C\delta^3}{2}K(K+1) \\
&\qquad + N^2 K(K+1)\delta^2 \|Q - \widetilde{Q}\|_{\ell_1},
\end{aligned}$$

where $K = \left[\frac{t}{\delta}\right] + 1$. Letting $\delta \downarrow 0$, we obtain that

$$\int_0^t \mathbb{P}(\Lambda_s \neq \tilde{\Lambda}_s)\mathrm{d}s \le N^2 t^2 \|Q - \widetilde{Q}\|_{\ell_1},$$

which concludes the proof. □

Remark 2.3 The perturbation theory of continuous-time Markov chains has been developed in many works; see, e.g. [16, 17] and references therein. According to this theory, one can get appropriate estimate of the distance between two transition semigroups by the distance between their corresponding transition rate matrices. Whereas, to control the term $\mathbb{E}\int_0^t \mathbf{1}_{\{\Lambda_s \neq \tilde{\Lambda}_s\}}\mathrm{d}s$ which concerns the behavior of Markov chains during a time interval $[0, t]$ rather than a fixed time t, one has to construct a suitable coupling process. One possible method is to use the optimal coupling for continuous-time Markov chains (cf. [7, Chapter 5]). But additional conditions on the generator of the coupling process are needed. However, we do not find an explicit condition in terms of the difference between Q and $\widetilde{Q}$, for example, $\|Q - \widetilde{Q}\|_{\ell_1}$ used in this work at current stage. Our result shows once again the significant effect of Skorkhod's representation of continuous-time Markov chains which has been applied in [22] to deal with state-dependent regime-switching processes.

Proof of Theorem 1.1 For simplicity of notation, let $Z_t = X_t - \widetilde{X}_t$. Then, due to (H1) and (H2), Itô's formula yields that

$$
\begin{aligned}
\mathrm{d}|Z_t|^2 &= \big\{2\langle Z_t, b(X_t,\Lambda_t) - b(\widetilde{X}_t,\tilde{\Lambda}_t)\rangle + \|\sigma(X_t,\Lambda_t) - \sigma(\widetilde{X}_t,\tilde{\Lambda}_t)\|_{\mathrm{HS}}^2\big\}\mathrm{d}t + \mathrm{d}M_t\\
&\le \big\{\kappa_{\Lambda_t}|Z_t|^2 + 2\langle Z_t, b(\widetilde{X}_t,\Lambda_t) - b(\widetilde{X}_t,\tilde{\Lambda}_t)\rangle + 2\|\sigma(\widetilde{X}_t,\Lambda_t) - \sigma(\widetilde{X}_t,\tilde{\Lambda}_t)\|_{\mathrm{HS}}^2\big\}\mathrm{d}t + \mathrm{d}M_t\\
&\le \big\{(\kappa_{\Lambda_t}+\varepsilon)|Z_t|^2 + \frac{1}{\varepsilon}\big(|b(\widetilde{X}_t,\Lambda_t)| + |b(\widetilde{X}_t,\tilde{\Lambda}_t)|\big)^2\mathbf{1}_{\{\Lambda_t\neq\tilde{\Lambda}_t\}}\\
&\qquad + 4\big(\|\sigma(\widetilde{X}_t,\Lambda_t)\|_{\mathrm{HS}}^2 + \|\sigma(\widetilde{X}_t,\tilde{\Lambda}_t)\|_{\mathrm{HS}}^2\big)\mathbf{1}_{\{\Lambda_t\neq\tilde{\Lambda}_t\}}\big\}\mathrm{d}t + \mathrm{d}M_t\\
&\le \big\{(\kappa_{\Lambda_t}+\varepsilon)|Z_t|^2 + \frac{4K}{\varepsilon}(1+|\widetilde{X}_t|^2)\mathbf{1}_{\{\Lambda_t\neq\tilde{\Lambda}_t\}} + 8K(1+|\widetilde{X}_t|^2)\mathbf{1}_{\{\Lambda_t\neq\tilde{\Lambda}_t\}}\big\}\mathrm{d}t + \mathrm{d}M_t
\end{aligned}
$$

for any $\varepsilon > 0$, where $M_t = \int_0^t 2\langle Z_s, (\sigma(X_s,\Lambda_s) - \sigma(\widetilde{X}_s,\tilde{\Lambda}_s))\mathrm{d}W_s\rangle$ for $t \ge 0$ is a martingale. Taking the expectation w.r.t. $\mathbb{P}_1$ on both sides of the previous inequality, we get

$$
\begin{aligned}
\mathrm{d}\,\mathbb{E}_{\mathbb{P}_1}[|Z_t|^2](\omega_2) \le{}& \big(4\varepsilon^{-1}+8\big)K\mathbb{E}_{\mathbb{P}_1}\big[1+|\widetilde{X}_t|^2\big](\omega_2)\mathbf{1}_{\{\Lambda_t\neq\tilde{\Lambda}_t\}}(\omega_2)\mathrm{d}t\\
&+ (\kappa_{\Lambda_t}+\varepsilon)(\omega_2)\mathbb{E}_{\mathbb{P}_1}[|Z_t|^2](\omega_2)\mathrm{d}t.
\end{aligned}
\tag{2.10}
$$

To proceed, let us recall an elementary inequality. Let $u(t)$ be a real-valued differentiable function, $\alpha(t)$ and $\beta(t)$ real-valued integrable functions (not necessary nonnegative). If

$$u'(t) \le \alpha(t) + \beta(t)u(t),$$

then

$$u(t) \le u(0)\mathrm{e}^{\int_0^t \beta(s)\mathrm{d}s} + \int_0^t \alpha(s)\mathrm{e}^{\int_s^t \beta(r)\mathrm{d}r}\mathrm{d}s.$$

Using this inequality to (2.10), and invoking the estimate in Lemma 2.1, we obtain that

$$\mathbb{E}_{\mathbb{P}_1}[|Z_t|^2](\omega_2) \le \big(4\varepsilon^{-1}+8\big)K\int_0^t\Big(1+(|x_0|^2+2Ks)\mathrm{e}^{(2K+1)s}\Big)\mathbf{1}_{\{\Lambda_s\neq\tilde{\Lambda}_s\}}\,\mathrm{e}^{\int_s^t(\kappa_{\Lambda_r}+\varepsilon)(\omega_2)\mathrm{d}r}\mathrm{d}s.$$

Taking the expectation w.r.t. $\mathbb{P}_2$ and using Hölder's inequality, we get

$$
\begin{aligned}
\mathbb{E}|Z_t|^2 \le \int_0^t &\Big\{(4\varepsilon^{-1}+8)K\big[1+(|x_0|^2+2Ks)\mathrm{e}^{(2K+1)s}\big]\\
&\cdot\big(\mathbb{E}\mathbf{1}_{\{\Lambda_s\neq\tilde{\Lambda}_s\}}(\omega_2)\big)^{\frac{1}{q}}\big(\mathbb{E}\mathrm{e}^{p\int_s^t(\kappa_{\Lambda_r}+\varepsilon)(\omega_2)\mathrm{d}r}\big)^{\frac{1}{p}}\Big\}\mathrm{d}s
\end{aligned}
\tag{2.11}
$$

for p, $q > 1$ with $1/p + 1/q = 1$.

