Donsker-Varadhan Large Deviations for Path-Distribution Dependent SPDEs *

Panpan Ren[b)], **Feng-Yu Wang**[a)]

[a)] Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

[b)] Department of Mathematics, Swansea University, Bay Campus, SA1 8EN, United Kingdom

673788@swansea.ac.uk, wangfy@tju.edu.cn

November 24, 2021

Abstract

As an important tool characterizing the long time behavior of Markov processes, the Donsker-Varadhan LDP (large deviation principle) does not directly apply to distribution dependent SDEs/SPDEs since the solutions are not standard Markovian. We establish this type LDP for several different models of distribution dependent SDEs/SPDEs which may also with memories, by comparing the original equations with the corresponding distribution independent ones. As preparations, the existence, uniqueness and exponential convergence are also investigated for path-distribution dependent SPDEs which should be interesting by themselves.

AMS subject Classification: 60B05, 60B10.
Keywords: Donsker-Varadhan LDP, path-distribution dependent SDEs, Warsserstein distance.

1 Introduction

The LDP (large deviation principle) is a fundamental tool characterizing the asymptotic behaviour of probability measures $\{\mu_\varepsilon\}_{\varepsilon>0}$ on a topological space E, see [5] and references within. Recall that μ_ε for small $\varepsilon>0$ is said to satisfy the LDP with speed $\lambda(\varepsilon)\to+\infty$ (as $\varepsilon\to 0$) and rate function $I:E\to[0,+\infty]$, if I has compact level sets (i.e. $\{I\le r\}$ is compact for $r\in\mathbb{R}^+$), and for any Borel subset A of E,

$$-\inf_{A^o} I\le \liminf_{\varepsilon\to 0}\frac{1}{\lambda(\varepsilon)}\log\mu_\varepsilon(A)\le\limsup_{\varepsilon\to 0}\frac{1}{\lambda(\varepsilon)}\log\mu_\varepsilon(A)\le -\inf_{\bar A} I,$$

where A^o and $\bar A$ stand for the interior and the closure of A in E respectively. The following two different type LDPs have been studied in the literature.

*Supported in part by NNSFC (11771326, 11831014, 11921001).

The Freidlin-Wentzell type small noise LDP [7]: μ_ε stands for the distribution of the solution to a dynamic system perturbed by a noise with small intensity $\varepsilon > 0$, i.e. SDE (stochastic differential equation) with small noise. In this case, E is the path space for the solutions of the SDE. This type LDP describes, as $\varepsilon \to 0$, the convergence of stochastic systems to the corresponding deterministic system.

The Donsker-Varadhan type long time LDP [6]: μ_ε stands for the distribution of $L_{\varepsilon^{-1}}$, where

$$L_t := \frac{1}{t}\int_0^t \delta_{X(s)}\mathrm{d}s, \quad t > 0$$

is the empirical measure for a stochastic process $\{X(t)\}_{t\geq 0}$. This type LDP describes the behaviour of L_t as $t \to \infty$. In this case, E is the set of all probability measures on the state space of the process, on which both the weak topology (induced by bounded continuous functions) and the τ-topology (induced by bounded measurable functions) are considered in the literature.

In this paper, we study the Donsker-Varadhan LDP for path-distribution dependent SDEs (stochastic differential equations) on a separable Hilbert space $\mathbb{H}$. Inspired by Kac's programme for Vlasov systems in kinetic theory [11], McKean [14] introduced distribution dependent SDEs. According to Sznitman [17], under the global Lipschtiz condition, these type SDEs can be derived as the limit of mean-field particle systems when the number of particles tends to infinity. Therefore, distribution dependent SDEs are also called Mckean-Vlasov SDEs and mean-field SDEs.

In applications, the distribution of a stochastic process can be regarded as a macro property, while the path of the process up to a time t stands for the history of the system before this time. Since the evolution of a stochastic system may depend on both the macro environment and the history, it is reasonable to investigate path-distribution dependent SDEs. Moreover, because in many cases the configuration space for particle systems is infinite-dimensional, we consider path-distribution dependent SDEs on Hilbert spaces, and in this case the SDEs are called SPDEs (stochastic partial differential equations).

In recent years, distribution dependent SDEs have been intensively investigated. Among many other papers in this field, [15] established the Freidlin-Wentzell LDP for distribution dependent SDEs. However, up to our best knowledge, there is no any result on the Donsker-Varadhan LDP for this type SDEs. Since the solution is not standard Markovian, existing results on the Donsker-Varadhan LDP derived for Markov processes do not apply. Indeed, the definition of the rate function (the Donsker-Varadhan level 2 entropy function) depends on the standard Markov property of the process, for which the law of the process starting at an initial distribution ν is given by

$$P^\nu = \int_E P^x \nu(\mathrm{d}x),$$

where P^x is the law of the process starting at x, see Subsection 3.2 for details.

The framework. We investigate path-distribution dependent SDEs on a separable Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle, |\cdot|)$. For a fixed constant $r_0 > 0$, a path $\xi \in \mathscr{C} := C([-r_0, 0]; \mathbb{H})$ stands for a sample of the history with time length r_0. Recall that $\mathscr{C}$ is a Banach space with the uniform norm

$$\|\xi\|_\infty := \sup_{\theta\in[-r_0,0]} |\xi(\theta)|, \quad \xi \in \mathscr{C}.$$

For any map $\xi(\cdot) : [-r_0, \infty) \to \mathbb{H}$ and any time $t \geq 0$, its segment $\xi_\cdot : [0, \infty) \to \mathscr{C}$ is defined by

$$\xi_t(\theta) := \xi(t+\theta), \quad \theta \in [-r_0, 0], t \geq 0.$$

Let $\mathscr{P}(\mathscr{C})$ denote the space of all probability measures on $\mathscr{C}$ equipped with the weak topology, and let $\mathscr{L}_\eta$ stand for the distribution of a random variable η. Consider the following path-distribution dependent SPDE on $\mathbb{H}$:

$$\mathrm{d}X(t) = \{AX(t) + b(X_t, \mathscr{L}_{X_t})\}\mathrm{d}t + \sigma(\mathscr{L}_{X_t})\mathrm{d}W(t), \quad t \geq 0, \tag{1.1}$$

where

- $(A, \mathscr{D}(A))$ is a negative definite self-adjoint operator on $\mathbb{H}$;
- $W(t)$ is the cylindrical Brownian motion on a separable Hilbert space $\tilde{\mathbb{H}}$; i.e.

$$W(t) = \sum_{i=1}^\infty B_i(t)\tilde{e}_i, \quad t \geq 0$$

  for an orthonormal basis $\{\tilde{e}_i\}_{i\geq1}$ on $\tilde{\mathbb{H}}$ and a sequence of independent one-dimensional Brownian motions $\{B_i\}_{i\geq1}$ on a complete filtration probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq0}, \mathbb{P})$, where $\mathscr{F}_0$ is rich enough such that for any $\pi \in \mathscr{P}(\mathscr{C} \times \mathscr{C})$ there exists a $\mathscr{C} \times \mathscr{C}$-valued random variable ξ on $(\Omega, \mathscr{F}_0, \mathbb{P})$ such that $\mathscr{L}_\xi = \pi$.
- $b : \mathscr{C} \times \mathscr{P}(\mathscr{C}) \to \mathbb{H}, \quad \sigma : \mathscr{P}(\mathscr{C}) \to \mathbb{L}(\tilde{\mathbb{H}}; \mathbb{H})$ are measurable.

In Section 3, the well-posedness of a more general equation (3.1) will be presented. However, to establish the Donsker-Varadhan LDP using the comparing method proposed in Theorem 5.5 below, we have to assume that the noise term only depends on the distribution $\mathscr{L}_{X_t}$ rather than the solution X_t.

Let X_t^ν denote the mild segment solution of (1.1) with initial distribution $\nu \in \mathscr{P}(\mathscr{C})$, which is a continuous adapted process on $\mathscr{C}$, see Definition 3.1 below for details. We study the long time LDP for the empirical measure

$$L_t^\nu := \frac{1}{t}\int_0^t \delta_{X_s^\nu}\mathrm{d}s, \quad t > 0.$$

Definition 1.1. Let $\mathscr{P}(\mathscr{C})$ be equipped with the weak topology, let $\mathscr{A} \subset \mathscr{P}(\mathscr{C})$, and let $J : \mathscr{P}(\mathscr{C}) \to [0, \infty]$ have compact level sets, i.e. $\{J \leq r\}$ is compact in $\mathscr{P}(\mathscr{C})$ for any $r > 0$.

(1) $\{L_t^\nu\}_{\nu\in\mathscr{A}}$ is said to satisfy the upper bound uniform LDP with rate function J, denoted by $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP_u(J)$, if for any closed $A \subset \mathscr{P}(\mathscr{C})$,

$$\limsup_{t\to\infty} \frac{1}{t} \sup_{\nu\in\mathscr{A}} \log \mathbb{P}(L_t^\nu \in A) \leq -\inf_A J.$$

(2) $\{L_t^\nu\}_{\nu\in\mathscr{A}}$ is said to satisfy the lower bound uniform LDP with rate function J, denoted by $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP_l(J)$, if for any open $A \subset \mathscr{P}(\mathscr{C})$,

$$\liminf_{t\to\infty} \frac{1}{t} \inf_{\nu\in\mathscr{A}} \log \mathbb{P}(L_t^\nu \in A) \geq -\inf_A J.$$

(3) $\{L_t^\nu\}_{\nu\in\mathscr{A}}$ is said to satisfy the uniform LDP with rate function J, denoted by $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP(J)$, if $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP_u(J)$ and $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP_l(J)$.

Main idea of the study. To establish the Donsker-Varadhan type LDP for a distribution dependent SDE/SPDE, we choose a reference SDE/SPDE whose solution is Markovian so that existing results on the Donsker-Varadhan LDP apply. When (1.1) is well-posed, let $P_t^*\mu = \mathscr{L}_{X_t}$ for $\mathscr{L}_{X_0} = \mu \in \mathscr{P}(\mathscr{C})$. If P_t^* has a unique invariant probability measure $\bar{\mu}$, we choose the following stationary equation as the reference SPDE

$$\mathrm{d}X(t) = \{AX(t) + b(X_t, \bar{\mu})\}\mathrm{d}t + \sigma(\bar{\mu})\mathrm{d}W(t), \quad t \geq 0. \tag{1.2}$$

The solution to this equation is a standard Markov process and hence its LDP can be illustrated using existing results. By comparing the original equation (1.1) with (1.2) in the sense of LDP, see Lemma 5.4 below, we establish the desired Donsker-Varadhan LDP.

Since the theory of LDP has already been well developed for Markov processes (see Wu [22, 23]), the main point of the present study is to verify assumptions in Wu's results and the comparison theorem (see Theorem 5.5), rather than to make development in methodology. However, it is non-trivial to check these assumptions for the present path-distribution dependent model. Indeed, when $r_0 > 0$ and σ is non-constant, we are not able to prove the exponential estimate in Theorem 5.5, so that in this case the Donsker-Varadhan LDP for (1.1) remains open. In conclusion, new aspects presented in the paper include:

(a) The well-posedness for path-distribution dependent SPDEs (see Theorem 3.1);

(b) The exponential ergodicity of (1.1), and the hypercontractivity, irreducibility and strong Feller property for the stationary equation (1.2);

(c) The exponential estimate (5.7) for solutions to the original equation (1.1) and the reference equation (1.2).

To derive the hypercontractivity, irreducibility and strong Feller property, a crucial tool is the dimension-free Harnack inequality due to the second named author.

The remainder of the paper is orgnized as follows. In Section 2 we state the main results of the paper and illustrate them by specific examples; in Section 3 we investigate the existence and uniqueness for path-distribution dependent SDEs/SPDEs; and in Section 4 we prove the main results. Finally, for readers convenience, we recall in Appendix some results on LDP due to [22, 23], and present an exponential estimate for the equivalence of LDPs, which are applied in our proofs.

2 Main results and Examples

Path-Distribution dependent SDE is a probability model characterizing nonlinear Fokker-Planck equations on the path space, see [10] for details. A fundamental problem in ergodic theory is to characterize the stationary distribution of a stochastic system (i.e. invariant solution of the associated Fokker-Planck equation). Although the existence and uniqueness of the stationary distribution can be confirmed by standard techniques (for instance, coupling method), the exact formulation of this distribution is however unknown. In applications, one uses empirical measures to simulate the stationary distribution, this leads to the study of limit theory on empirical measures of Markov processes, for instance the law of large numbers, central limit theorems and large deviations, see [12]. In the path dependent case, limit theorems have been studied in [2]. In this section, we state our results on Donsker-Varadhan LDP for path-distribution dependent SPDEs and present concrete examples to illustrate them.

To state our main results, we recall some notations. For a Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$, let $\mathscr{P}(\mathbb{B})$ be the space of all probability measures on $\mathbb{B}$ equipped with the weak topology. For any constant $p > 0$, let

$$\mathscr{P}_p(\mathbb{B}) := \big\{\mu \in \mathscr{P}(\mathbb{B}) : \|\mu\|_p := \mu(\|\cdot\|_{\mathbb{B}}^p)^{\frac{1}{p\vee 1}} < \infty\big\},$$

which is a Polish space under the metric

$$\mathbf{W}_p(\mu_1, \mu_2) := \inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)} \bigg(\int_{\mathbb{B}\times\mathbb{B}} \|x-y\|_{\mathbb{B}}^p \pi(\mathrm{d}\xi, \mathrm{d}\eta)\bigg)^{\frac{1}{p}},$$

where $\mathscr{C}(\mu_1, \mu_2)$ is the set of all couplings for μ_1 and μ_2. Next, let $\mathscr{B}_b(\mathscr{B})$ (resp. $C_b(\mathscr{B})$) be the space of bounded measurable (resp. continuous) real functions on $\mathscr{B}$. A sub-Markov operator P on $\mathscr{B}_b(\mathscr{C})$ is called Feller if $PC_b(\mathscr{C}) \subset C_b(\mathscr{C})$, strong Feller if $P\mathscr{B}_b(\mathscr{C}) \subset C_b(\mathscr{C})$, and μ-irreducible for some $\mu \in \mathscr{P}(\mathscr{B})$ if $\mu(1_A P 1_B) > 0$ holds for any $A, B \in \mathscr{B}(\mathscr{B})$ with $\mu(A)\mu(B) > 0$.

2.1 Distribution dependent SDE on $\mathbb{R}^d$

Let $r_0 = 0$, $\mathbb{H} = \mathbb{R}^d$ and $\tilde{\mathbb{H}} = \mathbb{R}^m$ for some $d, m \in \mathbb{N}$. In this case, we combine the linear term Ax with the drift term $b(x, \mu)$, so that (1.1) reduces to

$$\mathrm{d}X(t) = b(X(t), \mathscr{L}_{X(t)})\mathrm{d}t + \sigma(\mathscr{L}_{X(t)})\mathrm{d}W(t), \tag{2.1}$$

where $b : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^m$ and $W(t)$ is the m-dimensional Brownian motion. We assume

(H_1) b is continuous, σ is bounded and continuous such that

$$2\langle b(x,\mu) - b(y,\nu), x-y\rangle + \|\sigma(\mu) - \sigma(\nu)\|_{HS}^2 \le -\kappa_1|x-y|^2 + \kappa_2 \mathbf{W}_2(\mu,\nu)^2$$

holds for some constants $\kappa_1 > \kappa_2 \ge 0$ and all $x, y \in \mathbb{R}^d$, $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$.

