Wasserstein Convergence Rate for Empirical Measures of Markov Processes

Feng-Yu Wang

Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

wangfy@tju.edu.cn

May 2, 2025

Abstract

The convergence rate in Wasserstein distance is estimated for empirical measures of ergodic Markov processes, and the estimate can be sharp in some specific situations. The main result is applied to subordinations of typical models excluded by existing results, which include: stochastic Hamiltonian systems on $\mathbb{R}^n \times \mathbb{R}^m$, spherical velocity Langevin processes on $\mathbb{R}^n \times \mathbb{S}^{n-1}$, multi-dimensional Wright-Fisher type diffusion processes, and stable type jump processes.

AMS subject Classification: 60B05, 60B10.
Keywords: Convergence rate, empirical measure, Wasserstein distance, Markov process.

1 Introduction

The purpose of this paper is to provide a general result on the Wasserstein convergence rate of empirical measures, which applies to a broad class of ergodic Markov processes including typical models beyond the range of existing results.

1.1 Problem in existing study

In recent years the Wasserstein convergence rate has been intensively investigated for the empirical measures of continuous time stochastic systems, see [30, 31, 32, 33, 19] for symmetric diffusion processes, [16, 17, 15, 18, 34, 36] for subordinate diffusion processes, [13, 18] for the fractional Brownian motion on flat torus. See also [40] for the study of weighted empirical measures of symmetric diffusions on compact manifolds.

In these references, the symmetric part of the generator has discrete spectrum with positive spectral gap, see [34, (2.6)]. In particular, there exists a constant $c > 0$ such that the following

Poincaré inequality holds:

$$\mu(f^2) - \mu(f)^2 \le c\,\hat{\mathscr{E}}(f,f),\quad \hat{\mathscr{E}}(f,f) := -\mu(fLf). \tag{1.1}$$

However, this restriction excludes degenerate models where $\mathscr{E}(f,f)$ is reducible.

For instance, consider the following stochastic Hamiltonian system $X_t := (X_t^{(1)}, X_t^{(2)})$ on $\mathbb{R}^d \times \mathbb{R}^d = \mathbb{R}^{2d}$:

$$\begin{cases} \mathrm{d}X_t^{(1)} = X_t^{(2)}\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = \sqrt{2}\,\mathrm{d}W_t - \big\{\nabla V(X_t^{(1)}) + X_t^{(2)}\big\}\mathrm{d}t, \end{cases} \tag{1.2}$$

where W_t is the Brownian motion on $\mathbb{R}^d$, and $V \in C^2(\mathbb{R}^d)$ such that $Z_V := \int_{\mathbb{R}^d} \mathrm{e}^{-V(x)}\mathrm{d}x < \infty$ and $\|\nabla^2 V\| \le C(1 + |\nabla V|)$ holds or some constant $C > 0$. In this case, the solution of (1.2) is a diffusion process having invariant probability measure $\mu = \mu_V \times \mathscr{N}_1$, where $\mu_V(\mathrm{d}x) = Z_V^{-1}\mathrm{e}^{-V(x)}\mathrm{d}x$ and $\mathscr{N}_1$ is the standard Gaussian measure on $\mathbb{R}^d$. The energy form associated with (1.2) is

$$\hat{\mathscr{E}}(f,f) = \mu\Big(\big|\nabla_{x^{(2)}} f(x^{(1)},\cdot)\big|^2(x^{(2)})\Big),\quad f \in C_0^2(\mathbb{R}^{2d}).$$

Since the gradient is only taken for the second variable, when f is a non-constant function only depending on $x^{(1)}$, the Poincaré inequality (1.1) does not hold.

On the other hand, the Markov process $X_t = (X_t^{(1)}, X_t^{(2)})$ solving (1.2) may be exponential ergodic (see [24]), so it is natural to ask for the convergence rate of the empirical measure with respect to a reasonable Wasserstein distance.

1.2 New idea of the present work

To derive sharp estimates on Wasserstein distance of empirical measures, we need to regularize the empirical measures such that analytic inequalities apply. In previous references, the empirical measures are regularized by using the semigroup $\hat{P}_t$ generated by the symmetric part of the Markov generator under study, for which we need to make assumptions on $\hat{P}_t$.

For instance, let $\hat{L} := \Delta + \nabla V$ on a compete connected Riemannian manifold M for a smooth function V such that $\mu(\mathrm{d}x) := \mathrm{e}^{V(x)}\mathrm{d}x$ is a probability measure, where $\mathrm{d}x$ stands for the Riemannian volume measure, and let $\mathbb{W}_p$ be the p-Wasserstein distance induced by the Riemannian distance ρ, see (2.12) below. Then by [14, Theorem 2], for any probability measure $\nu(\mathrm{d}x) := f(x)\mu(\mathrm{d}x)$ on M, we have

$$\mathbb{W}_p(\nu,\mu) \le p\Big(\mu\big(|\nabla(-\hat{L})^{-1}(f-1)|^p\big)\Big)^{\frac{1}{p}}. \tag{1.3}$$

See also [2] for a refined estimate for $\mathbb{W}_2$. Since the empirical measure of the diffusion process generated by $\hat{L}$ is singular with respect to μ, to apply the estimate (1.3), one regularizes the empirical measure by the diffusion semigroup $\hat{P}_t$, see [2] and [30]-[33].

However, the above technique does not apply to degenerate models like stochastic Hamiltonian systems arising from kinetic mechanics, where the symmetric part $\hat{L}$ of the generator does not induce any distance.