In order to estimate the term $\mathbb{E}\,\mathrm{e}^{p\int_0^t(\kappa_{\Lambda_s}+1)\mathrm{d}s}$, we need the following notation. Let

$$Q_p = Q + p\,\mathrm{diag}(\kappa_0, \kappa_1, \ldots, \kappa_N),$$

and

$$\eta_p = -\max\big\{\mathrm{Re}(\gamma);\ \gamma\in\mathrm{spec}(Q_p)\big\}.$$

According to [2, Proposition 4.1], for any $p>0$, there exist two positive constants $C_1(p)$ and $C_2(p)$ such that

$$C_1(p)\mathrm{e}^{-\eta_p t} \le \mathbb{E}\,\mathrm{e}^{p\int_0^t \kappa_{\Lambda_s}\mathrm{d}s} \le C_2(p)\mathrm{e}^{-\eta_p t},\quad t>0. \tag{2.12}$$

The term $\int_0^t \mathbb{E}\mathbf{1}_{\{\Lambda_s\neq\tilde\Lambda_s\}}\mathrm{d}s$ is estimated in Lemma 2.2. Consequently, substituting the estimates (2.12) and (2.6) into (2.11), we get

$$\begin{aligned}\mathbb{E}[|Z_t|^2] \le \big(4\varepsilon^{-1}+8\big)KC_2(p)^{\frac1p}\Big(N^2t^2\|Q-\widetilde Q\|_{\ell_1}\Big)^{\frac1q}\\ \cdot\Big(\int_0^t\big[1+(|x_0|^2+2Ks)\mathrm{e}^{(2K+1)s}\big]^p\mathrm{e}^{-(\eta_p-\varepsilon p)(t-s)}\mathrm{d}s\Big)^{\frac1p}.\end{aligned} \tag{2.13}$$

Note that the solutions of (2.3) and (2.4) exist uniquely. Then the distribution of $(X_t,\widetilde X_t)$ on $\mathbb{R}^d\times\mathbb{R}^d$ is a coupling of $\mathcal{L}(X_t)$ and $\mathcal{L}(\widetilde X_t)$. By the definition of the Wasserstein distance, it follows

$$\begin{aligned}W_2(\mathcal{L}(X_t),\mathcal{L}(\widetilde X_t))^2 &\le \mathbb{E}[|X_t-\widetilde X_t|^2]\\ &\le \big(4\varepsilon^{-1}+8\big)KC_2(p)^{\frac1p}N^{\frac2q}t^{\frac2q}\|Q-\widetilde Q\|_{\ell_1}^{\frac1q}\\ &\quad\cdot\Big(\int_0^t\big[1+(|x_0|^2+2Ks)\mathrm{e}^{(2K+1)s}\big]^p\mathrm{e}^{-(\eta_p-\varepsilon p)(t-s)}\mathrm{d}s\Big)^{\frac1p},\end{aligned}$$

which is the desired estimate (1.8).

When b and σ are bounded satisfying (1.10), we have a simple estimate

$$\mathrm{d}|Z_t|^2 \le \big\{(\kappa_{\Lambda_t}+\varepsilon)|Z_t|^2+4K(2+\varepsilon^{-1})\mathbf{1}_{\{\Lambda_t\neq\tilde\Lambda_t\}}\big\}\mathrm{d}t+\mathrm{d}M_t,$$

where $M_t=\int_0^t 2\langle Z_s,(\sigma(X_s,\Lambda_s)-\sigma(\widetilde X_s,\tilde\Lambda_s))\mathrm{d}W_s\rangle$, $t\ge 0$. This yields

$$\mathbb{E}|Z_t|^2 \le (4\varepsilon^{-1}+8)K\Big(\int_0^t\mathbb{P}(\Lambda_s\neq\tilde\Lambda_s)\mathrm{d}s\Big)^{\frac1q}\Big(\int_0^t\mathbb{E}\mathrm{e}^{p\int_s^t(\kappa_{\Lambda_r}+\varepsilon)\mathrm{d}r}\mathrm{d}s\Big)^{\frac1p}.$$

Then, (1.11) can be established by following the same procedure to deduce (1.8). □

Proof of Theorem 1.2 To emphasize the idea, we give out the proof in the situation $E = \mathcal{S}\backslash\{0\}$. For the given transition rate matrices $Q = (q_{ij})_{i,j\in\mathcal{S}}$ on $\mathcal{S}$ and $\widehat{Q} = (\hat{q}_{ij})_{i,j\in E}$ on E, write Q in the form

$$Q = \begin{pmatrix} -q_0 & \alpha \\ \beta & Q_1 \end{pmatrix}, \tag{2.14}$$

where $\alpha = \{q_{0i}; 1 \le i \le N\}$ and $\beta = \{q_{j0}; 1 \le j \le N\}$ are the row and column vectors on E. Let (Λ_t) and $(\widehat{\Lambda}_t)$ be the Markov chains on $\mathcal{S}$ and E with the transition rate matrices Q and $\widehat{Q}$ respectively. Consider

$$\mathrm{d}\hat{X}_t = b(\hat{X}_t, \hat{\Lambda}_t)\mathrm{d}t + \sigma(\hat{X}_t, \hat{\Lambda}_t)\mathrm{d}W_t, \quad \hat{X}_0 = x_0, \quad \hat{\Lambda}_0 = i_0 \in E. \tag{2.15}$$

In order to employ the method used in Theorem 1.1, we propose the following extension

$$\widetilde{Q} = \begin{pmatrix} -q_0 & \alpha \\ 0 & \widehat{Q} \end{pmatrix}. \tag{2.16}$$

It is easy to see that $\widetilde{Q}$ is conservative. Hence, there is a unique semigroup $(\widetilde{P}_t)_{t\ge0}$ on $\mathcal{S}$ corresponding to the generator $\widetilde{Q}$. This $(\tilde{\Lambda}_t)$ helps us to define another dynamical system $(\widetilde{X}_t)$ by the following SDE:

$$\mathrm{d}\widetilde{X}_t = b(\widetilde{X}_t, \tilde{\Lambda}_t)\mathrm{d}t + \sigma(\widetilde{X}_t, \tilde{\Lambda}_t)\mathrm{d}W_t, \quad \widetilde{X}_0 = x_0, \ \tilde{\Lambda}_0 = i_0 \in E. \tag{2.17}$$

Under the conditions (H1) and (H2), the solutions of SDEs (2.15) and (2.17) are uniquely determined. Due to the definition of $\widetilde{Q}$ in (2.16), the process $(\tilde{\Lambda}_t)$ starting from $i_0 \in E$ will never reach the point 0, thus $\tilde{\Lambda}_t = \hat{\Lambda}_t$, $t > 0$, a.s. when $\tilde{\Lambda}_0 = \hat{\Lambda}_0 = i_0 \in E$. As a consequence,

$$\widetilde{X}_t = \hat{X}_t, \quad t > 0, \quad a.s. \tag{2.18}$$

Moreover, by virtue of (2.14) and (2.16), it holds

$$\|Q - \widetilde{Q}\|_{\ell_1} \le \|\beta\|_{\ell_1} + \|Q_1 - \widehat{Q}\|_{\ell_1}. \tag{2.19}$$