We will prove that under (H_1) the equation (2.1) is well-posed and has a unique invariant probability measure $\bar\mu \in \mathscr{P}_2(\mathbb{R}^d)$ such that

$$\mathbf{W}_2(P_t^*\nu, \bar\mu)^2 \le \mathrm{e}^{-(\kappa_1-\kappa_2)t}\mathbf{W}_2(\nu,\bar\mu)^2, \quad t\ge 0, \nu\in\mathscr{P}_2(\mathbb{R}^d), \tag{2.2}$$

where $P_t^*\mu = \mathscr{L}_{X(t)}$ for initial distribution $\mathscr{L}_{X(0)} = \mu$. Consider the reference SDE

$$\mathrm{d}\bar X(t) = b(\bar X(t), \bar\mu)\mathrm{d}t + \sigma(\bar\mu)\mathrm{d}W(t). \tag{2.3}$$

It is standard that under (H_1) the equation (2.3) has a unique solution $\bar X^x(t)$ for any starting point $x\in\mathbb{R}^d$, and $\bar\mu$ is the unique invariant probability measure of the associated Markov semigroup

$$\bar P_t f(x) := \mathbb{E}[f(\bar X^x(t))], \quad t\ge 0, x\in\mathbb{R}^d, f\in\mathscr{B}_b(\mathbb{R}^d).$$

Consequently, $\bar P_t$ uniquely extends to $L^\infty(\bar\mu)$. If $f\in L^\infty(\bar\mu)$ satisfies

$$\bar P_t f = f + \int_0^t \bar P_s g\mathrm{d}s, \quad \bar\mu\text{-a.e.}$$

for some $g\in L^\infty(\bar\mu)$ and all $t\ge 0$, we write $f\in\mathscr{D}(\bar{\mathscr{A}})$ and denote $\bar{\mathscr{A}}f = g$. Then $\mathscr{D}(\bar{\mathscr{A}})\supset C_c^\infty(\mathbb{R}^d) := \{f\in C_b^\infty(\mathbb{R}^d) : \nabla f \text{ has compact support}\}$ and

$$\bar{\mathscr{A}}f(x) = \frac{1}{2}\sum_{i,j=1}^d \{\sigma\sigma^*\}_{ij}(\bar\mu)\partial_i\partial_j f(x) + \sum_{i=1}^\infty b_i(x,\bar\mu)\partial_i f(x), \quad f\in C_c^\infty(\mathbb{R}^d).$$

According to Section 3, the Donsker-Varadhan level 2 entropy function J for the diffusion process generated by $\bar{\mathscr{A}}$ has compact level sets in $\mathscr{P}(\mathbb{R}^d)$ under the τ and weak topologies, and by (5.2) below we have

$$J(\nu) = \begin{cases} \sup\big\{ \int_{\mathbb{R}^d} \frac{-\bar{\mathscr{A}}f}{f}\mathrm{d}\nu : \ 1\le f\in\mathscr{D}(\bar{\mathscr{A}})\big\}, & \text{if } \nu\lambda\mu,\\ \infty, & \text{otherwise.}\end{cases}$$

Theorem 2.1. *Assume (H_1). For any $r, R>0$, let $\mathscr{B}_{r,R} = \big\{\nu\in\mathscr{P}(\mathbb{R}^d) : \nu(\mathrm{e}^{|\cdot|^r}) \le R\big\}$.*

1. *(2.1) is well-posed for initial distributions in $\mathscr{P}_2(\mathbb{R}^d)$ and P_t^* has a unique invariant probability measure $\bar\mu\in\mathscr{P}_2(\mathbb{R}^d)$ such that (2.2) holds.*
2. *We have $\{L_t^\nu\}_{\nu\in\mathscr{B}_{r,R}} \in LDP_u(J)$ for all $r,R>0$. If $\bar P_t$ is strong Feller and $\bar\mu$-irreducible for some $t>0$, then $\{L_t^\nu\}_{\nu\in\mathscr{B}_{r,R}}\in LDP(J)$ for all $r,R>0$.*
3. *If there exist constants $\varepsilon, c_1, c_2>0$ such that*

$$\langle x, b(x,\nu)\rangle \le c_1 - c_2|x|^{2+\varepsilon}, \quad x\in\mathbb{R}^d, \nu\in\mathscr{P}_2(\mathbb{R}^d), \tag{2.4}$$

then $\{L_t^\nu\}_{\nu\in\mathscr{P}_2(\mathbb{R}^d)}\in LDP_u(J)$. If moreover $\bar P_t$ is strong Feller and $\bar\mu$-irreducible for some $t>0$, then $\{L_t^\nu\}_{\nu\in\mathscr{P}_2(\mathbb{R}^d)}\in LDP(J)$.

To apply this result, we first recall some facts on the strong Feller property and the irreducibility of diffusion semigroups.

Remark 2.1. We make some comments on the strong Feller property and the $\bar\mu$-irreducibility used in the above theorem for hypo-elliptic or elliptic diffusion processes.

(1) Let $\bar P_t$ be the (sub-)Markov semigroup generated by the second order differential operator

$$\bar{\mathscr{A}} := \sum_{i=1}^m U_i^2 + U_0,$$

where $\{U_i\}_{i=1}^m$ are C^1-vector fields and U_0 is a continuous vector field. Here, as a convention in differential geometry, each vector field U is regarded as a first order differential operator by letting $Uf := \langle \nabla f, U\rangle$ for a differentiable function f. Assume that $\bar{\mathscr{A}}$ satisfies the Hörmander condition, i.e. there exists a natural number $k \geq 1$ such that at each point we have $\mathrm{span}\mathscr{H}_k = \mathbb{R}^d$, where $\mathscr{H}_1 := \{U_i : 1 \leq i \leq m\}$ and for each $n \geq 1$ we set

$$\mathscr{H}_{n+1} := \mathscr{H}_n \cup \big\{[U, U_i] := UU_i - U_iU,\ 0 \leq i \leq m\big\}.$$

Then according to [13, Theorem 5.1], $\bar P_t$ satisfies the Harnack inequality

$$\bar P_t f(x) \leq \psi(t,s,x,y)\bar P_{t+s} f(y), \quad t,s>0, x,y\in\mathbb{R}^d, f\in\mathscr{B}^+(\mathbb{R}^d)$$

for some map $\psi : (0,\infty)^2 \times (\mathbb{R}^d)^2 \to (0,\infty)$. Consequently, if $\bar P_t$ has an invariant probability measure $\bar\mu$, then $\bar P_t$ is $\bar\mu$-irreducible for any $t > 0$. Finally, if $\{U_i\}_{0\leq i\leq m}$ are smooth with bounded derivatives of all orders, then the above Hörmander condition implies that $\bar P_t$ has smooth heat kernel with respect to the Lebesgue measure, in particular it is strong Feller for any $t > 0$.

(2) Let $\bar P_t$ be the Markov semigroup generated by

$$\bar{\mathscr{A}} := \sum_{i,j=1}^d \bar a_{ij}\partial_i\partial_j + \sum_{i=1}^d \bar b_i \partial_j,$$

where $(\bar a_{ij}(x))$ is strictly positive definite for any x, $\bar a_{ij} \in H^{p,1}_{loc}(\mathrm{d}x)$ and $\bar b_i \in L^p_{loc}(\mathrm{d}x)$ for some $p > d$ and all $1 \leq i,j \leq d$. Moreover, let $\bar\mu$ be an invariant probability measure of $\bar P_t$. Then by [3, Theorem 4.1], $\bar P_t$ is strong Feller for all $t > 0$. Moreover, as indicated in (1) that [13, Theorem 5.1] ensures the $\bar\mu$-irreducibility of $\bar P_t$ for $t > 0$.

We present below two examples to illustrate this result, where the first is a distribution dependent perturbation of the Ornstein-Ulenbeck process, and the second is the distribution dependent stochastic Hamiltonian system.

Example 2.1. Let $\sigma(\nu) = I + \varepsilon\sigma_0(\nu)$ and $b(x,\nu) = -x + \varepsilon b_0(x,\nu)$, where I is the identity matrix, $\varepsilon > 0$ and $\sigma_0 : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^d$, $b_0 : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ are Lipschitz continuous. When $\varepsilon > 0$ is small enough, assumption (H_1) holds and that $\bar P_t$ satisfies conditions in Remark 2.1(2). So, Theorem 2.1(1) implies $\{L_t^\nu\}_{\nu\in\mathscr{B}_{r,R}} \in LDP(J)$ for all $r, R > 0$.

If we take the above σ but $b(x,\nu) = -x - c|x|^\theta x + \varepsilon b_0(x,\nu)$ for some constants $c, \theta > 0$ and a Lipschitz continuous map $b_0 : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ such that $|b_0(x,\nu)| \leq c_0(1 + |x|)$ holds for some constant $c_0 > 0$, then when $\varepsilon > 0$ is small enough (H_1) and (2.4) are satisfied, so that Theorem 2.1(2) and Remark 2.1(2) imply $\{L_t^\nu\}_{\nu\in\mathscr{P}_2(\mathbb{R}^d)} \in LDP(J)$.

Example 2.2. Let $d = 2m$ and consider the following distribution dependent SDE for $X(t) = (X^{(1)}(t), X^{(2)}(t))$ on $\mathbb{R}^m \times \mathbb{R}^m$:

$$\begin{cases} \mathrm{d}X^{(1)}(t) = \{X^{(2)}(t) - \lambda X^{(1)}(t)\}\mathrm{d}t \\ \mathrm{d}X^{(2)}(t) = \{Z(X(t), \mathscr{L}_{X(t)}) - \lambda X^{(2)}(t)\}\mathrm{d}t + \sigma \mathrm{d}W(t), \end{cases},$$

were $\lambda > 0$ is a constant, σ is an invertible $m \times m$-matrix, $W(t)$ is the m-dimensional Brownian motion, and $Z : \mathbb{R}^{2m} \times \mathscr{P}_2(\mathbb{R}^{2m}) \to \mathbb{R}^m$ satisfies

$$|Z(x_1, \nu_1) - Z(x_2, \nu_2)| \le \alpha_1 |x_1^{(1)} - x_2^{(1)}| + \alpha_2 |x_1^{(2)} - x_2^{(2)}| + \alpha_3 \mathbf{W}_2(\nu_1, \nu_2)$$

for some constants $\alpha_1, \alpha_2, \alpha_3 \ge 0$ and all $x_i = (x_i^{(1)}, x_i^{(2)}) \in \mathbb{R}^{2m}, \nu_i \in \mathscr{P}_2(\mathbb{R}^{2m}), 1 \le i, j \le 2$. If

$$4\lambda > \inf_{s>0} \big\{2\alpha_3 s + \alpha_3 s^{-1} + 2\alpha_2 + \sqrt{4(1+\alpha_1)^2 + (2\alpha_2 + \alpha_3 s^{-1})^2}\big\}, \tag{2.5}$$

then $\{L_t^\nu\}_{\nu \in \mathscr{B}_{r,R}} \in LDP(J)$ for all $r, R > 0$.

Indeed, $b(x, \nu) := (x^{(2)} - \lambda x^{(1)}, Z(x, \nu) - \lambda x^{(2)})$ satisfies

$$\begin{aligned} &2\langle b(x_1, \nu_1) - b(x_2, \nu_2), x_1 - x_2 \rangle \\ &\le -2\lambda |x_1^{(1)} - x_2^{(1)}|^2 - 2(\lambda - \alpha_2)|x_1^{(2)} - x_2^{(2)}|^2 \\ &\quad + 2|x_1^{(2)} - x_2^{(2)}| \big\{(1+\alpha_1)|x_1^{(1)} - x_2^{(1)}| + \alpha_3 \mathbf{W}_2(\nu_1, \nu_2)\big\} \\ &\le \alpha_3 s \mathbf{W}_2(\nu_1, \nu_2)^2 - \{2\lambda - \delta(1+\alpha_1)\}|x_1^{(1)} - x_2^{(1)}|^2 \\ &\quad - \{2\lambda - 2\alpha_2 - \delta^{-1}(1+\alpha_1) - \alpha_3 s^{-1}\}|x_1^{(2)} - x_2^{(2)}|^2, \quad s, \delta > 0 \end{aligned}$$

for all $x_1, x_2 \in \mathbb{R}^{2m}$ and $\nu_1, \nu_2 \in \mathscr{P}_2(\mathbb{R}^{2m})$. Taking

$$\delta = \frac{2\alpha_2 + \alpha_3 s^{-1} + \sqrt{4(1+\alpha_1)^2 + (2\alpha_2 + \alpha_3 s^{-1})^2}}{2(1+\alpha_1)}$$

such that $\delta(1+\alpha_1) = 2\alpha_2 + \delta^{-1}(1+\alpha_1) + \alpha_3 s^{-1}$, we see that (H_1) holds for some $\kappa_1 > \kappa_2$ provided $2\lambda - \delta(1+\alpha_1) > \alpha_3 s$ for some $s > 0$, i.e. (2.5) implies (H_1). Moreover, it is easy to see that conditions in Remark 2.1(1) hold, see also [8, 21] for Harnack inequalities and gradeint estimates on stochastic Hamiltonian systems which also imply the strong Feller and $\bar{\mu}$-irreducibility of $\bar{P}_t$. Therefore, the claimed assertion follows from Theorem 2.1(1).

2.2 Distribution dependent SPDE

Consider the following distribution-dependent SPDE on a separable Hilbert space $\mathbb{H}$:

$$\mathrm{d}X(t) = \{AX(t) + b(X(t), \mathscr{L}_{X(t)})\}\mathrm{d}t + \sigma(\mathscr{L}_{X(t)})\mathrm{d}W(t), \tag{2.6}$$

where $(A, \mathscr{D}(A))$ is a linear operator on $\mathbb{H}$, $b : \mathbb{H} \times \mathscr{P}_2(\mathbb{H}) \to \mathbb{H}$ and $\sigma : \mathscr{P}_2(\mathbb{H}) \to \mathbb{L}(\tilde{\mathbb{H}}; \mathbb{H})$ are measurable, and $W(t)$ is the cylindrical Brwonian motion on $\tilde{\mathbb{H}}$. We make the following assumption.

(H_2) $(-A, \mathscr{D}(A))$ is self-adjoint with discrete spectrum $0 < \lambda_1 \le \lambda_2 \le \cdots$ counting multiplicities such that $\sum_{i=1}^\infty \lambda_i^{\gamma-1} < \infty$ holds for some constant $\gamma \in (0,1)$.

Moreover, b is Lipschitz continuous on $\mathbb{H} \times \mathscr{P}_2(\mathbb{H})$, σ is bounded and there exist constants $\alpha_1, \alpha_2 \ge 0$ with $\lambda_1 > \alpha_1 + \alpha_2$ such that

$$2\langle x-y, b(x,\mu) - b(y,\nu)\rangle + \|\sigma(\mu) - \sigma(\nu)\|_{HS}^2 \le 2\alpha_1 |x-y|^2 + 2\alpha_2 \mathbf{W}_2(\mu,\nu)^2$$

holds for all $x, y \in \mathbb{H}$ and $\mu, \nu \in \mathscr{P}_2(\mathbb{H})$.

According to Theorem 3.1 below, assumption (H_2) implies that for any $X(0) \in L^2(\Omega \to \mathbb{H}, \mathscr{F}_0, \mathbb{P})$, the equation (2.6) has a unique mild solution $X(t)$. As before we denote by $X^\nu(t)$ the solution with initial distribution $\nu \in \mathscr{P}_2(\mathbb{H})$, and write $P_t^*\nu = \mathscr{L}_{X^\nu(t)}$. Moreover, by Itô's formula and $\kappa := \lambda_1 - (\alpha_1 + \alpha_2) > 0$, it is easy to see that P_t^* has a unique invariant probability measure $\bar\mu \in \mathscr{P}_2(\mathbb{H})$ and

$$\mathbf{W}_2(P_t^*\nu, \bar\mu) \le \mathrm{e}^{-\kappa t} \mathbf{W}_2(\nu, \bar\mu), \quad t \ge 0. \tag{2.7}$$

Consider the reference SPDE

$$\mathrm{d}\bar X(t) = \{A\bar X(t) + b(\bar X(t), \bar\mu)\}\mathrm{d}t + \sigma(\bar\mu)\mathrm{d}W(t),$$

which is again well-posed for any initial value $\bar X(0) \in L^2(\Omega \to \mathbb{H}, \mathscr{F}_0, \mathbb{P})$. Let J be the Donsker-Varadhan level 2 entropy function for the Markov process $\bar X(t)$, see Section 3. For any $r, R > 0$ let

$$\mathscr{B}_{r,R} := \big\{\nu \in \mathscr{P}(\mathbb{H}) : \nu(\mathrm{e}^{|\cdot|^r}) \le R\big\}.$$

Theorem 2.2. *Assume (H_2). If there exist constants $\varepsilon \in (0,1)$ and $c > 0$ such that*

$$\langle (-A)^{\gamma-1}x, b(x,\mu)\rangle \le c + \varepsilon |(-A)^{\frac{\gamma}{2}}x|^2, \quad x \in \mathscr{D}((-A)^{\frac{\gamma}{2}}), \tag{2.8}$$

then $\{L_t^\nu\}_{\nu \in \mathscr{B}_{r,R}} \in LDP_u(J)$ for all $r, R > 0$. If moreover $\bar P_t$ is strong Feller and $\bar\mu$-irreducible for some $t > 0$, then $\{L_t^\nu\}_{\nu \in \mathscr{B}_{r,R}} \in LDP(J)$ for all $r, R > 0$.

Assumption (H_2) is standard to imply the well-posedness of (2.6) and the exponential convergence of P_t^* in $\mathbf{W}_2$. Condition (2.8) is implied by

$$|(-A)^{\frac{\gamma}{2}-1} b(x,\mu)| \le \varepsilon' |(-A)^{\frac{\gamma}{2}}x| + c', \quad x \in \mathscr{D}((-A)^{\frac{\gamma}{2}}) \tag{2.9}$$

for some constants $\varepsilon' \in (0,1)$ and $c' > 0$. In particular, (2.8) holds if $|b(x,\mu)| \le c_1 + c_2|x|$ for some constants $c_1 > 0$ and $c_2 \in (0, \lambda_1)$. When σ is invertible with bounded σ^{-1} and $b(\cdot, \mu)$ is Lipschitz continuous, the dimension-free Harnack inequality established in [18, Theorem 3.4.1] implies the strong Feller property and $\bar\mu$-irreducibility of $\bar P_t$ for $t > 0$, see [18, Theorem 1.4.1] for more properties implied by this type Harnack inequality. Therefore, by Theorem 2.2, in this case (H_2) and (2.9) imply $\{L_t^\nu\}_{\nu \in \mathscr{B}_{r,R}} \in LDP(J)$ for all $r, R > 0$. See Example 2.4 below for the case where σ is non-invertible and b is possibly path-dependent.

2.3 Path-distribution dependent SPDE with additive noise

Let $\tilde{\mathbb{H}} = \mathbb{H}$ and $\sigma \in \mathbb{L}(\mathbb{H})$. Then (1.1) becomes

$$\mathrm{d}X(t) = \big\{AX(t) + b(X_t, \mathscr{L}_{X_t})\big\}\mathrm{d}t + \sigma \mathrm{d}W(t). \tag{2.10}$$

Below we consider this equation with σ being invertible and non-invertible respectively.