Following the procedure of the argument of Theorem 1.1, inserting (2.19) into (2.13), we obtain that

$$\begin{aligned} \mathbb{E}[|X_t - \widetilde{X}_t|^2] \le{}& \big(4\varepsilon^{-1}+8\big)KC_2(p)^{\frac{1}{p}}\big(Nt\big)^{\frac{2}{q}}\Big(\|\beta\|_{\ell_1} + \|Q_1 - \widehat{Q}\|_{\ell_1}\Big)^{\frac{1}{q}} \\ &\cdot \Big(\int_0^t \big[1 + (|x_0|^2 + 2Ks)\mathrm{e}^{(2K+1)s}\big]^p \mathrm{e}^{-(\eta_p - \varepsilon p)(t-s)}\mathrm{d}s\Big)^{\frac{1}{p}}. \end{aligned} \tag{2.20}$$

Due to (2.18), it follows that $\mathbb{E}[|X_t-\hat{X}_t|^2]=\mathbb{E}[|X_t-\widetilde{X}_t|^2]$. According to the definition of the Wasserstein distance, and using the estimate (2.20), we obtain

$$\begin{aligned}W_2(\mathcal{L}(X_t),\mathcal{L}(\hat{X}_t))^2 &\leq \big(4\varepsilon^{-1}+8\big)KC_2(p)^{\frac{1}{p}}\big(Nt\big)^{\frac{2}{q}}\Big(\|\beta\|_{\ell_1}+\|Q_1-\widehat{Q}\|_{\ell_1}\Big)^{\frac{1}{q}}\\ &\quad\cdot\Big(\int_0^t\big[1+(|x_0|^2+2Ks)\mathrm{e}^{(2K+1)s}\big]^p\mathrm{e}^{-(\eta_p-\varepsilon p)(t-s)}\mathrm{d}s\Big)^{\frac{1}{p}}.\end{aligned}\tag{2.21}$$

Analogously, if b and σ are bounded satisfying (1.10), we have

$$\begin{aligned}&W_2(\mathcal{L}(X_t),\mathcal{L}(\hat{X}_t))^2\leq\mathbb{E}[|X_t-\widetilde{X}_t|^2]\\ &\leq(4\varepsilon^{-1}+8)KC_2(p)^{\frac{1}{p}}\big(Nt\big)^{\frac{2}{q}}\Big(\|\beta\|_{\ell_1}+\|Q_1-\widehat{Q}\|_{\ell_1}\Big)^{\frac{1}{q}}\Big(\frac{1-\mathrm{e}^{-(\eta_p-\varepsilon p)t}}{\eta_p-\varepsilon p}\Big)^{\frac{1}{p}}.\end{aligned}\tag{2.22}$$

This completes the proof in the situation $E=\mathcal{S}\backslash\{0\}$. The general case can be proved in the same way, and the details are omitted.

2.2 SDEs with irregular coefficients

In this part, we consider the regime-switching processes with irregular drifts. Precisely, consider

$$\mathrm{d}X_t=b(X_t,\Lambda_t)\mathrm{d}t+\sigma(X_t)\mathrm{d}W_t,\quad X_0=x_0,\ \Lambda_0=i_0,\tag{2.23}$$

where $b:\mathbb{R}^d\times\mathcal{S}\to\mathbb{R}^d$ and $\sigma:\mathbb{R}^d\to\mathbb{R}^{d\times d}$. Here, we assume that the diffusion coefficient σ satisfies the Lipschitz condition: there exists $K>0$ such that

$$\|\sigma(x)-\sigma(y)\|_{\mathrm{HS}}^2\leq K|x-y|^2,\quad\forall\, x,y\in\mathbb{R}^d.\tag{2.24}$$

However, the drift b is assumed to satisfy certain integrability condition. Hence, it may be discontinuous. (Λ_t) is a continuous time Markov chain on $\mathcal{S}$ with the transition rate matrix $Q=(q_{ij})_{i,j\in\mathcal{S}}$. Consider the perturbation $\widetilde{Q}=(\tilde{q}_{ij})_{i,j\in\mathcal{S}}$ of Q and its associated Markov chain $(\tilde{\Lambda}_t)$. Let

$$\mathrm{d}\widetilde{X}_t=b(\widetilde{X}_t,\tilde{\Lambda}_t)\mathrm{d}t+\sigma(\widetilde{X}_t)\mathrm{d}W_t,\quad\widetilde{X}_0=x_0,\ \tilde{\Lambda}_0=i_0.\tag{2.25}$$

The integrability condition of type (1.22) is raised by Wang [28] to study the nonexplosion of the solutions of SDEs by using the dimension-free Harnack inequality. We will use the technique of [28] to analyze the stability of the regime-switching processes. Moreover, according to [28, Theorem 2.1] and using the technique to construct the regime-switching processes with Markovian switching (cf. e.g. [15]), it is standard to show the existence and uniqueness of the solutions of SDEs (2.23) and (2.25).

To proceed, we make some necessary preparations. Let (Y_t) be a process associated with the reference function $V \in C^2(\mathbb{R}^d)$:

$$\mathrm{d}Y_t = Z_0(Y_t)\mathrm{d}t + \sigma \mathrm{d}W(t), \quad Y_0 = x_0, \tag{2.26}$$

where the vector field Z_0 is defined by (1.19). Since Z_0 is globally Lipschitz continuous by condition (A), there is a unique nonexplosive solution to SDE (2.26). Via the process (Y_t), a new representation for (X_t) and $(\widetilde{X}_t)$ can be constructed with the help of the Girsanov theorem, which is verified by the dimension-free Harnack inequality for (Y_t) under appropriate integrability conditions.

Precisely, rewrite (2.26) as

$$\mathrm{d}Y_t = b(Y_t, \Lambda_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t^{(1)},$$

where

$$W_t^{(1)} = W_t - \int_0^t \sigma(Y_s)^{-1} Z(Y_s, \Lambda_s)\mathrm{d}s,\ \ Z(y,i) = b(y,i) - Z_0(y), t > 0, y \in \mathbb{R}^d, i \in \mathcal{S}. \tag{2.27}$$

If Novikov's condition

$$\mathbb{E}\mathrm{e}^{\frac{1}{2}\int_0^T |\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s)|^2 \mathrm{d}s} < \infty \tag{2.28}$$

holds, then

$$\mathbb{Q} := \exp\Big(\int_0^T \langle \sigma^{-1}(Y_s)Z(Y_s,\Lambda_s), \mathrm{d}W_s\rangle - \frac{1}{2}\int_0^T |\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s)|^2 \mathrm{d}s\Big)\mathbb{P} \tag{2.29}$$

is a new probability measure. Thus, the Girsanov theorem yields that $(W_t^{(1)})_{t\in[0,T]}$ is a new Brownian motion under the probability measure $\mathbb{Q}$. Note that the mutual independence between (W_t) and (Λ_t) has been used herein. Consequently, the uniqueness of the solution for the SDE (2.23) tells us that $(Y_t, \Lambda_t)_{t\in[0,T]}$ under $\mathbb{Q}$ has the same distribution as that of $(X_t, \Lambda_t)_{t\in[0,T]}$ under $\mathbb{P}$. To be more precise, let us show that (Λ_t) and $(W_t^{(1)})$ are mutually independent under $\mathbb{Q}$. For any bounded measurable functions f on $\mathcal{S}$ and g on $\mathbb{R}^d$, it holds

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}}\big[f(\Lambda_t)g(W_t^{(1)})\big] = \mathbb{E}_{\mathbb{P}}\Big[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}f(\Lambda_t)g(W_t^{(1)})\Big] \\
&= \mathbb{E}_{\mathbb{P}_2}\Big[f(\Lambda_t)\mathbb{E}_{\mathbb{P}_2}\Big[\mathbb{E}_{\mathbb{P}_1}\Big(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}g(W_t^{(1)})\Big)\Big|\mathscr{F}_T^{\Lambda}\Big]\Big] \\
&= \mathbb{E}_{\mathbb{P}_2}\big[f(\Lambda_t)\mathbb{E}_{\mathbb{P}_1}\big[g(W_t)\big]\big] = \mathbb{E}_{\mathbb{P}_2}\big[f(\Lambda_t)\big]\mathbb{E}_{\mathbb{P}}\big[g(W_t)\big] \\
&= \mathbb{E}_{\mathbb{P}}\big[f(\Lambda_t)\big]\mathbb{E}_{\mathbb{Q}}\big[g(W_t^{(1)})\big],
\end{aligned} \tag{2.30}$$

where $\mathscr{F}_t^{\Lambda}$ denotes the σ-field generated by the process (Λ_s) up to time t, and $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ denotes the Radon-Nikodym derivative. Applying again the Grisanov theorem, we have $\mathbb{E}_{\mathbb{P}_1}\Big(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big) = 1$ and

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\big[f(\Lambda_t)\big] &= \mathbb{E}_{\mathbb{P}}\Big[f(\Lambda_t)\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big] \\
&= \mathbb{E}_{\mathbb{P}_2}\Big[f(\Lambda_t)\mathbb{E}_{\mathbb{P}_2}\Big[\mathbb{E}_{\mathbb{P}_1}\Big(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big)\Big|\mathscr{F}_T^{\Lambda}\Big]\Big] \\
&= \mathbb{E}_{\mathbb{P}_2}\big[f(\Lambda_t)\big] = \mathbb{E}_{\mathbb{P}}\big[f(\Lambda_t)\big].
\end{aligned}$$

Combining this with the previous equality (2.30), we have

$$\mathbb{E}_{\mathbb{Q}}\big[f(\Lambda_t)g(W_t^{(1)})\big] = \mathbb{E}_{\mathbb{Q}}\big[f(\Lambda_t)\big]\mathbb{E}_{\mathbb{Q}}\big[g(W_t^{(1)})\big],$$

and hence (Λ_t) and $(W_t^{(1)})$ are mutually independent.

Analogously, rewrite (Y_t) as

$$\mathrm{d}Y_t = b(Y_t,\tilde{\Lambda}_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}\widetilde{W}_t,$$

where

$$\widetilde{W}_t = W_t - \int_0^t \sigma(Y_s)^{-1}Z(Y_s,\tilde{\Lambda}_s)\mathrm{d}s. \tag{2.31}$$

If Novikov's condition

$$\mathbb{E}e^{\frac{1}{2}\int_0^T|\sigma^{-1}(Y_s)Z(Y_s,\tilde{\Lambda}_s)|^2\mathrm{d}s} < \infty \tag{2.32}$$

holds, then

$$\widetilde{\mathbb{Q}} := \exp\Big(\int_0^T\langle\sigma^{-1}(Y_s)Z(Y_s,\tilde{\Lambda}_s),\mathrm{d}W_s\rangle - \frac{1}{2}\int_0^T|\sigma^{-1}(Y_s)Z(Y_s,\tilde{\Lambda}_s)|^2\mathrm{d}s\Big)\mathbb{P} \tag{2.33}$$

is a new probability measure. Moreover, $(Y_t,\tilde{\Lambda}_t)_{t\in[0,T]}$ under $\widetilde{\mathbb{Q}}$ has the same distribution as that of $(\widetilde{X}_t,\tilde{\Lambda}_t)$ under $\mathbb{P}$.

Lemma 2.4 *Let $G : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}_+$ be a measurable function and $\beta > 0$ be a constant. Let $T > 0$ be fixed.*

(i) *If there exists a constant $\xi > d$ such that $\max_{i\in\mathcal{S}} \mu_0\big(G^\xi(\cdot,i)\big) < \infty$, then*

$$\mathbb{E}\Big[\int_0^T G(Y_s,\Lambda_s)\mathrm{d}s\Big] \le C \max_{i\in\mathcal{S}} \mu_0\big(G^\xi(\cdot,i)\big)^{\frac{1}{\xi}} < \infty \tag{2.34}$$

for some constant $C = C(T,\xi,K_0) > 0$.

(ii) *If there exists a constant η such that $\eta > \beta T d$ and $\max_{i\in\mathcal{S}} \mu_0\big(\mathrm{e}^{\eta G(\cdot,i)}\big) < \infty$, then*

$$\mathbb{E}\Big[\mathrm{e}^{\beta\int_0^T G(Y_s,\Lambda_s)\mathrm{d}s}\Big] < \infty. \tag{2.35}$$

Proof. We first prove (ii), then (i) follows easily from the derivation of (ii). Let P_t^0 denote the semigroup corresponding to the process $(Y(t))$ defined by (2.26) with initial value $Y(0) = x$. Hence, the semigroup P_t^0 is symmetric w.r.t. μ_0. Since V satisfies condition (A), according to [27, Theorem 1.1], for $p > 1$, the following Harnack inequality holds:

$$\Big(P_t^0 f(x)\Big)^p \le P_t^0 f^p(y) \exp\Big[\frac{K_0\sqrt{p}}{\sqrt{p}-1}\cdot\frac{|x-y|^2}{1-\mathrm{e}^{-K_0 t}}\Big], \quad \forall\, f \in \mathscr{B}_b^+(\mathbb{R}^d). \tag{2.36}$$

Applying the Harnack inequality (2.36) and the mutual independence between (Λ_t) and (W_t), we get for any $\gamma > 0$ and $K > 0$

$$\begin{aligned}\Big\{\mathbb{E}\Big[\mathrm{e}^{\gamma G(Y_t,\Lambda_t)\wedge K}\Big|\mathscr{F}_t^\Lambda\Big]\Big\}^p &= \Big\{P_t^0 \mathrm{e}^{\gamma G(\cdot,\Lambda_t)\wedge K}\Big\}^p(x)\\ &\le \Big\{P_t^0 \mathrm{e}^{\gamma p G(\cdot,\Lambda_t)\wedge K}\Big\}(y)\exp\Big[\frac{K_0\sqrt{p}}{\sqrt{p}-1}\cdot\frac{|x-y|^2}{1-\mathrm{e}^{-K_0 t}}\Big].\end{aligned}$$