2.3.1 Invertible σ

Since σ is constant, we are able to establish LDP for $b(\xi, \cdot)$ being Lipshcitz continuous in $\mathbf{W}_p$ for some $p \geq 1$ rather than just for $p = 2$ as in the last two results.

(H_3) $\sigma \in \mathbb{L}(\mathbb{H})$ is constant and $(A, \mathscr{D}(A))$ satisfies the corresponding condition in (H_2). Moreover, there exist constants $p \geq 1$ and $\alpha_1, \alpha_2 \geq 0$ such that

$$|b(\xi, \mu) - b(\eta, \nu)| \leq \alpha_1 \|\xi - \eta\|_\infty + \alpha_2 \mathbf{W}_p(\mu, \nu), \quad \xi, \eta \in \mathscr{C}, \mu, \nu \in \mathscr{P}_p(\mathscr{C}).$$

Since (H_3) implies assumption **(A)** in Theorem 3.1 below, for any $X_0^\nu \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ with $\nu = \mathscr{L}_{X_0^\nu}$, the equation (1.1) has a unique mild segment solution X_t^ν with

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|X_t^\nu\|_\infty^p\Big] < \infty, \quad T > 0.$$

Let $P_t^* \nu = \mathscr{L}_{X_t^\nu}$ for $t \geq 0$ and $\nu \in \mathscr{P}_p(\mathscr{C})$.

When P_t^* has a unique invariant probability measure $\bar{\mu} \in \mathscr{P}_p(\mathscr{C})$, we consider the reference functional SPDE

$$\mathrm{d}\bar{X}(t) = \big\{A\bar{X}(t) + b(\bar{X}_t, \bar{\mu})\big\}\mathrm{d}t + \sigma \mathrm{d}W(t). \tag{2.11}$$

By Theorem 3.1 below, this reference equation is well-posed for any initial value in $L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$. For any $\varepsilon, R > 0$, let

$$\mathscr{I}_{\varepsilon,R} = \big\{\nu \in \mathscr{P}(\mathscr{C}) : \nu(\mathrm{e}^{\varepsilon\|\cdot\|_\infty^2}) \leq R\big\}.$$

Theorem 2.3. *Assume (H_3). Let $\theta \in [0, \lambda_1]$ such that*

$$\kappa_p := \theta - (\alpha_1 + \alpha_2)\mathrm{e}^{p\theta r_0} = \sup_{r\in[0,\lambda_1]} \big\{r - (\alpha_1 + \alpha_2)\mathrm{e}^{prr_0}\big\}.$$

(1) *For any $\nu_1, \nu_2 \in \mathscr{P}_p(\mathscr{C})$,*

$$\mathbf{W}_p(P_t^*\nu_1, P_t^*\nu_2)^p \leq \mathrm{e}^{p\theta r_0 - p\kappa_p t} \mathbf{W}_p(\nu_1, \nu_2)^p, \quad t \geq 0. \tag{2.12}$$

In particular, if $\kappa_p > 0$, then P_t^* has a unique invariant probability measure $\bar{\mu} \in \mathscr{P}_p(\mathscr{C})$ such that

$$\mathbf{W}_p(P_t^*\nu, \bar{\mu})^p \leq \mathrm{e}^{p\theta r_0 - p\kappa_p t} \mathbf{W}_p(\nu, \bar{\mu})^p, \quad t \geq 0, \nu \in \mathscr{P}_p(\mathscr{C}). \tag{2.13}$$

(2) *Let σ be invertible. If $\kappa_p > 0$ and $\sup_{s\in(0,\lambda_1]}(s - \alpha_1 \mathrm{e}^{sr_0}) > 0$, then $\{L_t^\nu\}_{\nu\in\mathscr{I}_{\varepsilon,R}} \in LDP(J)$ for any $\varepsilon, R > 0$, where J is the Donsker-Varadhan level 2 entropy function for the Markov process $\bar{X}_t$ on $\mathscr{C}$.*

Example 2.3. For a bounded domain $D \subset \mathbb{R}^d$, let $\mathbb{H} = L^2(D; \mathrm{d}x)$ and $A = -(-\Delta)^\alpha$, where Δ is the Dirichlet Laplacian on D and $\alpha > \frac{d}{2}$ is a constant. Let $\sigma = I$ be the identity operator on $\mathbb{H}$, and

$$b(\xi, \mu) = b_0(\mu) + \alpha_1 \int_{-r_0}^0 \xi(r)\Theta(\mathrm{d}r), \quad (\xi, \mu) \in \mathscr{C} \times \mathscr{P}_1(\mathscr{C}),$$

where $\alpha_1 \geq 0$ is a constant, Θ is a signed measure on $[-r_0, 0]$ with total variation 1 (i.e. $|\Theta|([-r_0, 0]) = 1$), and b_0 satisfies

$$|b_0(\mu) - b_0(\nu)| \leq \alpha_2 \mathbf{W}_1(\mu, \nu), \quad \mu, \nu \in \mathscr{P}_1(\mathscr{C})$$

for some constant $\alpha_2 \geq 0$. Then (H_3) holds for $p = 1$, and as shown in the proof of Example 1.1 in [1] that

$$\lambda_1 \geq \lambda := \frac{(d\pi^2)^\alpha}{R(D)^{2\alpha}},$$

where $R(D)$ is the diameter of D. Therefore, all assertions in Theorem 2.3 hold provided

$$\sup_{r \in (0, \lambda]} \{r - (\alpha_1 + \alpha_2)\mathrm{e}^{r r_0}\} > 0.$$

In particular, under this condition $\{L_t^\nu\}_{\nu \in \mathscr{I}_{\varepsilon, R}} \in LDP(J)$ for any $\varepsilon, R > 1$.

2.3.2 Non-invertible σ

Let $\mathbb{H} = \mathbb{H}_1 \times \mathbb{H}_2$ for two separable Hilbert spaces $\mathbb{H}_1$ and $\mathbb{H}_2$, and consider the following path-distribution dependent SPDE for $X(t) = (X^{(1)}(t), X^{(2)}(t))$ on $\mathbb{H}$:

$$\begin{cases} \mathrm{d}X^{(1)}(t) = \{A_1 X^{(1)}(t) + BX^{(2)}(t)\}\mathrm{d}t, \\ \mathrm{d}X^{(2)}(t) = \{A_2 X^{(2)}(t) + Z(X_t, \mathscr{L}_{X_t})\}\mathrm{d}t + \sigma \mathrm{d}W(t), \end{cases} \tag{2.14}$$

where $(A_i, \mathscr{D}(A_i))$ is a densely defined closed linear operator on $\mathbb{H}_i$ generating a C_0-semigroup e^{tA_i} $(i = 1, 2)$, $B \in \mathbb{L}(\mathbb{H}_2; \mathbb{H}_1)$, $Z : \mathscr{C} \mapsto \mathbb{H}_2$ is measurable, $\sigma \in \mathbb{L}(\mathbb{H}_2)$, and $W(t)$ is the cylindrical Wiener process on $\mathbb{H}_2$. Then (2.14) reduces to (2.10) with $A = \mathrm{diag}\{A_1, A_2\}$ and $\mathrm{diag}\{0, \sigma\}$ replacing σ, i.e. (2.14) is a special case of (2.10) with non-invertible σ.

For any $\alpha > 0$ and $p \geq 1$, define

$$\mathbf{W}_{p,\alpha}(\nu_1, \nu_2) := \inf_{\pi \in \mathscr{C}(\nu_1, \nu_2)} \left(\int_{\mathscr{C} \times \mathscr{C}} \left(\alpha \|\xi_1^{(1)} - \xi_2^{(1)}\|_\infty + \|\xi_1^{(2)} - \xi_2^{(2)}\|_\infty\right)^p \pi(\mathrm{d}\xi_1, \mathrm{d}\xi_2) \right)^{\frac{1}{p}}.$$

We assume

(H_4) Let $p \geq 1$ and $\alpha > 0$.

(H_4^1) $(-A_2, \mathscr{D}(A_2))$ is self-adjoint with discrete spectrum $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ counting multiplicities such that $\sum_{i=1}^\infty \lambda_i^{\gamma - 1} < \infty$ for some $\gamma \in (0, 1)$. Moreover, $A_1 \leq \delta - \lambda_1$ for some constant $\delta \geq 0$; i.e., $\langle A_1 x, x \rangle \leq (\delta - \lambda_1)|x|^2$ holds for all $x \in \mathscr{D}(A_1)$.

(H_4^2) There exist constants $K_1, K_2 > 0$ such that

$$|Z(\xi_1,\nu_1)-b(\xi_2,\nu_2)| \le K_1\|\xi_1^{(1)}-\xi_2^{(1)}\|_\infty+K_2\|\xi_1^{(2)}-\xi_2^{(2)}\|_\infty+K_3\mathbf{W}_{p,\alpha}(\nu_1,\nu_2),\ \ (\xi_i,\nu_i)\in\mathscr{C}\times\mathscr{P}_p(\mathscr{C}).$$

(H_4^3) σ is invertible on $\mathbb{H}_2$, and there exists $A_0 \in \mathbb{L}(\mathbb{H}_1;\mathbb{H}_1)$ such that for any $t > 0$, $B\mathrm{e}^{tA_2} = \mathrm{e}^{tA_1}\mathrm{e}^{tA_0}B$ holds and

$$Q_t := \int_0^t \mathrm{e}^{sA_0}BB^*\mathrm{e}^{sA_0^*}\mathrm{d}s$$

is invertible on $\mathbb{H}_1$.

By Theorem 3.1 for $\mathbb{H}_0 = \mathbb{H}_2$ and $\mathrm{diag}\{0,\sigma\}$ replacing σ, (H_4) implies that for any $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ this equation has a unique mild segment solution. Let $P_t^*\nu = \mathscr{L}_{X_t}$ for $\mathscr{L}_{X_0} = \nu \in \mathscr{P}_p(\mathscr{C})$.

Theorem 2.4. *Assume (H_4) for some constants $p \ge 1$ and $\alpha > 0$ satisfying*

$$\alpha \le \alpha' := \frac{1}{2\|B\|}\Big\{\delta - K_2 + \sqrt{(\delta-K_2)^2+4K_1\|B\|}\Big\}, \tag{2.15}$$

where $\|\cdot\|$ is the operator norm. If

$$\sup_{s\in(0,\lambda_1]} s\mathrm{e}^{-sr_0} > K_2 + \alpha'\|B\| + K_3, \tag{2.16}$$

then P_t^* has a unique invariant probability measure $\bar\mu$ such that

$$\mathbf{W}_p(P_t^*\nu,\bar\mu)^2 \le c_1\mathrm{e}^{-c_2t}\mathbf{W}_p(\nu,\bar\mu),\quad \nu\in\mathscr{P}_p(\mathscr{C}), t\ge 0 \tag{2.17}$$

holds for some constants $c_1, c_2 > 0$, and $\{L_t^\nu\}_{\nu\in\mathscr{I}_{\varepsilon,R}} \in LDP(J)$ for any $\varepsilon, R > 1$, where J is the Donsker-Varadhan level 2 entropy function for the associated reference equation for $\bar X(t)$.

Example 2.4. Consider the following equation for $X(t) = (X^{(1)}(t), X^{(2)}(t))$ on $\mathbb{H} = \mathbb{H}_0 \times \mathbb{H}_0$ for a separable Hilbert space $\mathbb{H}_0$:

$$\begin{cases}\mathrm{d}X^{(1)}(t) = \{\alpha_1 X^{(2)}(t) - AX^{(1)}(t)\}\mathrm{d}t\\ \mathrm{d}X^{(2)}(t) = \{Z(X(t),\mathscr{L}_{X(t)}) - AX^{(2)}(t)\}\mathrm{d}t + \mathrm{d}W(t),\end{cases}$$

where $\alpha_1 \in \mathbb{R}\setminus\{0\}$, $W(t)$ is the cylindrical Brownian motion on $\mathbb{H}_0$, A is a self-adjoint operator on $\mathbb{H}_0$ with discrete spectrum such that all eigenvalues $0 < \lambda_1 \le \lambda_2 \le \cdots$ counting multiplicities satisfy

$$\sum_{i=1}^\infty \lambda_i^{\gamma-1} < \infty$$

for some $\gamma \in (0,1)$, and Z satisfies

$$|Z(\xi_1,\nu_1) - Z(\xi_2,\nu_2)| \le \alpha_2\|\xi_1-\xi_2\|_\infty + \alpha_3\mathbf{W}_2(\nu_1,\nu_2),\quad (\xi_i,\nu_i)\in\mathscr{C}\times\mathscr{P}_2(\mathscr{C}), i=1,2.$$

Let

$$\alpha = \frac{1}{2\alpha_1}\Big(\sqrt{\alpha_2^2 + 4\alpha_1\alpha_2} - \alpha_2\Big).$$

Then P_t^* has a unique invariant probability measure $\bar\mu \in \mathscr{P}_2(\mathscr{C})$, and $\{L_t^\nu\}_{\nu\in\mathscr{I}_{R,q}} \in LDP(J)$ for any $R, q > 1$ if

$$\sup_{s\in[0,\lambda_1]} s\mathrm{e}^{-sr_0} > \alpha_2 + \alpha_1\alpha + \frac{\alpha_3}{1\wedge\alpha}. \tag{2.18}$$

Indeed, it is easy to see that assumption (H_4) holds for $p = 2$, $\delta = 0, K_1 = K_2 = \alpha_2$, $K_3 = \frac{\alpha_3}{1\wedge\alpha}$, $A_1 = A_2 = -A, B = \alpha_1$ and $A_0 = 0$. So, we have $\alpha = \alpha'$ and (2.18) is equivalent to (2.16). Then the desired assertion follows from Theorem 2.4.

3 Well-posedness of path-distribution dependent SPDEs

Consider the following path-distribution dependent SPDE on $\mathbb{H}$:

$$\mathrm{d}X(t) = \big\{AX(t) + b_t(X_t, \mathscr{L}_{X_t})\big\}\mathrm{d}t + \sigma_t(X_t, \mathscr{L}_{X_t})\mathrm{d}W(t), \tag{3.1}$$

where $(A, \mathscr{D}(A))$ is a negative self-adjoint operator on $\mathbb{H}$, and

$$b : [0,\infty)\times\mathscr{C}\times\mathscr{P}(\mathscr{C}) \to \mathbb{H}, \quad \sigma : [0,\infty)\times\mathscr{C}\times\mathscr{P}(\mathscr{C}) \to \mathbb{L}(\mathbb{H};\tilde{\mathbb{H}})$$

are measurable, and $W(t)$ is the cylindrical Brownian motion on $\tilde{\mathbb{H}}$.

Definition 3.1. An adapted continuous process $(X_t)_{t\ge0}$ on $\mathscr{C}$ is called a mild segment (or functional) solution of (3.1), if

$$\mathbb{E}\int_0^t \{|\mathrm{e}^{(t-s)A}b_s(X_s, \mathscr{L}_{X_s})| + \|\mathrm{e}^{(t-s)A}\sigma_s(X_s, \mathscr{L}_{X_s})\|_{HS}^2\}\mathrm{d}s < \infty, \quad t \ge 0,$$

and the process $X(t) := X_t(0)$ satisfies $\mathbb{P}$-a.s.

$$X(t) = \mathrm{e}^{At}X(0) + \int_0^t \mathrm{e}^{(t-s)A}b_s(X_s, \mathscr{L}_{X_s})\mathrm{d}s + \int_0^t \mathrm{e}^{(t-s)A}\sigma_s(X_s, \mathscr{L}_{X_s})\mathrm{d}W(s), \quad t \ge 0.$$

In this case, we call $(X(t))_{t\ge0}$ a mild solution of (3.1) with initial value X_0.

To ensure the existence and uniqueness of mild solutions with $X_0 \in L^p(\Omega\to\mathscr{C}, \mathscr{F}_0, \mathbb{P})$ for some $p > 0$, we make the following assumption.

(A) Let $p \in (0,\infty)$. There exists a subspace $\mathbb{H}_0$ of $\mathbb{H}$ such that $\sigma(\xi,\nu)\tilde{\mathbb{H}} \subset \mathbb{H}_0$ for any $(\xi,\nu) \in \mathscr{C}\times\mathscr{P}(\mathscr{C})$, and the orthogonal projection $\pi_0 : \mathbb{H} \to \mathbb{H}_0$ satisfies $A\pi_0 = \pi_0 A$ on $\mathscr{D}(A)$. Moreover, there exist $\gamma \in (0,1)$ and $1 \le K \in L^1_{loc}([0,\infty)\to[0,\infty))$ such that

(A_1) $\int_0^t s^{-\gamma}\|\mathrm{e}^{sA}\pi_0\|_{HS}^2\mathrm{d}s < \infty$, $t \in (0,\infty)$.

(A_2) There exists $p_0 > 2$ such that for any $t \geq 0, \xi, \eta \in \mathscr{C}$ and $\mu, \nu \in \mathscr{P}_p(\mathscr{C})$,

$$|b_t(\xi,\mu) - b_t(\eta,\nu)| \leq K(t)\big(\|\xi-\eta\|_\infty + \mathbf{W}_p(\mu,\nu)\big),$$
$$\|\sigma_t(\xi,\mu) - \sigma_t(\eta,\nu)\|^p \leq K(t)^{1\wedge\frac{p}{p_0}}\big(\|\xi-\eta\|_\infty^p + \mathbf{W}_p(\mu,\nu)^p\big).$$

(A_3) $|b_t(0,\delta_0)| + \|\sigma_t(0,\delta_0)\|^{p\vee p_0} \leq K(t), \quad t \geq 0.$

In many references (A_1) is replaced by $\int_0^t s^{-\gamma}\|e^{As}\|_{HS}^2 ds < \infty$, see for instance [4]. The present weaker version allows us to cover more examples with degenerate noise.

Remark 3.1. By **(A)** we have $e^{A(t-s)}\sigma_s = e^{A(t-s)}\pi_0\sigma_s$, so that using $e^{A(t-s)}\pi_0$ to replace the semigroup $S(t-s)$ in the proof of [4, Proposition 7.9], if $\Phi(s)$ is an adapted process on $\mathbb{L}(\mathbb{H};\tilde{\mathbb{H}})$ such that $\mathbb{E}\int_0^t \|\Phi(s)\|^q ds < \infty$ for some $q > 2$, then

$$W_\Phi(t) := \int_0^t e^{A(t-s)}\pi_0\Phi(s)dW(s), \quad t \geq 0$$

is an adapted continuous process on $\mathbb{H}$ such that

$$\mathbb{E}\bigg[\bigg|\sup_{s\in[0,t]}\int_0^t e^{A(s-r)}\pi_0\Phi(r)dW(r)\bigg|^q\bigg] \leq c\mathbb{E}\int_0^t \|\Phi(s)\|^q ds$$

holds for some constant $c > 0$.

In the following result, "$p = 2$" is included in both (2) and (3), but conditions in these two situations are incomparable: comparing with (2), (3) allows $\sigma(\xi,\mu)$ depending on ξ which is more general on the one hand, but assumes the Lipschitz condition in the Hilbert-Schmidt norm which is more restrictive on the other hand.

Theorem 3.1. *Assume* **(A)** *and let* $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$. *Then* (3.1) *has a unique mild segment solution* $\{X_t\}_{t\geq 0}$ *starting at* X_0 *with*

$$\mathbb{E}\Big[\sup_{t\in[0,T]}\|X_t\|_\infty^p\Big] < \infty, \quad T \in (0,\infty),$$

provided one of following conditions holds:

(1) $p > 2$.

(2) $p \in (0,2]$ *and* $\sigma_s(\xi,\mu)$ *does not depend on* ξ.

(3) $p = 2$ *and for any* $s \geq 0, \xi, \eta \in \mathscr{C}$ *and* $\mu, \nu \in \mathscr{P}_2(\mathscr{C})$,

$$\|\sigma_s(\xi,\mu) - \sigma_s(\eta,\nu)\|_{HS}^2 \leq K(s)\{\|\xi-\eta\|_\infty^2 + \mathbf{W}_2(\mu,\nu)^2\}.$$

Proof. We consider cases (1)-(3) respectively.

Proof for Case (1). Let $p > 2$.
The existence. We adopt an iteration argument as in [20]. It suffices to prove that for any fixed $T > 0$, the SPDE has a unique mild segment solution up to time T satisfying

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|X_t\|_\infty^p\Big] < \infty. \tag{3.2}$$

(1a) We first consider the case that X_0 is bounded. Let $X_t^0 = X_0$ and $\mu_t^0 = \mathscr{L}_{X_t^0}$ for $t \geq 0$. By Remark 3.1,

$$X^1(t) := \mathrm{e}^{At}X(0) + \int_0^t \mathrm{e}^{(t-s)A} b_s(X_s^0, \mu_s^0)\mathrm{d}s + \int_0^t \mathrm{e}^{(t-s)A}\sigma_s(X_s^0, \mu_s^0)\mathrm{d}W(s), \quad t \geq 0,$$

is an adapted continuous process on $\mathbb{H}$ such that

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|X_t^1\|_\infty^q\Big] < \infty, \quad q > 0, \tag{3.3}$$

where $X_t^1(r) := X^1(t+r)1_{\{t+r\geq 0\}} + X_0(t+r)1_{\{t+r<0\}}$.

Now, assume that for some $n \geq 1$ we have constructed a continuous adapted process $\{X_t^n\}_{t\in[0,T]}$ on $\mathscr{C}$ with $X_0^n = X_0$ and

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|X_t^n\|_\infty^q\Big] < \infty, \quad q > 0.$$

By Remark 3.1,

$$X^{n+1}(t) := \mathrm{e}^{At}X(0) + \int_0^t \mathrm{e}^{(t-s)A} b_s(X_s^n, \mu_s^n)\mathrm{d}s + \int_0^t \mathrm{e}^{(t-s)A}\sigma_s(X_s^n, \mu_s^n)\mathrm{d}W(s), \quad t \in [0,T] \tag{3.4}$$

for $\mu_s^n := \mathscr{L}_{X_s^n}$ is an adapted continuous process on $\mathbb{H}$, and the segment process X_t^{n+1} given by

$$X_t^{n+1}(r) := X^{n+1}(t+r)1_{\{t+r\geq 0\}} + X_0(t+r)1_{\{t+r<0\}} \text{ for } r \in [-r_0, 0], \quad t \geq 0 \tag{3.5}$$

satisfies

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|X_t^{n+1}\|_\infty^q\Big] < \infty, \quad q > 0.$$

It suffices to find a constant $t_0 > 0$ independent of X_0 such that $\{X_{[0,t_0]}^n\}_{n\geq 1}$ is a Cauchy sequence in $L^p(\Omega \to C([0,t_0];\mathscr{C}), \mathbb{P})$. This together with assumption **(A)** imply that the limit $X_{[0,t_0]} := \lim_{n\to\infty} X_{[0,t_0]}^n$ gives rise to a mild segment solution of (3.1) up to time t_0. By repeating the procedure with initial time it_0 and initial value X_{it_0} for $i \geq 1$, in finite many steps we may construct a mild segment solution of (3.1) up to time T, such that (3.2) holds.

For any $n \geq 1$, by (3.4), (3.5) and assumption **(A)** we have

$$\begin{aligned} \psi_n(t) &:= \sup_{s\in[0,t]} \|X_s^{n+1} - X_s^n\|_\infty = \sup_{s\in[0,t]} |X^{n+1}(s) - X^n(s)| \\ &\leq \int_0^t K(s)\big\{\|X_s^n - X_s^{n-1}\|_\infty + \mathbf{W}_p(\mu_s^n, \mu_s^{n-1})\big\}\mathrm{d}s + \sup_{s\in[0,t]}\Big|\int_0^s \mathrm{e}^{A(s-r)}\Phi_n(r)\mathrm{d}W(r)\Big|, \end{aligned} \tag{3.6}$$

where $\Phi_n(r) := \sigma_r(X_r^n, \mu_r^n) - \sigma_r(X_r^{n-1}, \mu_r^{n-1})$ satisfies

$$\|\Phi_n(r)\|^p \le K(r)\big\{\|X_r^n - X_r^{n-1}\|_\infty^p + \mathbf{W}_p(\mu_r^n, \mu_r^{n-1})^p\big\}.$$

Combining this with $\mathbf{W}_p(\mu_r^n, \mu_r^{n-1})^p \le \mathbb{E}\|X_r^n - X_r^{n-1}\|_\infty^p$, and noting that Remark 3.1 implies

$$\mathbb{E}\bigg[\sup_{s\in[0,t]}\bigg|\int_0^s \mathrm{e}^{A(s-r)}\Phi_n(r)\mathrm{d}W(r)\bigg|^p\bigg] \le c\mathbb{E}\int_0^t \|\Phi_n(s)\|^p\mathrm{d}s \tag{3.7}$$

for some constant $c > 0$, we find constants $C_1, C_2 > 0$ such that

$$\begin{aligned}\mathbb{E}[\psi_n(t)^p] &\le C_1\mathbb{E}\bigg(\int_0^t K(s)\big\{\psi_{n-1}(s) + \mathbf{W}_p(\mu_s^n, \mu_s^{n-1})\mathrm{d}s\bigg)^p \\ &\quad + C_1\mathbb{E}\int_0^t K(s)\big\{\|X_s^n - X_s^{n-1}\|_\infty^p + \mathbf{W}_p(\mu_s^n, \mu_s^{n-1})^p\big\}\mathrm{d}s \\ &\le C_2\varepsilon(t)\mathbb{E}[\psi_{n-1}^p(t)], \quad \varepsilon(t) := \bigg(\int_0^t K(s)\mathrm{d}s\bigg)^p + \int_0^t K(s)\mathrm{d}s.\end{aligned} \tag{3.8}$$

Taking $t_0 \in (0, T]$ such that $C_2\varepsilon(t_0) \le \frac{1}{2}$, we obtain

$$\mathbb{E}[\psi_n^p(t_0)] \le 2^{-n}\mathbb{E}[\psi_0^p(t_0)] < \infty, \quad n \ge 1.$$

Thus, $\{X_{[0,t_0]}^n\}_{n\ge1}$ is a Cauchy sequence in $L^p(\Omega \to C([0, t_0]; \mathscr{C}), \mathbb{P})$ as desired.

(1b) In general, for $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ and $N \in \mathbb{N}$ let $X_0^{(N)} = X_0 1_{\{\|X_0\|_\infty \le N\}}$. By (1a), for any $N \ge 1$ we have constructed a mild segment solution $(X_t^{(N)})_{t\in[0,T]}$ for (3.1) satisfying (3.2) with initial value $X_0^{(N)}$:

$$X^N(t) := \mathrm{e}^{At}X^{(N)}(0) + \int_0^t \mathrm{e}^{(t-s)A}b_s(X_s^{(N)}, \mu_s^{(N)})\mathrm{d}s + \int_0^t \mathrm{e}^{(t-s)A}\sigma_s(X_s^{(N)}, \mu_s^{(N)})\mathrm{d}W(s), \quad t \in [0, T],$$

where $\mu_s^{(N)} = \mathscr{L}_{X_s^{(N)}}$. By the above argument for $X^{(N)}(t) - X^{(M)}(t)$ instead of $X^{n+1}(t) - X^n(t)$, we find a constant $C > 0$ such that for any $N, M \ge 1$, the process

$$\psi_{N,M}(t) := \sup_{s\in[0,t]} \|X_s^{(N)} - X_s^{(M)}\|_\infty^p, \quad t \in [0, T]$$

satisfies

$$\mathbb{E}[\psi_{N,M}(t)] \le C\mathbb{E}\big[\|X_0\|_\infty^p 1_{\{\|X_0\|_\infty > N\wedge M\}}\big] + C\varepsilon(t)\mathbb{E}[\psi_{N,M}(t)], \quad t \in [0, T]. \tag{3.9}$$

Taking $t_0 \in (0, T]$ such that $C\varepsilon(t_0) \le \frac{1}{2}$, we obtain

$$\mathbb{E}[\psi_{N,M}(t_0)] \le 2C\mathbb{E}\big[\|X_0\|_\infty^p 1_{\{\|X_0\|_\infty > N\wedge M\}}\big], \quad N, M \ge 1,$$

so that, $\{X_{[0,t_0]}^{(N)}\}_{N\ge1}$ is a Cauchy sequence in $L^p(\Omega \to C([0, t_0]; \mathscr{C}), \mathbb{P})$, and it is easy to see that its limit as $N \to \infty$ is a solution of (3.1) up to time t_0. As explained before that by repeating

the procedure we construct a mild segment solution of (3.1) up to time T satisfying (3.2).
The uniqueness. Let X_t and Y_t be two mild segment solutions with initial value X_0 satisfying

$$\mathbb{E}\bigg[\sup_{t\in[0,T]}\big(\|X_t\|_\infty^p+\|Y_t\|_\infty^p\big)\bigg]<\infty.$$

Similarly to (3.8) we have

$$\mathbb{E}\Big[\sup_{s\in[0,t]}\|X_s-Y_s\|_\infty^p\Big]\le C\varepsilon(t)\mathbb{E}\Big[\sup_{s\in[0,t]}\|X_s-Y_s\|_\infty^p\Big],\quad t\in[0,T].$$

This implies $X_t=Y_t$ up to time $t_0\in(0,T]$ such that $C\varepsilon(t_0)<1$. Since this t_0 does not depend on the initial value, repeating the same argument leads to $X_t=Y_t$ for all $t\in[0,T]$.

Proof for Case (2). Let $p\in(0,2]$. Again we first assume that X_0 is bounded and let X^n,μ^n,ψ_n be defined in step (1a). Since $\sigma_s(\xi,\mu)$ does not depend on ξ and $K(s)\ge 1$, by (A_2), $\Phi_n(s)$ in (3.6) satisfies

$$\|\Phi_n(s)\|^{p_0}\le K(s)\mathbf{W}_p(\mu_s^n,\mu_s^{n-1})^{p_0}.$$

Combining this with Remark 3.1 for $q=p_0>2$, and using $\mathbf{W}_p(\mu_s^n,\mu_s^{n-1})^p\le\mathbb{E}\|X_s^n-X_s^{n-1}\|_\infty^p$, we find a constant $C_1>0$ such that

$$\begin{aligned}\mathbb{E}[\psi_n^p(t)]&\le C_1\bigg(\mathbb{E}\int_0^t K(s)\psi_{n-1}(s)\mathrm{d}s\bigg)^p+C_1\bigg(\int_0^t K(s)\mathbf{W}_p(\mu_s^n,\mu_s^{n-1})^{p_0}\mathrm{d}s\bigg)^{\frac{p}{p_0}}\\&\le C_1\delta(t)\mathbb{E}[\psi_{n-1}^p(t)],\quad t\in[0,T],n\ge 1\end{aligned}$$

holds for $\delta(t):=(\int_0^t K(s)\mathrm{d}s)^p+(\int_0^t K(s)\mathrm{d}s)^{\frac{p}{p_0}}$. Then the remainder of the proof, including the existence and uniqueness for bounded X_0, and the extension to general $X_0\in L^p(\Omega\to\mathscr{C},\mathscr{F}_0,\mathbb{P})$, is similar to that in Case (1).

Proof for Case (3). Let $p=2$. As explained above we only consider bounded X_0. In this case, let X^n,μ^n,ψ_n be defined in step (1a). By **(A)** and Itô's formula to $|X^{n+1}(t)-X^n(t)|^2$, we find a constant $c>0$ such that

$$\begin{aligned}&\mathrm{d}|X^{n+1}(t)-X^n(t)|^2\\&\le cK(t)\Big\{|X^{n+1}(t)-X^n(t)|^2+\|X_t^n-X_t^{n-1}\|_\infty^2+\mathbf{W}_2(\mu_t^n,\mu_t^{n-1})^2\Big\}\mathrm{d}t+\mathrm{d}M^n(t),\end{aligned}\tag{3.10}$$

where

$$M^n(t):=2\int_0^t\big\langle X^{n+1}(s)-X^n(s),\{\sigma_s(X_s^n,\mu_s^n)-\sigma_s(X_s^{n-1},\mu_s^{n-1})\}\mathrm{d}W(s)\big\rangle$$

satisfies

$$\mathrm{d}\langle M^n(t)\rangle\le 4K(t)|X^{n+1}(t)-X^n(t)|^2\big\{\|X_t^n-X_t^{n-1}\|_\infty^2+\mathbf{W}_2(\mu_t^n,\mu_t^{n-1})^2\big\}\mathrm{d}t.$$

By (3.10), we obtain

$$|X^{n+1}(t)-X^n(t)|^2\mathrm{e}^{-c\int_0^t K(s)ds}$$

$$\leq \int_0^t K(s)\mathrm{e}^{-c\int_0^s K(r)\mathrm{d}r}\big(\|X_s^n - X_s^{n-1}\|_\infty^2 + \mathbf{W}_2(\mu_s^n,\mu_s^{n-1})^2\big)\mathrm{d}s + \int_0^t \mathrm{e}^{-\int_0^s cK(r)\mathrm{d}r}\mathrm{d}M^n(s)$$

for $t \in [0,T]$. Therefore, by the BDG inequality, there exist constants $C_1, C_2 > 0$ depending only on T such that

$$\begin{aligned}
&\mathbb{E}[\psi_n^2(t)] \leq \mathrm{e}^{c\int_0^T K(s)\mathrm{d}s}\mathbb{E}\Big[\sup_{s\in[0,t]}|X^{n+1}(s) - X^n(s)|^2\mathrm{e}^{-c\int_0^s K(r)dr}\Big]\\
&\leq C_1\int_0^t K(s)\big\{\mathbb{E}[\psi_{n-1}^2(s)] + \mathbf{W}_2(\mu_s^n,\mu_s^{n-1})^2\big\}\mathrm{d}s\\
&\qquad + C_1\mathbb{E}\bigg[\bigg(\int_0^t K(s)\psi_n^2(s)\big\{\psi_{n-1}^2(s) + \mathbf{W}_2(\mu_s^n,\mu_s^{n-1})^2\big\}\mathrm{d}s\bigg)^{\frac{1}{2}}\bigg]\\
&\leq \frac{1}{2}\mathbb{E}[\psi_n^2(t)] + C_2\int_0^t K(s)\big\{\mathbb{E}[\psi_{n-1}^2(s)] + \mathbf{W}_2(\mu_s^n,\mu_s^{n-1})^2\big\}\mathrm{d}s,\quad t\in[0,T].
\end{aligned}$$

Noting that $\mathbf{W}_2(\mu_s^n,\mu_s^{n-1})^2 \leq \mathbb{E}[\psi_{n-1}^2(s)]$, this implies

$$\mathbb{E}[\psi_n^2(t)] \leq 4C_2\int_0^t K(s)\mathbb{E}[\psi_{n-1}^2(s)]\mathrm{d}s \leq 4C_2\mathbb{E}[\psi_{n-1}^2(t)]\int_0^t K(s)\mathrm{d}s,\quad t\in[0,T], n\geq 1.$$

Taking $t_0 \in (0,T]$ such that $4C_2\int_0^{t_0} K(s)\mathrm{d}s \leq \frac{1}{2}$, we obtain

$$\mathbb{E}[\psi_n^2(t_0)] \leq 2^{-(n-1)}\mathbb{E}[\psi_0^2(t_0)] < \infty,\quad n\geq 1.$$

Thus, $\{X_{[0,t_0]}^n\}_{n\geq 1}$ is a Cauchy sequence in $L^2(\Omega \to C([0,t_0];\mathscr{C}),\mathbb{P})$ as desired. The remainder of the proof is similar to that in Case (1). □

4 Proofs of main results

4.1 Proof of Theorem 2.1

We first prove the well-posedness for the following more general SDE such that the first assertion follows:

$$\mathrm{d}X(t) = b(X(t),\mathscr{L}_{X(t)})\mathrm{d}t + \sigma(X(t),\mathscr{L}_{X(t)})\mathrm{d}W(t), \tag{4.1}$$

where $W(t)$ is the m-dimensional Brownian motion, and

$$b:\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}^d,\quad \sigma:\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}^d\otimes\mathbb{R}^m$$

are continuous such that

$$\begin{aligned}
&\langle b(x,\mu) - b(y,\nu), x-y\rangle + \frac{1}{2}\|\sigma(x,\mu)-\sigma(y,\nu)\|_{HS}^2 \leq K_1|x-y|^2 + K_2\mathbf{W}_2(\mu,\nu)^2,\\
&\|\sigma(x,\mu)\| \leq K_3\big\{|x-y| + \mathbf{W}_2(\mu,\nu)\big\},\quad x,y\in\mathbb{R}^d, \mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)
\end{aligned}\tag{4.2}$$

holds for some constants $K_1\in\mathbb{R}, K_2, K_3 \geq 0$. We have the following result.