Passing to the limit as $K \to +\infty$, it follows from Fatou's lemma that

$$\Big\{P_t^0 \mathrm{e}^{\gamma G(\cdot,\Lambda_t)}\Big\}^p(x) \le \Big\{P_t^0 \mathrm{e}^{\gamma p G(\cdot,\Lambda_t)}\Big\}(y)\exp\Big[\frac{K_0\sqrt{p}}{\sqrt{p}-1}\cdot\frac{|x-y|^2}{1-\mathrm{e}^{-K_0 t}}\Big]. \tag{2.37}$$

Denote $B(x,r) = \{y \in \mathbb{R}^d; |y-x| \le r\}$ for $r > 0$, $x \in \mathbb{R}^d$. Integrating both sides of (2.37) w.r.t. μ_0 over the set $B(x,\sqrt{1-\mathrm{e}^{-K_0t}})$, we obtain

$$\begin{aligned}
&\Big\{P_t^0\mathrm{e}^{\gamma G(\cdot,\Lambda_t)}(x)\Big\}^p\mu_0\big(B\big(x,\sqrt{1-\mathrm{e}^{-K_0t}}\big)\big)\\
&\le\int_{B\left(x,\sqrt{1-\mathrm{e}^{-K_0t}}\right)}\big\{P_t^0\mathrm{e}^{\gamma pG(\cdot,\Lambda_t)}\big\}(y)\mathrm{e}^{\frac{K_0\sqrt{p}}{\sqrt{p}-1}\cdot\frac{|x-y|^2}{1-\mathrm{e}^{-K_0t}}}\mu_0(\mathrm{d}y)\\
&\le\int_{B\left(x,\sqrt{1-\mathrm{e}^{-K_0t}}\right)}\big\{P_t^0\mathrm{e}^{\gamma pG(\cdot,\Lambda_t)}\big\}(y)\mathrm{e}^{\frac{K_0\sqrt{p}}{\sqrt{p}-1}}\mu_0(\mathrm{d}y)\\
&\le\mathrm{e}^{\frac{K_0\sqrt{p}}{\sqrt{p}-1}}\mu_0(\mathrm{e}^{\gamma pG(\cdot,\Lambda_t)}).
\end{aligned}\tag{2.38}$$

Since μ_0 has strictly positive and continuous density e^{-V} w.r.t. the Lebesgue measure, there exists $\Gamma\in C(\mathbb{R}^d;(0,\infty))$ such that $\mu_0(B(x,t))\ge\Gamma(x)t^d$ for $t\in(0,1]$ and $x\in\mathbb{R}^d$. Invoking (2.38), we obtain

$$\mathbb{E}\mathrm{e}^{\gamma G(Y_t,\Lambda_t)}\le\Gamma(x)^{-\frac{1}{p}}\mathrm{e}^{\frac{K_0}{p-\sqrt{p}}}\max_{i\in\mathcal{S}}\mu_0\Big(\mathrm{e}^{\gamma pG(\cdot,i)}\Big)^{\frac{1}{p}}\frac{1}{\big(1-\mathrm{e}^{-K_0t}\big)^{d/p}},\qquad t\in(0,T].\tag{2.39}$$

Combining this with Jensen's inequality, one has

$$\begin{aligned}
\mathbb{E}\Big[\mathrm{e}^{\beta\int_0^TG(Y_t,\Lambda_t)\mathrm{d}t}\Big]&\le\frac{1}{T}\int_0^T\mathbb{E}\Big[\mathrm{e}^{\beta TG(Y_t,\Lambda_t)}\Big]\mathrm{d}t\\
&\le\frac{C}{\Gamma(x)^{1/p}}\max_{i\in\mathcal{S}}\mu_0\Big(\mathrm{e}^{\beta TpG(\cdot,i)}\Big)^{1/p}\int_0^T\frac{1}{(1-\mathrm{e}^{-K_0t})^{d/p}}\mathrm{d}t,
\end{aligned}\tag{2.40}$$

where $C=C(p,T,K_0)$ is a constant and x is the initial value of (Y_t). Taking $d<p<\frac{\eta}{\beta T}$ in (2.40), it follows from the assumed condition in (ii) that

$$\mathbb{E}\Big[\mathrm{e}^{\beta\int_0^TG(Y_t,\Lambda_t)\mathrm{d}t}\Big]<\infty.$$

In order to establish (2.34), noticing $\xi>d$, we obtain from (2.39) that

$$\mathbb{E}[G(Y_t,\Lambda_t)]\le\frac{\mathrm{e}^{\frac{K_0}{\xi-\sqrt{\xi}}}\max_{i\in\mathcal{S}}\mu_0(G^\xi(\cdot,i))^{\frac{1}{\xi}}}{\Gamma(x)^{\frac{1}{\xi}}(1-\mathrm{e}^{-K_0t})^{\frac{d}{\xi}}},\quad t\in(0,T],\tag{2.41}$$

and hence

$$\mathbb{E}\Big[\int_0^TG(Y_t,\Lambda_t)\mathrm{d}t\Big]\le\frac{\mathrm{e}^{\frac{K_0}{\xi-\sqrt{\xi}}}}{\Gamma(x)^{\frac{1}{\xi}}}\Big(\int_0^T\frac{1}{(1-\mathrm{e}^{-K_0t})^{d/\xi}}\mathrm{d}t\Big)\max_{i\in\mathcal{S}}\mu_0(G^\xi(\cdot,i))^{\frac{1}{\xi}}<\infty.$$

The proof is complete. $\square$

Proof of Theorem 1.3 For every Markov chain (Λ_t) with transition rate matrix Q, there is a unique strong solution to SDE (1.14) under the conditions imposed in this theorem, which of course implies the weak uniqueness of the solution to SDE (1.14). Similarly, weak uniqueness holds for SDE (1.16). In this proof, let (Λ_t) be the Markov chain given by (2.1), and $(\tilde\Lambda_t)$ be given by (2.2). All the results established in beginning of this section still hold for this special construction of Markov chains. We shall this coupling process $(\Lambda_t, \tilde\Lambda_t)$ to estimate $\mathbb{E}\int_0^T \mathbf{1}_{\{\Lambda_s\neq\tilde\Lambda_s\}}\mathrm{d}s$ using Lemma 2.2 in the following argument.