Lemma 4.1. *Let b and σ be continuous such that (4.2) holds for some constants $K_1 \in \mathbb{R}, K_2, K_3 \geq 0$. Then (4.1) has a unique solution for initial value with $\mathscr{L}_{X(0)} \in \mathscr{P}_2(\mathbb{R}^d)$, and the associated operator P_t^* satisfies*

$$\mathbf{W}_2(P_t^*\mu, P_t^*\nu) \leq \mathrm{e}^{(K_1+K_2)t}\mathbf{W}_2(\mu,\nu), \quad t \geq 0, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d). \tag{4.3}$$

Consequently, if $K_1 + K_2 < 0$ then P_t^* has a unique invariant probability measure $\bar{\mu} \in \mathscr{P}_2(\mathbb{R}^d)$.

Proof. For any $T > 0$ and initial value $X(0)$ with $\mathscr{L}_{X(0)} \in \mathscr{P}_2(\mathbb{R}^d)$, consider the space

$$\mathscr{D} := \big\{\mu : [0,T] \to \mathscr{P}_2(\mathbb{R}^d) \text{ is continuous, } \mu_0 = \mathscr{L}_{X(0)}\big\},$$

which is a complete metric space under

$$\mathbf{W}_{2,\lambda}(\mu,\nu) := \sup_{t\in[0,T]} \mathrm{e}^{-\lambda t}\mathbf{W}_2(\mu_t,\nu_t)$$

for $\lambda > 0$. For any $\mu_\cdot \in C([0,T]; \mathscr{P}_2(\mathbb{R}^d))$, consider the SDE

$$\mathrm{d}X^\mu(t) = b(X^\mu(t),\mu_t)\mathrm{d}t + \sigma(X^\mu(t),\mu_t)\mathrm{d}W(t), \quad t \geq 0, X^\mu(0) = X(0).$$

It is well known that the conditions on b and σ imply the well-posedness of this SDE, and $\mathscr{L}_{X_\cdot} \in \mathscr{D}$. So, we can define a map

$$\mathscr{D} \ni \mu \mapsto H(\mu) := \mathscr{L}_{X_\cdot} \in \mathscr{D}.$$

It remains to prove the contraction of H under the metric $W_{2,\lambda}$ for some $\lambda > 0$, which implies that H has a unique fixed point μ so that X^μ is the unique solution to (4.1). By (4.2) and Itô's formula, for any $\mu, \nu \in \mathscr{D}$ we have

$$\mathrm{d}|X^\mu(t) - X^\nu(t)|^2 \leq 2\big\{K_1|X^\mu(t) - X^\nu(t)|^2 + K_2\mathbf{W}_2(\mu_t,\nu_t)^2\big\}\mathrm{d}t + \mathrm{d}M(t)$$

for some martingale $M(t)$. This implies that for any $\lambda > K_1$,

$$\begin{aligned}&\mathrm{e}^{-2\lambda t}\mathbf{W}_2(H(\mu)(t),H(\nu)(t))^2 \leq \mathrm{e}^{-2\lambda t}\mathbb{E}|X^\mu(t) - X^\nu(t)|^2\\ &\leq 2K_2\int_0^t \mathrm{e}^{2(K_1-\lambda)(t-s)}\mathrm{e}^{-2\lambda s}\mathbf{W}_2(\mu_s,\nu_s)^2\mathrm{d}s \leq \frac{K_2}{\lambda - K_1}\mathbf{W}_{2,\lambda}(\mu,\nu)^2, \quad t \in [0,T].\end{aligned}$$

Therefore, H is contraction in $\mathbf{W}_{2,\lambda}$ for $\lambda > K_1 + K_2$, and hence, (4.1) has a unique solution for $\mathscr{L}_{X(0)} \in \mathscr{P}_2(\mathbb{R}^d)$.

Next, for two solutions $X(t)$ and $Y(t)$ of (4.1) with initial values satisfying

$$\mathscr{L}_{X(0)} = \mu, \quad \mathscr{L}_{Y(0)} = \nu, \quad \mathbf{W}_2(\mu,\nu)^2 = \mathbb{E}|X(0) - Y(0)|^2,$$

by (4.2), Itô's formula and Gronwall's lemma, we obtain

$$\mathbf{W}_2(P_t^*\mu, P_t^*\nu)^2 \leq \mathbb{E}|X(t) - Y(t)|^2 \leq \mathrm{e}^{2(K_1+K_2)t}\mathbb{E}|X(0) - Y(0)|^2 = \mathrm{e}^{2(K_1+K_2)t}\mathbf{W}_2(\mu,\nu)^2,$$

so that (4.3) holds. According to the proof of [20, Theorem 3.1], when $K_1 + K_2 < 0$ this implies that P_t^* has a unique invariant probability measure $\bar{\mu} \in \mathscr{P}_2(\mathbb{R}^d)$.

□

Below, we prove the other two assertions in Theorem 2.1 respectively.

Proof of (2). According to Theorem 2.1(1), (H_1) implies that the reference SDE

$$\mathrm{d}\bar{X}(t) = b(\bar{X}(t), \bar{\mu})\mathrm{d}t + \sigma(\bar{\mu})\mathrm{d}W(t)$$

is well-posed and the solution is a Markov Feller process, where $\bar{\mu}$ is the unique invariant probability measure of P_t^*. Let $\bar{X}^x(t)$ denote the solution starting at x. According to Theorem 5.3 and Theorem 5.5, we only need to prove the following assertions:

(2a) For any $\lambda > 0$, there exist a constant $s > 0$ and compact set $K \subset \mathbb{R}^d$, such that (5.4) holds for any compact set $K' \subset \mathbb{R}^d$ and

$$\tau_K^x := \inf\{t \geq 0 : \bar{X}^x(t) \in K\}, \quad x \in \mathbb{R}^d.$$

(2b) For any $N \geq 1$,

$$\sup_{\nu \in \mathscr{B}_{r,R}} \mathbb{E}\mathrm{e}^{N \int_0^\infty \{1 \wedge |X^\nu(s) - \bar{X}^0(s)|^2\}\mathrm{d}s} < \infty.$$

Indeed, by Theorem 5.3(1), (2a) implies the upper LDP (LDP if $\bar{P}_t$ is strong Feller and $\bar{\mu}$-irreducible) for $\bar{L}_t^x$ locally uniformly in x, in particular, L_t^0 satisfies the upper LDP (LDP if $\bar{P}_t$ is strong Feller and $\bar{\mu}$-irreducible). Combining this with (2b) and Theorem 5.5 for $\mathscr{I} = \mathscr{B}_{r,R}$ and $\Psi(\nu) := \delta_0$, we prove the desired assertion for L_t^ν with $\nu \in \mathscr{B}_{r,R}$.

Proof of (2a). By (H_1), there exist constants $\alpha, \beta > 0$ such that

$$\mathrm{d}|\bar{X}(t)|^2 \leq 2\{\alpha - \beta|\bar{X}(t)|^2\}\mathrm{d}t + 2\langle \bar{X}(t), \sigma(\bar{\mu})\mathrm{d}W(t)\rangle. \tag{4.4}$$

Let $\theta = \|\sigma\|_\infty^2$. Then for any $\varepsilon \in (0, \beta/\theta)$, there exist constants $c_1, c_2 > 0$ such that

$$\begin{aligned}\mathrm{d}\mathrm{e}^{\varepsilon|\bar{X}(t)|^2} &\leq 2\varepsilon\big\{\alpha - (\beta - \varepsilon\theta)|\bar{X}(t)|^2\big\}\mathrm{e}^{\varepsilon|\bar{X}(t)|^2}\mathrm{d}t + \mathrm{d}M(t)\\ &\leq \Big\{c_1 - c_2\mathrm{e}^{\varepsilon|\bar{X}(t)|^2}\Big\}\mathrm{d}t + \mathrm{d}M(t)\end{aligned}$$

for some martingale $M(t)$. So,

$$\mathbb{E}\mathrm{e}^{\varepsilon|\bar{X}^x(t)|^2} \leq \mathrm{e}^{\varepsilon|x|^2} + \frac{c_1}{c_2}, \quad x \in \mathbb{R}^d. \tag{4.5}$$

To estimate τ_K^x for $K := B_0(N)$, we take $N \geq N_0 := (2\alpha/\beta)^{\frac{1}{2}}$. Then (4.4) implies

$$\mathrm{d}|\bar{X}(t)|^2 \leq -\beta|\bar{X}(t)|^2\mathrm{d}t + 2\langle \bar{X}^x(t), \sigma(\bar{\mu})\mathrm{d}W(t)\rangle, \quad t \leq \tau_K^x.$$

For any $\delta > 0$, we obtain

$$\begin{aligned}&\mathbb{E}\mathrm{e}^{\delta\int_0^{t\wedge\tau_K^x}|\bar{X}^x(s)|^2\mathrm{d}s} \leq \mathrm{e}^{\delta\beta^{-1}|x|^2}\mathbb{E}\mathrm{e}^{2\delta\beta^{-1}\int_0^{t\wedge\tau_K^x}\langle\bar{X}^x(s), \sigma(\bar{X}^x(s),\mu)\mathrm{d}W(s)\rangle}\\ &\leq \mathrm{e}^{\delta\beta^{-1}|x|^2}\big(\mathbb{E}\mathrm{e}^{8\delta^2\beta^{-2}\theta\int_0^{t\wedge\tau_K^x}|\bar{X}^x(s)|^2\mathrm{d}s}\big)^{\frac{1}{2}}.\end{aligned}$$

Thus, taking $\delta \leq \frac{\beta^2}{8\theta}$ we arrive at

$$\mathbb{E}e^{\delta N^2(t\wedge\tau_K^x)} \leq \mathbb{E}e^{\delta\int_0^{t\wedge\tau_K^x}|\bar{X}^x(s)|^2\mathrm{d}s} \leq e^{2\delta\beta^{-1}|x|^2}.$$

Letting $t \uparrow \infty$ implies

$$\mathbb{E}e^{\delta N^2\tau_K^x} \leq e^{2\delta\beta^{-1}|x|^2}, \quad x \in \mathbb{R}^d, N \geq N_0. \tag{4.6}$$

Combining this with the Markov property and (4.5), when $\delta \leq \frac{\varepsilon\beta}{2}$ we have

$$\mathbb{E}e^{\delta N^2\tau_K^{\bar{X}^x(s)}} \leq \mathbb{E}e^{2\delta\beta^{-1}|\bar{X}^x(s)|^2} \leq e^{\varepsilon|x|^2} + \frac{c_1}{c_2}, \quad x \in \mathbb{R}^d, s \geq 0, N \geq N_0.$$

Therefore, for any $\lambda > 0$ there exists compact $K \subset \mathbb{R}^d$ such that (5.4) holds.

Proof of (2b). Simply denote $X(t) = X^\nu(t), \bar{X}(t) = \bar{X}^0(t)$ and $\nu_t = \mathscr{L}_{X^\nu(t)} = P_t^*\nu$ for $\nu \in \mathscr{B}_{r,R}$. By (H_1), (2.2) and Itô's formula, we obtain

$$\begin{aligned}\mathrm{d}|X(t)-\bar{X}(t)|^2 \leq& \big\{-\kappa_1|X(t)-\bar{X}(t)|^2 + \kappa_2 e^{-(\kappa_1-\kappa_2)t}\mathbf{W}_2(\bar{\mu},\nu)^2\big\}\mathrm{d}t \\ &+ 2\langle X(t)-\bar{X}(t), \{\sigma(\nu_t)-\sigma(\bar{\mu})\}\mathrm{d}W(t)\rangle.\end{aligned}$$

Letting $\gamma(t) = \frac{|X(t)-\bar{X}(t)|^2}{1+|X(t)-\bar{X}(t)|^2}$, we derive

$$\begin{aligned}&\mathrm{d}\log(1+|X(t)-\bar{X}(t)|^2) \leq \big\{-\kappa_1\gamma(t) + \kappa_2 e^{-(\kappa_1-\kappa_2)t}\mathbf{W}_2(\bar{\mu},\nu)^2\big\}\mathrm{d}t \\ &+ \frac{2}{1+|X(t)-\bar{X}(t)|^2}\langle X(t)-\bar{X}(t), \{\sigma(\nu_t)-\sigma(\bar{\mu})\}\mathrm{d}W(t)\rangle.\end{aligned}$$

We deduce from this and (2.2) that for any $\lambda > 0$,

$$\begin{aligned}&e^{-\frac{\lambda\kappa_2}{\kappa_1-\kappa_2}\mathbf{W}_2(\bar{\mu},\nu)^2}\mathbb{E}\big[e^{\lambda\kappa_1\int_0^t\gamma(s)\mathrm{d}s}\big] \\ &\leq \mathbb{E}\Big[(1+|X(0)|^2)^\lambda e^{\lambda\int_0^t\frac{2}{1+|X(s)-\bar{X}(s)|^2}\langle X(s)-\bar{X}(s),\{\sigma(\nu_s)-\sigma(\mu)\}\mathrm{d}W(s)\rangle}\Big] \\ &\leq \mathbb{E}\Big[(1+|X(0)|^2)^\lambda\Big(\mathbb{E}\big[e^{8\kappa_2\lambda^2\int_0^t\gamma(s)\mathbf{W}_2(\nu_s,\bar{\mu})^2\mathrm{d}s}\big]\Big|\mathscr{F}_0\Big)^{\frac{1}{2}}\Big] \\ &\leq \big\{\nu\big((1+|\cdot|^2)^{2\lambda}\big)\big\}^{\frac{1}{2}}\Big(\mathbb{E}\big[e^{8\kappa_2\lambda^2\mathbf{W}_2(\nu,\bar{\mu})^2\int_0^t\gamma(s)e^{-(\kappa_1-\kappa_2)s}\mathrm{d}s}\big]\Big)^{\frac{1}{2}} \\ &\leq C(\lambda,R)\Big(\mathbb{E}\big[e^{\lambda\kappa_1\int_0^t\gamma(s)\mathrm{d}s}\big]\Big)^{\frac{1}{2}}, \quad t > 0\end{aligned} \tag{4.7}$$

holds for some constant $C(\lambda,R) > 0$, where the last step is due to $\gamma(s) \leq 1$ and $\nu \in \mathscr{B}_{r,R}$. Therefore,

$$\sup_{\nu\in\mathscr{B}_{r,R}}\mathbb{E}\Big[e^{\lambda\kappa_1\int_0^\infty\frac{|X^\nu(s)-\bar{X}^0(s)|^2}{1+|X^\nu(s)-\bar{X}^0(s)|^2}\mathrm{d}s}\Big] < \infty, \quad \lambda > 0,$$

which implies (2b). □

Proof of (3). Assume (2.4). For any $\lambda > 0$, it suffices to find a compact set $K \subset \mathbb{R}^d$ such that (5.5) holds for $\bar{X}$, and

$$\sup_{\nu \in \mathscr{P}_2(\mathbb{R}^d)} \mathbb{E} e^{N \int_0^\infty \{1 \wedge |X^\nu(s) - \bar{X}^\nu(s)|^2\} ds} < \infty, \quad N \geq 1.$$

Indeed, by Theorem 5.3(2) and Theorem 5.5 with $\mathscr{I} = \mathscr{P}_2(\mathbb{R}^d)$ and $\Psi(\nu) = \nu$, this implies the upper LDP (LDP if $\bar{P}_t$ is strong Feller and $\bar{\mu}$-irreducible) for L_t^ν uniformly in $\nu \in \mathscr{P}_2(\mathbb{R}^d)$.

By (2.4), there exist constants $c_1, c_2 > 0$ such that

$$\mathrm{d}e^{|\bar{X}(t)|^2} \leq \big\{c_1 - c_2 |\bar{X}(t)|^{2+\varepsilon} e^{|\bar{X}(t)|^2}\big\} \mathrm{d}t + 2 e^{|\bar{X}(t)|^2} \langle \bar{X}(t), \sigma(\bar{\mu}) \mathrm{d}W(t) \rangle. \tag{4.8}$$

This implies

$$h_x(t) := \mathbb{E} e^{|\bar{X}^x(t)|^2} \leq c_1 t + e^{|x|^2} < \infty, \quad t \geq 0, x \in \mathbb{R}^d.$$

Moreover, by Jensen's inequality and the convexity of $[1, \infty) \ni r \mapsto r \log^{1+\varepsilon/2} r$, we deduce from (4.8) that

$$h_x(t) \leq h_x(0) + c_1 t - c_2 \int_0^t h_x(s) \log^{1+\varepsilon/2} h_x(s) \mathrm{d}s, \quad t \geq 0.$$

This and the comparison theorem imply $h_x(t) \leq \psi(t)$, where $\psi(t)$ solves the ODE

$$\psi'(t) = c_1 - c_2 \psi(t) \log^{1+\varepsilon/2} \psi(t), \quad \psi(0) = h_x(0) = e^{|x|^2}.$$

So,

$$\sup_{x \in \mathbb{R}^d} h_x(t) \leq \sup_{\psi(0) \geq 1} \psi(t) =: c(t) < \infty. \tag{4.9}$$

On the other hand, by (4.8), there exist constants $N_0, \beta > 0$ such that for any $N \geq N_0$ and $K = B_0(N)$, we have

$$\mathrm{d}e^{|\bar{X}^x(t)|^2} \leq -\beta |\bar{X}^x(t)|^{2+\varepsilon} e^{|\bar{X}^x(t)|^2} \mathrm{d}t + 2 e^{|\bar{X}^x(t)|^2} \langle \bar{X}^x(t), \sigma(\bar{\mu}) \mathrm{d}W(t) \rangle, \quad t \leq \tau_K^x. \tag{4.10}$$

Combining this with (4.6) and using the Markov property, when $2\delta \leq \beta^2$ we arrive at

$$\begin{aligned}
\mathbb{E}[e^{\delta N^2 \tau_K^x}] &\leq e^{\delta N^2} + \mathbb{E}\big[e^{\delta N^2 \tau_K^x} 1_{\{\tau_K^x \geq 1\}}\big] \\
&\leq e^{\delta N^2} + \mathbb{E}\big[e^{\delta N^2 (1 + \tau_K^{\bar{X}^x(1)})} 1_{\{\tau_K^x \geq 1\}}\big] \\
&\leq e^{\delta N^2} (1 + \mathbb{E} e^{|\bar{X}^x(1)|^2}) \leq e^{\delta N^2} (1 + c(1)) < \infty, \quad x \in \mathbb{R}^d, N \geq N_0.
\end{aligned}$$

Therefore, for any $\lambda > 0$, there exists compact set K such that (5.5) holds.

Finally, repeating the proof of (4.9) using $X^\nu(t)$ replacing $\bar{X}^x(t)$, we derive

$$\sup_{\nu \in \mathscr{P}_2(\mathbb{R}^d)} \mathbb{E}[e^{|X^\nu(1)|^2}] < \infty.$$

This together with (4.9) yields

$$\sup_{\nu \in \mathscr{P}_2(\mathbb{R}^d)} \mathbb{E}\big[e^{|X^\nu(1)|^2} + e^{|\bar{X}^\nu(1)|^2}\big] < \infty. \tag{4.11}$$

On the other hand, as in (4.7) but integrating from time 1, we obtain

$$\begin{aligned}
&\mathrm{e}^{-\frac{\lambda\kappa_2}{\kappa_1-\kappa_2}\mathbf{W}_2(\bar\mu,\nu)^2}\mathbb{E}\Big[\mathrm{e}^{\lambda\kappa_1\int_1^t\frac{|X^\nu(s)-\bar X^\nu(s)|^2}{1+|X^\nu(s)-\bar X^\nu(s)|^2}\mathrm{d}s}\Big]\\
&\le \mathbb{E}\bigg[(1+|X^\nu(1)-\bar X^\nu(1)|^2)^\lambda \mathrm{e}^{\lambda\int_1^t\frac{2}{1+|X^\nu(s)-\bar X^\nu(s)|^2}\langle X^\nu(s)-\bar X^\nu(s),\{\sigma(\nu_s)-\sigma(\bar\mu)\}\mathrm{d}W(s)\rangle}\bigg]\\
&\le \big\{\mathbb{E}\big[(1+|X^\nu(1)-\bar X^\nu(1)|^2)^{2\lambda}\big]\big\}^{\frac12}\Big(\mathbb{E}\Big[\mathrm{e}^{\lambda\kappa_1\mathbf{W}_2(P_1^*\nu,\bar\mu)^2\int_1^t\frac{|X^\nu(s)-\bar X^\nu(s)|^2\mathrm{e}^{-(\kappa_1-\kappa_2)s}}{1+|X^\nu(s)-\bar X^\nu(s)|^2}\mathrm{d}s}\Big]\Big)^{\frac12},\quad t>1.
\end{aligned}$$

Combining this with (4.11), we derive

$$\sup_{\nu\in\mathscr{P}_2(\mathbb{R}^d)}\mathbb{E}\mathrm{e}^{\lambda\kappa_1\int_1^\infty\frac{|X^\nu(s)-\bar X^\nu(s)|^2}{1+|X^\nu(s)-\bar X^\nu(s)|^2}\mathrm{d}s}<\infty,\quad \lambda\ge 1.$$

Therefore, the desired assertion holds. □

4.2 Proof of Theorem 2.2

As explained in the proof of Theorem 2.1(2)-(3) that, by Theorems 5.3 and 5.5, it suffices to prove the following assertions:

(a) For any $\lambda>0$, there exist a constant $s>0$ and compact set $K\subset\mathbb{R}^d$, such that (5.4) holds for any compact set $K'\subset\mathbb{H}$ and

$$\tau_K^x:=\inf\{t\ge 0:\bar X^x(t)\in K\},\quad x\in\mathbb{H}.$$

Moreover, for any $N\ge 1$,

$$\sup_{\nu\in\mathscr{B}_{r,R}}\mathbb{E}\mathrm{e}^{N\int_0^\infty\{1\wedge|X^\nu(s)-\bar X^0(s)|^2\}\mathrm{d}s}<\infty.$$

(b) For any $\lambda>0$, there exists a compact set $K\subset\mathbb{H}$ such that (5.5) holds for $\bar X$, and

$$\sup_{\nu\in\mathscr{P}_2(\mathbb{R}^d)}\mathbb{E}\mathrm{e}^{N\int_0^\infty\{1\wedge|X^\nu(s)-\bar X^\nu(s)|^2\}\mathrm{d}s}<\infty,\quad N\ge 1.$$

Comparing with the finite-dimensional case, the main difficulty is that bounded sets are no longer compact. To construct compact sets, let $\{e_i\}_{i\ge1}$ be the eigenbasis of A; i.e. it is an orthonromal basis of $\mathbb{H}$ such that $Ae_i=-\lambda_i e_i, i\ge 1$. For any $N>0$, the set

$$K:=B_{0,\gamma}(N)=\Big\{x\in\mathbb{H}:|x|_\gamma^2:=\sum_{i=1}^\infty\langle x,e_i\rangle^2\lambda_i^\gamma\le N^2\Big\}$$

is a compact set in $\mathbb{H}$.

Proof of (a). Simply denote $\bar X(t) = \bar X^x(t)$ and $\tau_K = \tau_K^x := \inf\{t \ge 0 : \bar X^x(t) \in K\}$. By (H_2) and (2.8), we may apply Itô's formula to

$$\psi(\bar X(t)) := \langle (-A)^{\gamma-1}\bar X(t), \bar X(t)\rangle = \sum_{i=1}^{\infty} \langle \bar X(t), e_i\rangle^2 \lambda_i^{\gamma-1},$$

such that for some constants $d_1, d_2 > 0$ and $\|b\|_{lip} \sum_{i=1}^{\infty} \lambda_i^{\gamma-1}\lambda_1^{\gamma} < 1$

$$\mathrm{d}\psi(\bar X(t)) \le (d_1 - d_2|\bar X(t)|_\gamma^2)\mathrm{d}t + \mathrm{d}M(t), \tag{4.12}$$

where $M(t) := 2\sum_{i=1}^{\infty} \lambda_i^{\gamma-1}\langle \bar X(t), e_i\rangle\langle \sigma(\bar\mu)\mathrm{d}W(t), e_i\rangle$ for an orthonormal basis $\{e_i\}_{i\ge 1}$ of $\mathbb{H}$. Let $N \ge N_0 := (2d_1/d_2)^{\frac{1}{2}}$, and consider τ_K for $K = B_{0,\gamma}(N)$. Then

$$d_1 - d_2|\bar X(t)|_\gamma^2 \le -d_1|\bar X(t)|_\gamma^2, \quad t \le \tau_K. \tag{4.13}$$

Since σ is bounded, by (H_2) there exists a constant $c > 0$ such that

$$\langle M\rangle(t) \le c\int_0^t |\bar X(s)|^2 \mathrm{d}s, \quad t \ge 0.$$

So, letting $\tau_n := \inf\{t \ge 0 : |\bar X(t)| \ge n\}$, we deduce form (4.12) and (4.13) that

$$\begin{aligned}\mathbb{E}\mathrm{e}^{\int_0^{t\wedge\tau_n\wedge\tau_K} \delta d_1|X(s)|_\gamma^2 \mathrm{d}s} &\le \mathrm{e}^{\delta\psi(x)}\big(\mathbb{E}\mathrm{e}^{2\delta^2\langle M\rangle(t\wedge\tau_n\wedge\tau_K)}\big)^{\frac{1}{2}}\\ &\le \mathrm{e}^{\delta\psi(x)}\big(\mathbb{E}\mathrm{e}^{2c\delta^2\int_0^{t\wedge\tau_n\wedge\tau_K} |\bar X(s)|_\gamma^2 \mathrm{d}s}\big)^{\frac{1}{2}} < \infty, \quad n \ge 1.\end{aligned}$$

Taking $\delta \le (2c)^{-1}$ leads to

$$\mathbb{E}\mathrm{e}^{\delta d_1 N^2(t\wedge\tau_n\wedge\tau_K)} \le \mathbb{E}\mathrm{e}^{\int_0^{t\wedge\tau_n\wedge\tau_K} \delta d_1|\bar X(s)|_\gamma^2 \mathrm{d}s} \le \mathrm{e}^{2\delta\psi(x)}, \quad t \ge 0, n \ge 1.$$

Letting $t, n \to \infty$ we derive

$$\mathbb{E}\mathrm{e}^{\delta N^2 d_1 \tau_K} \le \mathrm{e}^{2\delta\psi(x)}, \quad x \in \mathbb{H}.$$

Combining this with the Markov property, we obtain

$$\mathbb{E}\mathrm{e}^{\delta N^2 d_1 \tau_K^{\bar X(s)}} \le \mathbb{E}\mathrm{e}^{2\delta\psi(\bar X(s))},$$

and it is easy to see from (4.12) that the upper bound is locally bounded in x when δ is small enough. Therefore, condition (a) is satisfied, since $N \ge N_0$ is arbitrary.

Proof of (b). By (H_2) and Itô's formula, we have

$$\begin{aligned}\mathrm{d}|X^\nu(t) - \bar X^0(t)|^2 \le &\big\{-2(\lambda_1 - \alpha_1)|X^\nu(t) - \bar X^0(t)|^2 + 2\alpha_2 \mathbf{W}_2(P_t^*\nu, \bar\mu)^2\big\}\mathrm{d}t\\ &+ 2\langle X^\nu(t) - \bar X^0(t), \{\sigma(P_t^*\nu) - \sigma(\bar\mu)\}\mathrm{d}W(t)\rangle.\end{aligned}$$

The remainder of the proof is completely similar to that of the proof of Theorem 2.1(3).

4.3 Proof of Theorem 2.3

Let $\theta \in [0, \lambda_1]$ such that $\kappa_p = \theta - (\alpha_1 + \alpha_2)\mathrm{e}^{p\theta r_0}$.

4.3.1 Proof of Theorem 2.3(1)

For any $\nu_1, \nu_2 \in \mathscr{P}_p(\mathscr{C})$, take $X_0^{\nu_i} \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ such that $\mathscr{L}_{X_0^{\nu_i}} = \nu_i, i = 1, 2$, and

$$\mathbb{E}\big[\|X_0^{\nu_1} - X_0^{\nu_2}\|_\infty^p\big] = \mathbf{W}_p(\nu_1, \nu_2)^p. \tag{4.14}$$

Since σ is constant, we have

$$\mathrm{d}(X^{\nu_1}(t) - X^{\nu_2}(t)) = \big\{A(X^{\nu_1}(t) - X^{\nu_2}(t)) + b(X_t^{\nu_1}, P_t^*\nu_1) - b(X_t^{\nu_2}, P_t^*\nu_2)\big\}\mathrm{d}t, \quad t \geq 0.$$

By (H_3) and noting that $\theta \in [0, \lambda_1]$, we obtain

$$\begin{aligned}
&\mathrm{d}\big\{|X^{\nu_1}(t) - X^{\nu_2}(t)|^p\mathrm{e}^{p\theta t}\big\}\\
&= p\mathrm{e}^{p\theta t}|X^{\nu_1}(t) - X^{\nu_2}(t)|^{p-2}\big\{\langle X^{\nu_1}(t) - X^{\nu_2}(t), (\theta + A)(X^{\mu}(t) - X^{\nu_2}(t))\rangle\\
&\qquad\qquad + \langle X^{\nu_1}(t) - X^{\nu_2}(t), b(X_t^{\nu_1}, P_t^*\nu_1) - b(X_t^{\nu_2}, P_t^*\nu_2)\rangle\big\}\mathrm{d}t\\
&\leq p|X^{\nu_1}(t) - X^{\nu_2}(t)|^{p-1}\mathrm{e}^{p\theta t}\big\{\alpha_1\|X_t^{\nu_1} - X_t^{\nu_1}\|_\infty + \alpha_2\mathbf{W}_p(P_t^*\nu_1, P_t^*\nu_2)\big\}\mathrm{d}t, \quad t \geq 0.
\end{aligned}$$

Letting $\psi(t) = \|X_t^{\nu_1} - X_t^{\nu_2}\|_\infty^p\mathrm{e}^{p\theta t}$, we derive

$$\begin{aligned}
\psi(t) &\leq \mathrm{e}^{p\theta r_0}\sup_{s\in[(t-r_0),t]}|X^{\nu_1}(s) - X^{\nu_2}(s)|^p\mathrm{e}^{p\theta s}\\
&\leq \mathrm{e}^{p\theta r_0}\|X_0^{\nu_1} - X_0^{\nu_2}\|_\infty^p + p\mathrm{e}^{p\theta r_0}\int_0^t\big\{\alpha_1\psi(s) + \alpha_2\mathrm{e}^{\theta s}\mathbf{W}_p(P_s^*\nu_1, P_s^*\nu_2)\psi(s)^{\frac{p-1}{p}}\big\}\mathrm{d}s.
\end{aligned} \tag{4.15}$$

Combining this with (4.14) and

$$\mathbf{W}_p(P_t^*\nu_1, P_t^*\nu_2)^p \leq \mathbb{E}\|X_t^{\nu_1} - X_t^{\nu_2}\|_\infty^p, \quad t \geq 0,$$

we arrive at

$$\mathbb{E}[\psi(t)] \leq \mathrm{e}^{p\theta r_0}\mathbf{W}_p(\nu_1, \nu_2)^p + p\mathrm{e}^{p\theta r_0}(\alpha_1 + \alpha_2)\int_0^t\mathbb{E}[\psi(s)]\mathrm{d}s, \quad t \geq 0.$$

By Theorem 3.1 we have $\mathbb{E}[\psi(t)] < \infty, t > 0$. Then Gronwall's lemma yields

$$\mathbb{E}[\psi(t)] \leq \{\mathbf{W}_p(\nu_1, \nu_2)\}^p\mathrm{e}^{p\theta r_0 + p(\alpha_1+\alpha_2)\mathrm{e}^{p\theta r_0}t}, \quad t \geq 0.$$

Therefore,

$$\mathbf{W}_p(P_t^*\nu_1, P_t^*\nu_2)^p \leq \mathrm{e}^{-p\theta t}\mathbb{E}[\psi(t)] \leq \{\mathbf{W}_p(\nu_1, \nu_2)\}^p\mathrm{e}^{p\theta r_0 - p\kappa_p t}, \quad t \geq 0.$$

When $\kappa_p > 0$, it is standard that (2.12) implies the existence and uniqueness of P_t^*-invariant probability measure μ such that (2.13) holds, see, for instance, [20, Proof of Theorem 3.1(2)].