By Lemma 2.4, Novikov's conditions (2.28) and (2.32) are verified under the assumption of this theorem. Therefore, $(X_t,\Lambda_t)_{t\in[0,T]}$ and $(\widetilde X_t,\tilde\Lambda_t)_{t\in[0,T]}$ can be represented in terms of $(Y_t,\Lambda_t)_{t\in[0,T]}$ and $(Y_t,\tilde\Lambda_t)_{t\in[0,T]}$. Denote the initial value of (Y_t) by x_0. It follows that for any measurable f with $\|f\|_{\mathrm{Lip}}+\|f\|_\infty\le 1$, and any $t\in[0,T]$,

$$\begin{aligned}|\mathbb{E}f(X_t)-\mathbb{E}f(\widetilde X_t)| &= \big|\mathbb{E}_{\mathbb{Q}}f(Y_t)-\mathbb{E}_{\widetilde{\mathbb{Q}}}f(Y_t)\big|\\ &=\Big|\mathbb{E}\Big[\Big(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}-\frac{\mathrm{d}\widetilde{\mathbb{Q}}}{\mathrm{d}\mathbb{P}}\Big)f(Y_t)\Big]\Big|\le \mathbb{E}\Big|\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}-\frac{\mathrm{d}\widetilde{\mathbb{Q}}}{\mathrm{d}\mathbb{P}}\Big|.\end{aligned}\tag{2.42}$$

Setting

$$M_t=\int_0^t\langle\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s),\mathrm{d}W_s\rangle,\quad \widetilde M_t=\int_0^t\langle\sigma^{-1}(Y_s)Z(Y_s,\tilde\Lambda_s),\mathrm{d}W_s\rangle,$$

and

$$\langle M\rangle_t=\int_0^t|\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s)|^2\mathrm{d}s,\quad \langle\widetilde M\rangle_t=\int_0^t|\sigma^{-1}(Y_s)Z(Y_s,\tilde\Lambda_s)|^2\mathrm{d}s$$

for $t\in[0,T]$, by the inequality $|\mathrm{e}^x-\mathrm{e}^y|\le(\mathrm{e}^x+\mathrm{e}^y)|x-y|$ for all x, $y\in\mathbb{R}$, we obtain that

$$\begin{aligned}&|\mathbb{E}f(X_t)-\mathbb{E}f(\widetilde X_t)|\\ &\le\mathbb{E}\Big[\Big(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}+\frac{\mathrm{d}\widetilde{\mathbb{Q}}}{\mathrm{d}\mathbb{P}}\Big)^p\Big]^{\frac1p}\mathbb{E}\big[|M_T-\widetilde M_T-\frac12\langle M\rangle_T+\frac12\langle\widetilde M\rangle_T|^q\big]^{\frac1q}\end{aligned}\tag{2.43}$$

for p, $q>1$ with $1/p+1/q=1$.

For the first term in (2.43), since $\eta>2Td$, we can choose $p=p_0>1$ such that $q_0=p_0/(p_0-1)>2$ and $2p_0^2Td<\eta$.

$$\mathbb{E}\Big[\Big(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big)^{p_0}\Big]=\mathbb{E}\big[\exp\big(p_0M_T-\frac{p_0}{2}\langle M\rangle_T\big)\big]$$

$$\le \mathbb{E}\big[\exp(2p_0M_T - 2p_0^2\langle M\rangle_T)\big]^{\frac{1}{2}}\mathbb{E}\big[\exp(p_0(2p_0-1)\langle M\rangle_T)\big]^{\frac{1}{2}}.$$

According to Lemma 2.1,

$$\mathbb{E}\Big[\mathrm{e}^{2p_0^2\langle M\rangle_T}\Big] < \infty, \quad \mathbb{E}\Big[\mathrm{e}^{p_0(2p_0-1)\langle M\rangle_T}\Big] < \infty.$$

Hence, $t \mapsto \exp\big(2p_0M_t - 2p_0^2\langle M\rangle_t\big)$ is an exponential martingale for $t \in [0, T]$ and

$$\mathbb{E}\Big[\Big(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big)^{p_0}\Big] \le \frac{C}{\Gamma(x_0)^{\frac{1}{p_1}}}\max_{i\in\mathcal{S}}\mu_0\Big(\mathrm{e}^{\eta|\sigma^{-1}(\cdot)Z(\cdot,i)|^2}\Big)^{\frac{1}{p_1}}\int_0^T\frac{1}{(1-\mathrm{e}^{-K_0t})^{\frac{d}{p_1}}}\mathrm{d}t < \infty, \tag{2.44}$$

where $p_1 > d$ satisfies $2p_0^2p_1T < \eta$, and $C = C(p_1, T, K_0)$.

We proceed to estimate the second term in (2.43). We shall estimate $\mathbb{E}[|M_T - \widetilde{M}_T|^{q_0}]$ and $\mathbb{E}[|\langle M\rangle_T - \langle\widetilde{M}\rangle_T|^{q_0}]$ separately. Since $q_0 > 2$, it follows from Burkholder-Davis-Gundy's inequality and Jensen's inequality that

$$\begin{aligned}
&\mathbb{E}[|M_T - \widetilde{M}_T|^{q_0}]\\
&\le C_{q_0}\mathbb{E}\Big[\Big(\int_0^T|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s))|^2\mathrm{d}s\Big)^{\frac{q_0}{2}}\Big]\\
&\le C_{q_0}T^{\frac{q_0}{2}-1}\mathbb{E}\Big[\int_0^T|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s))|^{q_0}\mathrm{d}s\Big]\\
&= C_{q_0}T^{\frac{q_0}{2}-1}\mathbb{E}\Big[\int_0^T|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s))|^{q_0}\mathbf{1}_{\{\Lambda_s\neq\tilde{\Lambda}_s\}}\mathrm{d}s\Big]\\
&\le C_{q_0}T^{\frac{q_0}{2}-1}\int_0^T\mathbb{E}\Big[|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s))|^{2q_0}\Big]^{\frac{1}{2}}\mathbb{P}\big(\Lambda_s\neq\tilde{\Lambda}_s\big)^{\frac{1}{2}}\mathrm{d}s\\
&\le C_{q_0}T^{\frac{q_0}{2}-1}\Big(\int_0^T\mathbb{E}\Big[|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s))|^{2q_0}\Big]\mathrm{d}s\Big)^{\frac{1}{2}}\Big(\int_0^T\mathbb{P}(\Lambda_s\neq\tilde{\Lambda}_s)\mathrm{d}s\Big)^{\frac{1}{2}}.
\end{aligned}$$

By (2.41) of Lemma 2.1,

$$\begin{aligned}
&\mathbb{E}\Big[\big|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s))\big|^{2q_0}\Big]\\
&\le 2^{2q_0-1}\Big(\mathbb{E}\Big[\big|\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s)\big|^{2q_0}\Big]+\mathbb{E}\Big[\big|\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s)\big|^{2q_0}\Big]\Big)\\
&\le 2^{2q_0}\mathrm{e}^{\frac{K_0}{\xi-\sqrt{\xi}}}\frac{\max_{i\in\mathcal{S}}\mu_0\Big(|\sigma^{-1}(\cdot)Z(\cdot,i)|^{2q_0\xi}\Big)^{\frac{1}{\xi}}}{\Gamma(x)(1-\mathrm{e}^{-K_0s})^{\frac{d}{\xi}}}, \quad \xi > d.
\end{aligned} \tag{2.45}$$

Note that the finiteness of $\max_{i\in\mathcal{S}}\mu_0\Big(|\sigma^{-1}(\cdot)Z(\cdot,i)|^{2q_0\xi}\Big)$ follows easily from the assumption

$$\max_{i\in\mathcal{S}}\mu_0\Big(\mathrm{e}^{\eta|\sigma^{-1}(\cdot)Z(\cdot,i)|^2}\Big)<\infty.$$