4.3.2 Proof of Theorem 2.3(2)

Let $\kappa_p > 0$. To prove the LDP, let $\bar{P}_t$ be the Markov semigroup for the stationary equation (2.11) and consider the LDP for $\bar{L}_t^\nu$. Since $\lambda > 0$ implies $\sup_{r\in[0,\lambda_1]}(r - \alpha_1 \mathrm{e}^{r r_0}) > 0$ and noting that (H_3) implies

$$|b(\xi,\mu) - b(\eta,\mu)| \le \alpha_1 \|\xi - \eta\|_\infty,$$

by [1, Theorem 1.2] and $\kappa_1 \ge \kappa_p > 0$, the Markov semigroup $\bar{P}_t$ is hypercontractive. Thus, by the semigroup property and the interpolation theorem, for any $q > 1$ there exists $t_q > 0$ such that $\bar{P}_{t_q}$ is uniformly integrable in $L^q(\mu)$. Moreover, according to [18, Theorem 4.2.4], assumption (H_3) implies that for any $t > r_0$, there exists a constant $c > 0$ such that the following Harnack inequality holds:

$$\big(\bar{P}_{t_0} f(\eta)\big)^2 \le (\bar{P}_{t_0} f^2(\xi))\mathrm{e}^{c\|\xi-\eta\|_\infty^2}, \quad \xi,\eta \in \mathscr{C}, f \in \mathscr{B}_b(\mathscr{C}). \tag{4.16}$$

Since $P_t^* \bar{\mu} := \mathscr{L}_{X_t^{\bar{\mu}}} = \bar{\mu}$, $X_t^{\bar{\mu}}$ solves the the stationary equation (2.11), so that by the uniqueness conclude that $\bar{\mu}$ is also $\bar{P}_t$-invariant. Then for any $B \in \mathscr{B}(\mathscr{C})$ such that $\bar{\mu}(B) > 0$, we have $\bar{\mu}(\bar{P}_{t_0} 1_B) = \mu(B) > 0$, and hence $\bar{P}_{t_0} 1_B(\eta) > 0$ holds for some $\eta \in \mathscr{C}$. Then (4.16) implies $\bar{P}_{t_0} 1_B(\xi) > 0$ for all $\xi \in \mathscr{C}$, so that $\bar{\mu}(1_A \bar{P}_{t_0} 1_B) > 0$ for $\bar{\mu}(A), \bar{\mu}(B) > 0$, i.e. $\bar{P}_{t_0}$ is $\bar{\mu}$-irreducible. Therefore, by Theorem 5.2,

$$\bar{L}_t^\nu \in LDP(J) \ \text{ uniformly in } \nu = h\bar{\mu} \in \mathscr{P}(\mathscr{C}) \text{ with } \|h\|_{L^q(\bar{\mu})} \le R, \quad R > 0. \tag{4.17}$$

Combining this with Lemma 5.4, it remains to show that for any $\varepsilon, R > 0$,

(I) $\{\bar{L}_t^\nu\}_{\nu \in \mathscr{I}_{\varepsilon,R}} \in LDP(J)$;

(II) For any $\delta > 0$,

$$\lim_{t\to\infty} \frac{1}{t} \sup_{\nu \in \mathscr{I}_{\varepsilon,R}} \log \mathbb{P}\bigg(\frac{1}{t}\int_0^t \{1 \wedge \|X_s^\nu - \bar{X}_s^\nu\|_\infty\}\mathrm{d}s > \delta\bigg) = -\infty.$$

For (I). Observing that for any $\xi, \eta \in \mathscr{C}$ we have

$$\mathrm{d}(\bar{X}^\xi(t) - \bar{X}^\eta(t)) = \big\{A(\bar{X}^\xi(t) - \bar{X}^\eta(t)) + b(\bar{X}_t^\xi, \bar{\mu}) - b(\bar{X}_t^\eta, \bar{\mu})\big\}\mathrm{d}t,$$

by the same reason leading to (4.15) we obatin

$$\|\bar{X}_t^\xi - \bar{X}_t^\eta\|_\infty^p \mathrm{e}^{p\theta t} \le \mathrm{e}^{p\theta r_0}\|\xi - \eta\|_\infty^p + \alpha_1 p \mathrm{e}^{p\theta r_0} \|\bar{X}_s^\xi - \bar{X}_s^\eta\|_\infty^p \mathrm{e}^{p\theta s}\mathrm{d}s, \quad t \ge 0.$$

Noting that $\kappa_p \le \theta - \alpha_1 \mathrm{e}^{p\theta r_0}$, by Gronwall's inequality we get

$$\|\bar{X}_t^\xi - \bar{X}_t^\eta\|_\infty^p \le \mathrm{e}^{p\theta r_0 - p\{\theta - \alpha_1 \mathrm{e}^{p\theta r_0}\}t}\|\xi - \eta\|_\infty^p \le e^{p\theta r_0 - p\kappa_p\}t}\|\xi - \eta\|_\infty^p.$$

Combining this with (4.16) and using the semigroup property of $\bar{P}_t$, we find a constant $t_1 > t_0$ such that

$$\big(\bar{P}_{t_1} f(\eta)\big)^2 \le (\bar{P}_{t_1} f^2(\xi))\mathrm{e}^{\varepsilon\|\xi-\eta\|_\infty^2/2}, \quad \xi,\eta \in \mathscr{C}, f \in \mathscr{B}_b(\mathscr{C}).$$

This implies that the invariant probability measure $\bar\mu$ has full support on $\mathscr{C}$, so that there exists a constant $c>0$ such that

$$\sup_{\bar\mu(|f|^2)\le 1}(\bar P_{t_1}f(\xi))^2\le \frac{1}{\int_{\mathscr{C}}\mathrm{e}^{-\varepsilon\|\xi-\eta\|_\infty^2/2}\bar\mu(\mathrm{d}\eta)}\le c\mathrm{e}^{\varepsilon\|\xi\|_\infty^2},\quad \xi\in\mathscr{C}.$$

Therefore, $\bar P_{t_1}$ has a density $p_{t_1}(\xi,\eta)$ with respect to $\bar\mu$ satisfying

$$\int_{\mathscr{C}}p_{t_1}(\xi,\eta)^2\bar\mu(\mathrm{d}\eta)\le c\mathrm{e}^{\varepsilon\|\xi\|_\infty^2},\quad \xi\in\mathscr{C}.$$

Consequently, for any $\nu\in\mathscr{I}_{\varepsilon,R}$, $\bar\nu_{t_1}:=\mathscr{L}_{\bar X_{t_1}^\nu}$ has density

$$h(\eta):=\int_{\mathscr{C}}p_{t_1}(\xi,\eta)\nu(\mathrm{d}\xi)$$

with respect to $\bar\mu$ which satisfies

$$\bar\mu(|h|^2)\le\int_{\mathscr{C}\times\mathscr{C}}p_{t_1}(\xi,\eta)^2\nu(\mathrm{d}\xi)\bar\mu(\mathrm{d}\eta)\le c\nu(\mathrm{e}^{\varepsilon\|\cdot\|_\infty^2})\le cR.$$

Combining this with (4.17) and noting that the Markov property of $\bar X_t$ implies that the law of $\bar L_t^{\bar\nu_{t_1}}$ coincides with that of

$$\tilde L_t^\nu:=\frac{1}{t}\int_{t_1}^{t+t_1}\delta_{\bar X_s^\nu}\mathrm{d}s,$$

we prove

$$\{\tilde L_t^\nu\}_{\nu\in\mathscr{I}_{\varepsilon,R}}\in LDP(J). \tag{4.18}$$

On the other hand, consider the probability distance

$$\rho(\Lambda_1,\Lambda_2):=\inf_{\Pi\in\mathscr{C}(\Lambda_1,\Lambda_2)}\int_{\mathscr{C}\times\mathscr{C}}\big\{\|\xi-\eta\|_\infty\wedge 1\big\}\Pi(\mathrm{d}\xi,\mathrm{d}\eta) \tag{4.19}$$

on $\mathscr{P}(\mathscr{C})$. Then

$$\rho(\tilde L_t^\nu,\bar L_t^\nu)\le\frac{t_1}{t},\quad t>0.$$

So, by Lemma 5.4 and (4.18) we prove (I).

For (II). By (H_3) and (2.13), there exist constants $c>0$ such that for $\theta\in[0,\lambda_1]$,

$$\begin{aligned}\|X_t^\nu-\bar X_t^\nu\|_\infty\mathrm{e}^{\theta t}&\le \mathrm{e}^{\theta r_0}\sup_{s\in[(t-r_0)^+,t]}|X^\nu(s)-\bar X^\nu(s)|\mathrm{e}^{\theta s}\\ &\le \mathrm{e}^{\theta r_0}\int_0^t\mathrm{e}^{\theta s}\big\{\alpha_1\|X_s^\nu-\bar X_s^\nu\|_\infty+\alpha_2\mathbf{W}_p(P_s^*\nu,\mu)\big\}\mathrm{d}s\\ &\le c+\alpha_1\mathrm{e}^{\theta r_0}\int_0^t\mathrm{e}^{\theta s}\|X_s^\nu-\bar X_s^\nu\|_\infty\mathrm{d}s,\quad t\ge 0,\nu\in\mathscr{I}_{\varepsilon,R}.\end{aligned}$$

By Gronwall's inequality we obtain

$$\sup_{\nu\in\mathscr{I}_{\varepsilon,R}}\|X_t^\nu-\bar X_t^\nu\|_\infty\le c\exp\big[\{\alpha_1\mathrm{e}^{\theta r_0}-\theta\}t\big]\le c\mathrm{e}^{-\kappa_p t},\quad t>0.$$

This proves assertion (II).

4.4 Proof of Theorem 2.4

By (2.16), we take $\theta \in (0, \lambda_1]$ such that

$$\theta \mathrm{e}^{-\theta r_0} - K_2 - \alpha' \|B\| > K_3. \tag{4.20}$$

For any $\alpha > 0$, let

$$\rho_\alpha(\xi_1, \xi_2) := \alpha \|\xi_1^{(1)} - \xi_2^{(1)}\|_\infty + \|\xi_1^{(2)} - \xi_2^{(2)}\|_\infty, \quad \xi_1, \xi_2 \in \mathscr{C}.$$

We take $X_0, Y_0 \in L^2(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ such that $\mathscr{L}_{X_0} = \nu_1, \mathscr{L}_{Y_0} = \nu_2$ and

$$\mathbf{W}_{p,\alpha}(\nu_1, \nu_2)^p = \mathbb{E}\rho_\alpha(X_0, Y_0)^p. \tag{4.21}$$

Let $X(t)$ and $Y(t)$ solves (2.14) with initial values X_0 and Y_0 respectively. Then (H_4^1) implies $A_1 - \delta \le -\lambda_1 \le \theta$, so that

$$\begin{aligned} |X^{(1)}(t) - Y^{(1)}(t)| &\le |\mathrm{e}^{(A_1-\delta)t}\{X^{(1)}(0) - Y^{(1)}(0)\}| \\ &\quad + \int_0^t \big|\mathrm{e}^{(A_1-\delta)(t-s)}\{\delta(X^{(1)}(s) - Y^{(1)}(s)) + B(X^{(2)}(s) - Y^{(2)}(s))\}\big|\mathrm{d}s \\ &\le \mathrm{e}^{-\theta t}|X^{(1)}(0) - Y^{(1)}(0)| + \int_0^t \mathrm{e}^{-\theta(t-s)}\big\{\delta|X^{(1)}(s) - Y^{(1)}(s)| + \|B\| \cdot |X^{(2)}(s) - Y^{(2)}(s)|\big\}\mathrm{d}s. \end{aligned}$$

Equivalently,

$$\begin{aligned} \mathrm{e}^{\theta t}|X^{(1)}(t) - Y^{(1)}(t)| &\le |X^{(1)}(0) - Y^{(1)}(0)| \\ &\quad + \int_0^t \mathrm{e}^{\theta s}\big\{\delta|X^{(1)}(s) - Y^{(1)}(s)| + \|B\| \cdot |X^{(2)}(s) - Y^{(2)}(s)|\big\}\mathrm{d}s. \end{aligned}$$

Similarly, it follows from $A_2 \le -\lambda_1 \le -\theta$ and (H_4^2) that

$$\begin{aligned} \mathrm{e}^{\theta t}|X^{(2)}(t) - Y^{(2)}(t)| &\le |X^{(2)}(0) - Y^{(2)}(0)| \\ &\quad + \int_0^t \mathrm{e}^{\theta s}\big\{K_1\|X_s^{(1)} - Y_s^{(1)}\|_\infty + K_2\|X_s^{(2)} - Y_s^{(2)}\|_\infty + K_3\mathbf{W}_{p,\alpha}(P_s^*\nu_1, P_s^*\nu_2)\big\}\mathrm{d}s. \end{aligned}$$

Combining these with $\alpha' \ge \alpha$ and that $\lambda' := \frac{1}{2}\{\delta + K_2 + \sqrt{(K_2 - \delta)^2 + 4\|B\|}$ satisfies

$$\alpha'\delta + K_1 = \lambda'\alpha', \quad \alpha'\|B\| + K_2 = \lambda' > 0,$$

we derive

$$\begin{aligned} \mathrm{e}^{\theta t}\rho_{\alpha'}(X_t, Y_t) &\le \mathrm{e}^{\theta r_0} \sup_{s \in [t-r_0, t]} \{\alpha'|X^{(1)}(s) - Y^{(1)}(s)| + |X^{(2)}(s) - Y^{(2)}(s)|\}\mathrm{e}^{\theta s} \\ &\le \mathrm{e}^{\theta r_0}\rho_{\alpha'}(X_0, Y_0) + \mathrm{e}^{\theta r_0}\int_0^t \big\{(\delta\alpha' + K_1)\|X_s^{(1)} - Y_s^{(1)}\|_\infty \\ &\qquad + (\alpha'\|B\| + K_2)\|X_s^{(2)} - Y_s^{(2)}\|_\infty + K_3\mathbf{W}_{p,\alpha}(P_s^*\nu_1, P_s^*\nu_2)\big\}\mathrm{d}s \end{aligned}$$

$$= \mathrm{e}^{\theta r_0}\rho_{\alpha'}(X_0,Y_0) + \mathrm{e}^{\theta r_0}\int_0^t \mathrm{e}^{\theta s}\big\{\lambda'\rho_{\alpha'}(X_s,Y_s) + K_3\mathbb{E}[\rho_\alpha(X_s,Y_s)]\big\}\mathrm{d}s.$$

By Gronwall's lemma, for $\kappa := \theta - \lambda'\mathrm{e}^{\theta r_0} > 0$ we have

$$\rho_{\alpha'}(X_t,Y_t) \le \mathrm{e}^{\theta r_0 - \kappa t}\rho_{\alpha'}(X_0,Y_0) + 2\mathrm{e}^{\theta r_0}K_3\int_0^t \mathrm{e}^{-\kappa(t-s)}\mathbb{E}[\rho_\alpha(X_s,Y_s)]\mathrm{d}s.$$

Therefore, for any $\varepsilon > 0$ there exists a constant $C(\varepsilon) > 0$ such that

$$\rho_{\alpha'}(X_t,Y_t)^p \le C(\varepsilon)\rho_{\alpha'}(X_0,Y_0)^p\mathrm{e}^{-\kappa p t} + \frac{(2K_3)^p\mathrm{e}^{\theta r_0 p}(1+\varepsilon)}{\kappa^{p-1}}\int_0^t \mathrm{e}^{-\kappa(t-s)}\mathbb{E}[\rho_\alpha(X_s,Y_s)^p]\mathrm{d}s.$$

Combining this with $\rho_{\alpha'} \ge \rho_\alpha$ and $\mathbb{E}[\rho_{\alpha'}(X_t,Y_t)^p] < \infty$ due to Theorem 3.1, we deduce from this and Gronwall's lemma that

$$\begin{aligned}\mathbf{W}_{p,\alpha}(P_t^*\nu_1,P_t^*\nu_2)^p &\le \mathbb{E}[\rho_{\alpha'}(X_t,Y_t)^p]\\ &\le \frac{\alpha^p C(\varepsilon)}{(\alpha')^p}\mathbf{W}_{p,\alpha}(\nu_1,\nu_2)^p \exp\big[-\big(\kappa - (1+\varepsilon)(2K_3)^p\mathrm{e}^{\theta r_0 p}\kappa^{1-p}\big)t\big].\end{aligned}$$

It is easy to see that (4.20) implies $\kappa > K_3^p\mathrm{e}^{\theta r_0 p}\kappa^{1-p}$, so that by taking small enough $\varepsilon > 0$ we prove

$$\mathbf{W}_p(P_t^*\nu_1,P_t^*\nu_2) \le c_1\mathrm{e}^{-c_2 t},\quad t \ge 0, \nu_1,\nu_2 \in \mathscr{P}_p(\mathscr{C})$$

for some constants $c_1, c_2 > 0$. Consequently, P_t^* has a unique invariant probability measure $\bar\mu$ such that (2.17) holds.

Similarly, by (H_4) and (2.17), we find a constant $C > 0$ such that for any $X_0^\nu = \bar X_0^\nu \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$,

$$\int_0^\infty \|X_t^\nu - \bar X_t^\nu\|_\infty^2 \mathrm{d}t \le C,\quad \nu \in \mathscr{I}_{R,q}.$$

Moreover, it is easy to see that (2.16) implies the condition in [1, Theorem 1.3] for the reference equation with $\bar\mu$ replacing the distribution of solution, so that $\bar P_t$ is hypecontractive (hence uniformly integrable in $L^p(\bar\mu)$ for any $p > 1$) for large $t > 0$, and the Harnack in [1, Lemma 4.1] implies (4.16). Then the desired LDP can be proved in the same way as in the proof of Theorem 2.3.

5 Appendix: LDP for Markov processes

We first introduce the rate function, i.e. the Donsker-Varadhan level 2 entropy function for continuous Markov processes on a Polish space E.