Therefore,

$$\begin{aligned}&\mathbb{E}[|M_T-\widetilde{M}_T|^{q_0}]\\&\le C\mathrm{e}^{\frac{K_0}{2\xi-2\sqrt{\xi}}}\Big(\int_0^T\frac{\max_{i\in\mathcal{S}}\mu_0\Big(|\sigma^{-1}(\cdot)Z(\cdot,i)|^{2q_0\xi}\Big)^{\frac{1}{\xi}}}{(1-\mathrm{e}^{-K_0s})^{\frac{d}{\xi}}}\mathrm{d}s\Big)^{\frac12}\Big(\int_0^T\mathbb{P}(\Lambda_s\neq\tilde{\Lambda}_s)\mathrm{d}s\Big)^{\frac12}\end{aligned}\tag{2.46}$$

for some constant $C=C(q_0,T)>0$ and $\xi>d$. By Lemma 2.2, we obtain that

$$\begin{aligned}&\mathbb{E}[|M_T-\widetilde{M}_T|^{q_0}]\\&\le C\mathrm{e}^{\frac{K_0}{2\xi-2\sqrt{\xi}}}\Big(\int_0^T\frac{\max_{i\in\mathcal{S}}\mu_0\Big(|\sigma^{-1}(\cdot)Z(\cdot,i)|^{2q_0\xi}\Big)^{\frac{1}{\xi}}}{(1-\mathrm{e}^{-K_0s})^{\frac{d}{\xi}}}\mathrm{d}s\Big)^{\frac12}\cdot NT\|Q-\widetilde{Q}\|_{\ell_1}^{\frac12}.\end{aligned}\tag{2.47}$$

In the following, we shall estimate $\mathbb{E}[|\langle M\rangle_T-\langle\widetilde{M}\rangle_T|^{q_0}]$.

$$\begin{aligned}&\mathbb{E}\Big[\big|\langle M\rangle_T-\langle\widetilde{M}\rangle_T\big|^{q_0}\Big]\\&\le\mathbb{E}\Big[\Big(\int_0^T|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s)|\big(|\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s)|+|\sigma^{-1}(Y_s)Z(Y_s,\tilde{\Lambda}_s|)\big)\mathrm{d}s\Big)^{q_0}\Big]\\&\le\mathbb{E}\Big[\Big(\int_0^T|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s)|^{\gamma}\mathrm{d}s\Big)^{\frac{q_0}{\gamma}}\\&\qquad\cdot\Big(\int_0^T\big(|\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s)|+|\sigma^{-1}(Y_s)Z(Y_s,\tilde{\Lambda}_s)|\big)^{\gamma'}\mathrm{d}s\Big)^{\frac{q_0}{\gamma'}}\Big]\\&\le\mathbb{E}\Big[\Big(\int_0^T|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s)|^{\gamma}\mathrm{d}s\Big)^{q_0}\Big]^{\frac{1}{\gamma}}\\&\qquad\cdot\mathbb{E}\Big[\Big(\int_0^T\big(|\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s)|+|\sigma^{-1}(Y_s)Z(Y_s,\tilde{Y}_s)|\big)^{\gamma'}\mathrm{d}s\Big)^{q_0}\Big]^{\frac{1}{\gamma'}},\end{aligned}$$

where γ, $\gamma'>1$ satisfy $1/\gamma+1/\gamma'=1$. By Lemma 2.1, it is easy to see

$$\mathbb{E}\Big[\Big(\int_0^T\big(|\sigma^{-1}(Y_s)Z(Y_s,\Lambda_s)|+|\sigma^{-1}(Y_s)Z(Y_s,\tilde{Y}_s)|\big)^{\gamma'}\mathrm{d}s\Big)^{q_0}\Big]^{\frac{1}{\gamma'}}<\infty.$$

On the other hand,

$$\begin{aligned}&\mathbb{E}\Big[\Big(\int_0^T|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s))\mathrm{d}s\Big)^{q_0}\Big]\\&\le T^{q_0-1}\Big(\int_0^T\mathbb{E}\big[|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s))|^{2\gamma q_0}\big]\mathrm{d}s\Big)^{\frac12}\Big(\int_0^T\mathbb{P}(\Lambda_s\neq\tilde{\Lambda}_s)\mathrm{d}s\Big)^{\frac12}.\end{aligned}$$

By virtue of (2.45) and Lemma 2.2, we get

$$\begin{aligned}&\mathbb{E}\Big[\Big(\int_0^T|\sigma^{-1}(Y_s)(Z(Y_s,\Lambda_s)-Z(Y_s,\tilde{\Lambda}_s))\mathrm{d}s\Big)^{q_0}\Big]^{\frac{1}{\gamma}}\\&\le C\mathrm{e}^{\frac{K_0}{2\gamma(\xi-\sqrt{\xi})}}\Big(\int_0^T\frac{\max_{i\in\mathcal{S}}\mu_0\Big(|\sigma^{-1}(\cdot)Z(\cdot,i)|^{2q_0\gamma\xi}\Big)^{\frac{1}{\xi}}}{(1-\mathrm{e}^{-K_0s})^{\frac{d}{\xi}}}\mathrm{d}s\Big)^{\frac{1}{2\gamma}}N^{\frac{1}{\gamma}}\|Q-\widetilde{Q}\|_{\ell_1}^{\frac{1}{2\gamma}},\end{aligned}\tag{2.48}$$

where $C=C(T,x_0,q_0)$ is a positive constant.

In all, inserting the estimates (2.44), (2.47) and (2.48) into (2.43), we arrive at

$$\big|\mathbb{E}f(X_t)-\mathbb{E}f(\widetilde{X}_t)\big|\le C\big(\|Q-\widetilde{Q}\|_{\ell_1}^{\frac{1}{2q_0}}\vee\|Q-\widetilde{Q}\|_{\ell_1}^{\frac{1}{2q_0\gamma}}\big)$$

for some constant C depending on $N,T,x_0,\tau_1,K_0,\xi,\gamma,p_0,\max_{i\in\mathcal{S}}\mu_0\big(\mathrm{e}^{\eta|\sigma^{-1}(\cdot)Z(\cdot,i)|^2}\big)$, and $\gamma>1$. By virtue of the definition of $W_{bL}(\cdot,\cdot)$,

$$W_{bL}(\mathcal{L}(X_t),\mathcal{L}(\widetilde{X}_t))\le C\big(\|Q-\widetilde{Q}\|_{\ell_1}^{\frac{1}{2q_0}}\vee\|Q-\widetilde{Q}\|_{\ell_1}^{\frac{1}{2q_0\gamma}}\big).$$

This completes the proof.

Proof of Theorem 1.4 This theorem can be proved along the same line as Thereom 1.3 by noting $\|Q-\widetilde{Q}\|_{\ell_1}\le\|B\|_{\ell_1}+\|Q_1+\widehat{Q}\|_{\ell_1}$. The details are omitted.