Consider the path space

$$\mathbf{C}_E := C([0,\infty) \to E) = \{w : [0,\infty) \ni t \mapsto w(t) \in E \text{ is continuous}\}.$$

Let $\mathscr{P}(\mathbf{C}_E)$ be the set of all probability measures on $\mathbf{C}_E$, and $\mathscr{P}^s(\mathbf{C}_E)$ the set of all stationary (i.e. time-shift-invariant) elements in $\mathscr{P}(\mathbf{C}_E)$. For any $Q \in \mathscr{P}^s(\mathbf{C}_E)$, let $\bar Q$ be the unique stationary probability measure on $\bar{\mathbf{C}}_E := C(\mathbb{R} \to E)$ such that

$$\bar Q\big(\{w \in \bar{\mathbf{C}}_E : w(t_i) \in A_i, 1 \le i \le n\}\big) = Q\big(\{w \in \mathbf{C}_E : w(t_i + s) \in A_i, 1 \le i \le n\}\big)$$

holds for any $n \geq 1, -\infty < t_1 < t_2 < \cdots < t_n < \infty, s \geq -t_1$, and $\{A_i\}_{1\leq i\leq n} \subset \mathscr{B}(E)$. We call $\bar{Q}$ the stationary extension of Q to $\bar{\mathbf{C}}_E$. For any $s \leq t$, let $\mathscr{F}_t^s := \sigma(\bar{\mathbf{C}}_E \ni w \mapsto w(u) : s \leq u \leq t)$. For a probability measure $\bar{Q}$ on $\bar{\mathbf{C}}_E$, let $\bar{Q}_{w-}$ be the regular conditional distribution of $\bar{Q}$ given $\mathscr{F}_0^{-\infty}$. Moreover, let $\mathrm{Ent}_{\mathscr{F}_1^0}$ be the Kullback-Leibler divergence (i.e. relative entropy) on the σ-field $\mathscr{F}_1^0$; that is, for any two probability measures μ_1, μ_2 on $\mathbf{C}_E$,

$$\mathrm{Ent}_{\mathscr{F}_1^0}(\mu_1|\mu_2) := \begin{cases} \int_{\mathbf{C}_E} \big(h \log h\big)\mathrm{d}\mu_2, & \text{if } \mathrm{d}\mu_1|_{\mathscr{F}_1^0} = h\mathrm{d}\mu_2|_{\mathscr{F}_1^0}, \\ \infty, & \text{otherwise.} \end{cases}$$

Now, for a standard Markov process on E with $\{P^x : x \in E\} \subset \mathscr{P}(\mathbf{C}_E)$, where P^x stands for the distribution of the process starting at x, the process level entropy function of Donsker-Varadhan is given by

$$H(Q) := \begin{cases} \int_{\bar{\mathbf{C}}_E} \mathrm{Ent}_{\mathscr{F}_1^0}(\bar{Q}_{w-}|P^{w(0)})\bar{Q}(\mathrm{d}w), & \text{if } Q \in \mathscr{P}^s(\mathbf{C}_E), \\ \infty, & \text{otherwise.} \end{cases}$$

Then the Donsker-Varadhan level 2 entropy function is defined as

$$J(\nu) := \inf\big\{H(Q) : \ Q \in \mathscr{P}^s(\mathbf{C}_E), Q(w(0) \in \cdot) = \nu\big\}, \quad \nu \in \mathscr{P}(E). \tag{5.1}$$

This function has compact level sets in $\mathscr{P}(E)$ under the τ- (hence the weak) topology, see for instance [22, 23]. For any $\nu \in \mathscr{P}(E)$, let $(X_t^\nu)_{t\geq 0}$ be the Markov process with initial distribution ν. Consider its empirical measure

$$L_t^\nu := \frac{1}{t}\int_0^t \delta_{X_s^\nu}\mathrm{d}s, \quad t > 0.$$

When $\nu = \delta_x$, we denote $X_t^\nu = X_t^x$ and $L_t^\nu = L_t^x$. Let μ be an invariant probability measure of P_t, where P_t is the Markov semigroup given by

$$P_t f(x) = \mathbb{E}[f(X_t^x)], \quad x \in E, t \geq 0, f \in \mathscr{B}_b(E).$$

We write $f \in \mathscr{D}_\mu(\mathscr{A})$ if $f \in L^\infty(\mu)$ and there exists $g \in L^\infty(\mu)$ such that $P_t f - f = \int_0^t P_s g \mathrm{d}s$ holds μ-a.e. for all $t \geq 0$. In this case, we denote $\mathscr{A} f = g$. We have the following formula for J.

Theorem 5.1 ([23], Proposition B.10 and Corollary B.11)**.** *Assume that P_t has a unique invariant probability measure μ. Then*

$$J(\nu) = \begin{cases} \sup\big\{ \int_E \frac{-\mathscr{A} f}{f}\mathrm{d}\nu : \ 1 \leq f \in \mathscr{D}_\mu(\mathscr{A})\big\}, & \textit{if } \nu\lambda\mu, \\ \infty, & \textit{otherwise.} \end{cases} \tag{5.2}$$

In particular, if the Markov process is associated with a symmetric Dirichlet form $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ in $L^2(\mu)$, then

$$J(\nu) = \begin{cases} \mathscr{E}(h^{\frac{1}{2}}, h^{\frac{1}{2}}), & \textit{if } \nu = h\mu, h^{\frac{1}{2}} \in \mathscr{D}(\mathscr{E}), \\ \infty, & \textit{otherwise.} \end{cases} \tag{5.3}$$

We now recall another result due to [23] on the LDP for uniformly integrable Markov semigroups, which will be used in the proof of Theorem 2.3. Let $p \geq 1$ and let P be a bounded linear operator on $L^p(\mu)$. We call P uniformly integrable in $L^p(\mu)$ if

$$\lim_{R\to\infty} \sup_{\mu(|f|^p)\leq 1} \mu(|Pf|^p 1_{\{|Pf|>R\}}) = 0.$$

This LDP is established under the τ-topology induced by $f \in \mathscr{B}_b(E)$, and hence also holds under the weak topology. Let $\nu \in I_{q,L} := \{\nu = h\mu : \|h\|_{L^q(\mu)} \leq L\}$ for $q, L \in (1,\infty)$.

Theorem 5.2 ([23], Theorem 5.1)**.** *Assume that the Markov semigroup P_t has a unique invariant probability measure μ, and there exists $T \in (1,\infty)$ and $p \in (1,\infty)$ such that P_T is μ-irreducible and uniformly integrable in $L^p(\mu)$. Then $\{L_t^\nu\}_{\nu\in I_{q,L}} \in LDP(J)$ under the τ-topology for all $q, L \in (1,\infty)$.*

The next result due to [22] provides criteria on the LDP using the hitting time to compact sets, which will be used in the proofs of Theorem 2.1 and Theorem 2.2. For any set $K \subset E$ and any $x \in E$, let

$$\tau_K^x := \inf\{t \geq 0 : X^x(t) \in K\},$$

where $X^x(t)$ is the Markov process starting at x. We will use the following conditions where **(D1)** is weaker than **(D2)**:

(D1) For any $\lambda > 0$ there exist a constant $s > 0$ and a compact set $K \subset E$ such that for any compact set $K' \subset E$,

$$\sup_{x\in K} \mathbb{E}[\mathrm{e}^{\lambda \tau_K^{X^x(s)}}] < \infty, \quad \sup_{x\in K'} \mathbb{E}[\mathrm{e}^{\lambda \tau_K^x}] < \infty. \tag{5.4}$$

(D2) For any $\lambda > 0$ there exists a compact set $K \subset E$ such that

$$\sup_{x\in E} \mathbb{E}[\mathrm{e}^{\lambda \tau_K^x}] < \infty. \tag{5.5}$$

Theorem 5.3 ([22], Theorems 1.1,1.2)**.** *Assume that P_t is a Feller Markov semigroup.*

(2) **(D1)** *implies $\{L_t^x\}_{x\in D} \in LDP_u(J)$ for any compact set $D \subset E$, and the inverse holds provided E is locally compact. If P_t is strong Feller and μ-irreducible for some $t > 0$, then $\{L_t^x\}_{x\in D} \in LDP(J)$ for compact $D \subset E$ if and only if* **(D1)** *holds.*

(1) **(D2)** *implies $\{L_t^\nu\}_{\nu\in\mathscr{P}(E)} \in LDP_u(J)$, and the inverse holds when E is locally compact. If moreover P_t is strong Feller and μ-irreducible for some $t > 0$, then $\{L_t^\nu\}_{\nu\in\mathscr{P}(E)} \in LDP(J)$ if and only if* **(D2)** *holds.*

Moreover, we introduce the following approximation lemma which is easy to prove but useful in applications, see for instance [5, Theorems 4.2.16, 4.2.23], and see also [16, Theorem 3.2] for a stronger version called generalized contraction principle.

Lemma 5.4 (Approximation Lemma for LDP). *Let $\{(L_t^\nu)_{t>0}, (\bar L_t^\nu)_{t>0} : \nu \in \mathscr{I}\}$ be two families of stochastic processes on a Polish space (E,ρ) for an index set $\mathscr{I}$. If $(\bar L_t^\nu)_{\nu\in\mathscr{I}} \in LDP_u(J)$(respectively $LDP_l(J)$) and*

$$\lim_{t\to\infty} \frac{1}{\lambda(t)} \sup_{\nu\in\mathscr{I}} \log \mathbb{P}(\rho(L_t^\nu, \bar L_t^\nu) > \delta) = -\infty, \quad \delta > 0,$$

then $(L_t^\nu)_{\nu\in\mathscr{I}} \in LDP_u(J)$(respectively $LDP_l(J)$).

Finally, to establish the LDP for L_t^ν associated with (3.1), we consider a reference equation:

$$\mathrm{d}\bar X^\nu(t) = \big\{A\bar X^\nu(t) + \bar b(\bar X_t^\nu)\big\}\mathrm{d}t + \bar\sigma(\bar X_t^\nu)\mathrm{d}W(t), \quad \bar X_0^\nu = X_0^\nu, \tag{5.6}$$

where $\bar b : \mathscr{C} \to \mathbb{H}$, $\bar\sigma : \mathscr{C} \to \mathbb{L}(\mathbb{H})$ are measurable such that this equation has a unique mild segment solution for any initial value in $\mathscr{C}$, which is thus a Markov process on $\mathscr{C}$. In applications, the coefficients in (5.6) will be given by the limit of $b_t(\cdot,\nu_t)$ and $\sigma_t(\cdot,\nu_t)$ as $t\to\infty$, where b_t and σ_t are in (3.1) and $\nu_t := \mathscr{L}_{X_t^\nu}$. Now, let

$$\bar L_t^\nu = \frac{1}{t}\int_0^t \delta_{\bar X_s^\nu}\mathrm{d}s, \quad t > 0.$$

We have the following result which is more or less standard but we include a brief proof for complement.

Theorem 5.5. *Assume that* (3.1) *and* (5.6) *are well-posed for any initial value X_0 with $\mathscr{L}_{X_0} \in \mathscr{I}$ and $\mathscr{L}_{\bar X_0} \in \Psi(\mathscr{I})$ respectively, where $\mathscr{I}$ is a non-empty subset of $\mathscr{P}(\mathscr{C})$ and $\Psi : \mathscr{I} \to \mathscr{P}(\mathscr{C})$ is a map. If $\{\bar L_t^\nu\}_{\nu\in\Psi(\mathscr{I})} \in LDP_u(J)$(respectively $LDP_l(J)$) under the weak topology, and*

$$\sup_{\nu\in\mathscr{I}} \mathbb{E}\big[\mathrm{e}^{N\int_0^\infty \{\|X_s^\nu - \bar X_s^{\Psi(\nu)}\|_\infty \wedge 1\}\mathrm{d}s}\big] < \infty, \quad N \geq 1, \tag{5.7}$$

then $\{L_t^\nu\}_{\nu\in\mathscr{I}} \in LDP_u(J)$(respectively $LDP_l(J)$) under the weak topology.

Proof. Let ρ be in (4.19). It is well known that ρ induces the weak topology on $\mathscr{P}(\mathscr{C})$. Since

$$\rho(L_t^\nu, \bar L_t^{\bar\nu}) \leq \frac{1}{t}\int_0^t \big\{\|X_s^\nu - \bar X_s^{\Psi(\nu)}\|_\infty \wedge 1\big\}\mathrm{d}s, \quad t > 0,$$

(5.7) implies

$$\begin{aligned} &\lim_{t\to\infty} \frac{1}{t}\sup_{\nu\in\mathscr{I}} \log \mathbb{P}(\rho(L_t^\nu, \bar L_t^{\Psi(\nu)}) > \delta) \\ &\leq \lim_{t\to\infty} \frac{1}{t}\sup_{\nu\in\mathscr{I}} \log \mathbb{P}\Big(N\int_0^t \big\{\|X_s^\nu - \bar X_s^{\Psi(\nu)}\|_\infty \wedge 1\big\}\mathrm{d}s > tN\delta\Big) \\ &\leq -N\delta, \quad N \geq 1, \delta > 0. \end{aligned}$$

Therefore,

$$\lim_{t\to\infty} \frac{1}{t}\sup_{\nu\in\mathscr{I}} \log \mathbb{P}(\rho(L_t^\nu, \bar L_t^{\Psi(\nu)}) > \delta) = -\infty, \quad \delta > 0.$$

Then the desired assertion follows from Lemma 5.4 with $\bar L_t^{\Psi(\nu)}$ replacing $\bar L_t^\nu$. □

Acknowledgement. We would like to thank the reviewers and editor for helpful comments and corrections.

References

[1] J. Bao, F.-Y. Wang, C. Yuan, *Hypercontractivity for functional stochastic partial differential equations,* Comm. Electr. Probab. 20(2015), 1–15.

[2] J. Bao, F.-Y. Wang, C. Yuan, *Limit theorems for additive functional of path-distribution depedent SDEs,* Disc. Cont. Dyn. Sys.-A 40(2020), 5173–5188.

[3] V. Bogachev, N.V. Krylov, M. Röckner, *On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions,* Comm. Part. Diff. Equat. 26(2001), 2037–2080.

[4] Da Prato, G., Zabczyk, J., *Stochastic Equations in Infinite Dimensions,* Cambridge University Press, Cambridge, 1992.

[5] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications,* Second Edition, Springer, New York. 1998.

[6] M. D. Donsker and S. R. S. Varadhan, *Asymptotic evaluation of certain Markov process expectations for large time,* I-IV, Comm. Pure Appl. Math. 28(1975), 1–47, 279–301; 29(1976), 389–461; 36(1983), 183–212.

[7] M. I. Freidlin, A. D. Wentzell, *Random Perturbation of Dynamical Systems,* Translated by J. Szuc, Springer, 1984.

[8] A. Guillin, F.-Y. Wang, *Degenerate Fokker-Planck equations: Bismut formula, gradient estimate and Harnack inequality,* J. Diff. Equat. 253(2012), 20–40.

[9] M. Hairer, *On Malliavin's proof of Hörmander's theorem,* Bull. Sci. Math. 135(2011), 650–666.

[10] X. Huang, M. Röckner, F.-Y. Wang, *Nonlinear Fokker–Planck equations for probability measures on path space and path-distribution dependent SDEs,* Discrete Contin. Dyn. Syst. A. 39(2019), 3017–3035.

[11] M. Kac, *Foundations of kinetic theory,* In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III. University of California Press, Berkeley and Los Angeles, 1956, pp. 171–197.

[12] A. Kulik, *Ergodic behavior of Markov processes with applications to limit theorems*, De Gruyter Studies in Mathematics, 67, De Gruyter, Berlin, 2018.

[13] E. Lanconelli, S. Polidoro, *On a class of hypoelliptic evolution operator,* Rend. Sem. Mat. Univ. Pol. Torino 52(1994), 29–63.

[14] H. P. McKean, *A class of Markov processes associated with nonlinear parabolic equations,* In: Proceedings of the National Academy of Sciences of the United States of America 56.6(1966), p. 1907.

[15] G. D. Reis, W. Salkeld, J. Tugaut, *Freidlin-Wentzell LDP in path space for McKean-Vlasov equations and the functional iterated logarithm law,* Ann. Appl. Probab. 29(2019), 1487–1540.

[16] M. Röckner, F.-Y. Wang, L. Wu, *Large deviations for stochastic generalized porous media equations,* Stoch. Proc. Appl. 116(2006), 1677–1689.

[17] A.-S. Sznitman, *Topics in Propagation of Chaos,* Springer, 1991.

[18] F.-Y. Wang, *Harnack inequalities and Applications for Stochastic Partial Differential Equations*, Springer, 2013, Berlin.

[19] F.-Y. Wang, *Hypercontractivity and applications for stochastic Hamiltonian systems,* J. Funct. Anal. 272(2017), 5360–5383.

[20] F.-Y. Wang, *Distribution dependent SDEs for Landau type equations,* Stoch. Proc. Appl. 128(2018), 595–621.

[21] F.-Y. Wang, X. Zhang, *Derivative formula and applications for degenerate diffusion semi-groups,* J. Math. Pures Appl. 99(2013), 726–740.

[22] L. Wu, *Some notes on large deviations of Markov processes,* Acta Math. Sin. (English Ser.) 16(2000), 369–394.

[23] L. Wu, *Uniformly integrable operators and large deviations for Markov processes,* J. Funct. Anal. 172(2000), 301–376.