3 Further discussion

Recall the expression (2.16) of Q. The probabilistic meaning of q_0 is that the Markov chain (Λ_t) stays at the state "0" for a random period distributed as an exponential distribution with parameter q_0. So the larger the value of q_0 is, the shorter time period the process

(Λ_t) will stay at "0" in average. One may consider a limitation case that q_0 equals to $+\infty$, that is,

$$Q_\infty = \begin{pmatrix} -\infty & \alpha \\ \beta & Q_1 \end{pmatrix},$$

which means that the jump will occur immediately once the process (Λ_t) reaches the state "0". The state "0" in Q_∞ is called an instantaneous state. It seems also interesting to study the asymptotic behavior of Q to Q_∞ as q_0 tends to $+\infty$. Note that the continuous time Markov chain with instantaneous state produces new phenomenon compared with the Markov chains which are totally stable. For example, consider the well-known example provided by Kolmogorov [13]:

$$Q = \begin{pmatrix} -\infty & 1 & 1 & 1 & \dots \\ q_1 & -q_1 & 0 & 0 & \dots \\ q_2 & 0 & -q_2 & 0 & \dots \\ q_3 & 0 & 0 & -q_3 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

It was shown by Kendall and Reuter [12] that if

$$\sum_{j=1}^{\infty} q_j^{-1} < +\infty,$$

then there exists a Markov process with the generator Q. Notice that the state space of this Markov process is denumerable. Moreover, Chen and Reushaw [6] presented some sufficient conditions for the existence and uniqueness of continuous-time Markov chains with instantaneous states. According to [6, Corollary 3.2], Markov chains with a finite states have no instantaneous states. In the present work the state space $\mathcal{S}$ of Markov chain is finite, we have not consider that the Markov chain has the generator Q_∞, and hence the corresponding processes (Λ_t) and (X_t) have not been discussed. Therefore, to study the current problems for regime-switching processes with infinite state space $\mathcal{S}$ and instantaneous state is meaningful, and we leave it for further investigation.

Acknowledgement We are grateful to Professor Yonghua Mao for sharing his unpublished results, and valuable discussion on the works of [16, 17].

References

[1] J. Bao, J. Shao, Permanence and extinction of regime-switching predator-prey models, SIAM J. Math. Anal., 48 (2016), 725-739.

[2] J. Bardet, H. Guerin, F. Malrieu, Long time behavior of diffusions with Markov switching, ALEA Lat. Am. J. Probab. Math. Stat., 7 (2010), 151-170.

[3] G. Basak, A. Bisi, M. Ghosh, Stability of a random diffusion with linear drift, J. Math. Anal. Appl., 202 (1996), 604-622.

[4] G. Basak, A. Bisi, M. Ghosh, Stability of a degenerate diffusions with state-dependent switching, J. Math. Anal. Appl., 240 (1999), 219-248.

[5] S. Brown, P. Dybvig, The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates, Journal of Finance, 41 (1986), 617-630.

[6] A. Chen, E. Renshaw, Existence and uniqueness criteria for conservative uni-instantaneous denumerable Markov processes, Probab. Theory Relat. Fields, 94 (1993), 427-456.

[7] M.-F. Chen, From Markov chains to non-equilibrium particle systems, 2nd ed. Singapore: World Scientific, 2004.

[8] B. Cloez, M. Hairer, Exponential ergodicity for Markov processes with random switching, Bernoulli, 21 (2015), 505-536.

[9] de Saporta, J. Yao, Tail of a linear diffusion with Markov switching, Ann. Appl. Probab., 15 (2005), (1B), 992-1018.

[10] M. Ghosh, A. Arapostathis, S. Marcus, Optimal control of switching diffusions with application to flexible manufacturing systems, SIAM J. Contr. Optim., 30 (1992), 1-23.

[11] T. Hou, J. Shao, Heavy tail and light tail of Cox-Ingersoll-Ross processes with regime-switching, to appear in Sci. China Math. 2019 or arXiv:1709.01691.

[12] D. Kendall, G. Reuter, Some pathological Markov processes with a denumerable infinity of states and the associated semigroups of operators on ℓ, Proc. Intern. Congr. Math. Amsterdam, Vol. III, 377-415. Amserdam: North-Holland 1954.

[13] A. Kolmogorov, On the differentiability of the transition probabilities in homogeneous Markov processes with a denumerable number of states, Moskov. Gos. Univ. Ucenye Zapiski MGY 148 Mat. 4 (1951), 53-59.

[14] X. Mao, Stability of stochastic differential equations with Markovian switching, S-toch. Process. Appl., 79 (1999), 45-67.

[15] X. Mao, C. Yuan, Stochastic Differential Equations with Markovian Switching, Imperial College Press, London, 2006.

[16] A. Mitrophanov, Stability and exponential convergence of continuous-time Markov chains, J. Appl. Prob., 40 (2003), 970-979.

[17] A. Mitrophanov, The spectral gap and perturbation bounds for reversible continuous-time Markov chains, J. Appl. Prob., 41 (2004), 1219-1222.

[18] M. Pinsky, R. Pinsky, Transience recurrence and central limit theorem behavior for diffusions in random temporal environments, Ann. Probab., 21 (1993), 433–452.

[19] J. Shao, Ergodicity of one-dimensional regime-switching diffusion processes, Science China Math., 57 (2014), 2407-2414.

[20] J. Shao, Criteria for transience and recurrence of regime-switching diffusion processes, Electron. J. Probab., 20 (2015), 1-15.

[21] J. Shao, Strong solutions and strong Feller properties for regime-switching diffusion processes in an infinite state space, SIAM J. Control Optim., 53 (2015), 2462-2479.

[22] J. Shao, Invariant measures and Euler-Maruyama's approximations of state-dependent regime-switching diffusions. SIAM J. Control Optim., 56 (2018), no. 5, 3215-3238.

[23] J. Shao, F. Xi, Stability and recurrence of regime-switching diffusion processes, SIAM J. Control Optim., 52 (2014), 3496-3516.

[24] A. Skorokhod, Asymptotic Methods in the Theory of Stochastic Differential Equations, American Mathematical Society, Providence, RI. 1989.

[25] C. Villani, Topics in optimal transportation, American Mathematical Society, Providence, RI, 2003.

[26] C. Villani, Optimal transport, old and new, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer Berlin Heidelberg, 2009.

[27] F.Y. Wang, Harnack inequality for SDE with multiplicative noise and extension to Neumann semigroup on nonconvex manifolds, Ann. Probab., 39 (2011), 1149-1467.

[28] F.Y. Wang, Integrability conditions for SDEs and semilinear SPDEs, Ann. Probab., 45 (2017), 3223-3265.

[29] F. Xi, G. Yin, Stability of regime-switching jump diffusions, SIAM J. Control Optim., 48 (2010), 525-4549.

[30] G. Yin, C. Zhu, Hybrid switching diffusions: properties and applications, Vol. 63, Stochastic Modeling and Applied Probability, Springer, New York. 2010.

[31] A.I. Zeifman, D.L. Isaacson, On strong ergodicity for nonhomogeneous continuous-time Markov chains, Stochastic Process. Appl., 50 (1994), 263-273.

[32] X.C. Zhang, Stochastic differential equations with Sobolev diffusion and singular drift and applications, Annals Appl. Probab., 26 (2016), 2697-2732.