To overcome this problem, we choose a different symmetric semigroup $\hat{P}_t$, which is not generated by the symmetric part of the underlying Markov process, but has the same invariant probability measure μ. By choosing such a symmetric diffusion semigroup satisfying conditions needed in the study, we are able to apply (1.3) for the generator $\hat{L}$ of $\hat{P}_t$ to derive explicit convergence rate of the empirical measure with respect to the Wasserstein distance, which is induced by the intrinsic distance of $\hat{L}$.

1.3 Organization of the paper

In Section 2, we introduce the framework of the present study, state the main result for exponential ergodic Markov processes (Theorem 2.2), and an extension for non-exponential ergodic Markov processes (Theorem 2.3). The convergence rate presented in Theorem 2.2 is sharp for specific models shown by Examples 2.1-2.2 and Remark 6.1. Theorem 2.3 applies to any Markov process whose semigroup converges to the invariant probability measure at certain rate corresponding to the weak Poincaré inequality introduced in [21].

In Sections 3-6, we apply the main result to subordinations of several typical models: stochastic Hamiltonian systems, spherical velocity Langevin processes, Wright-Fisher type diffusion processes, and stable like processes. These models arise from different applied areas, and are not covered by existing results on Wasserstein convergence rate of empirical measures.

2 Framework and main result

Let (M,ρ) be a length space, let $\mathscr{P}(M)$ be the set of all probability measures on M. For any $p\in[1,\infty)$, the L^p-Wasserstein distance is defined as

$$\mathbb{W}_p(\nu_1,\nu_2) := \inf_{\pi\in\mathscr{C}(\nu_1,\nu_2)} \left(\int_{M\times M} \rho(x,y)^p \pi(\mathrm{d}x,\mathrm{d}y)\right)^{\frac{1}{p}}, \quad \nu_1,\nu_2\in\mathscr{P}(M), \tag{2.1}$$

where $\mathscr{C}(\nu_1,\nu_2)$ is the set of all couplings for ν_1 and ν_2. We study the convergence rate in $\mathbb{W}_p$ for empirical measures of ergodic Markov processes on M.

2.1 Subordinate Markov process

Let X_t be a standard time-homogenous Markov process on M having invariant probability measure $\mu\in\mathscr{P}(M)$. The associated Markov semigroup P_t is defined as $P_tf(x) := \mathbb{E}^x[f(X_t)]$ for $t\geq 0, f\in\mathscr{B}_b(M)$, where $\mathscr{B}_b(M)$ is the class of bounded measurable functions on M, $\mathbb{E}^x$ is the expectation taken for the underlying Markov process starting at point x. In general, for any $\nu\in\mathscr{P}(M)$, $\mathbb{E}^\nu$ denotes the expectation for the Markov process with initial distribution ν.

An important class of Markov jump processes are the subordinations (time changes) of diffusion processes induced by Bernstein functions. A typical model is the α-stable process generated by the fractional Laplacian, which is the time change of Brownian motion and has been used as Lévy noise in SDEs. See the monograph [6] and references therein for the study of subordinated Markov processes and applications.

To make time changes (i.e. the subordination) of the Markov process X_t, we introduce the class $\mathbf{B}$ of Bernstein functions B with $B(0)=0$. Recall that a Bernstein function is a function $B\in C([0,\infty))\cap C^\infty((0,\infty))$ satisfying $(-1)^{n-1}\frac{\mathrm{d}^n B(s)}{\mathrm{d}s^n}\ge 0,\quad s>0$. For any $\alpha\in(0,1]$, let

$$\mathbf{B}^\alpha := \Big\{B\in\mathbf{B}:\ \liminf_{r\to\infty} B(r)r^{-\alpha}>0\Big\}.$$

Obviously, $\mathbf{B}^0=\mathbf{B}$.

For each $B\in\mathbf{B}$, there exists a unique stable increasing process S_t^B on $[0,\infty)$ with Laplace transform

$$\mathbb{E}\big[\mathrm{e}^{-rS_t^B}\big]=\mathrm{e}^{-B(r)t},\quad t,r\ge 0, \tag{2.2}$$

see for instance [22]. Let S_t^B be independent of X_t. Consider the subordinate diffusion process $X_t^B:=X_{S_t^B}, t\ge 0$, and its empirical measures

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

We investigate the convergence rate of $\mathbb{W}_p(\mu_t^B,\mu)\to 0$ as $t\to\infty$.

2.2 Reference symmetric diffusion process

Let $\hat X_t$ be a reversible Markov process on M with the same invariant probability measure μ, and with ρ as the intrinsic distance. Heuristically, $\hat X_t$ has symmetric Dirichlet form $(\hat{\mathscr{E}},\mathscr{D}(\hat{\mathscr{E}}))$ in $L^2(\mu)$ satisfying

$$\hat{\mathscr{E}}(f,f)=\int_M |\nabla f|^2\mathrm{d}\mu,\quad f\in C_{b,L}(M)\subset\mathscr{D}(\hat{\mathscr{E}}),$$

where $C_{b,L}(M)$ be the set of all bounded Lipschitz continuous functions on M, and

$$|\nabla f(x)|:=\limsup_{y\to x}\frac{|f(y)-f(x)|}{\rho(x,y)},\quad x\in M.$$

More precisely, we assume that $C_{b,L}(M)$ is a dense subset of $\mathscr{D}(\hat{\mathscr{E}})$ and

$$\hat{\mathscr{E}}(f,g)=\int_M \Gamma(f,g)\mathrm{d}\mu,\quad f,g\in C_{b,L}(M),$$

where

$$\Gamma: C_{b,L}(M)\times C_{b,L}(M)\to\mathscr{B}_b(M)$$

is a symmetric local square field (champ de carré), i.e. for any $f,g,h\in C_{b,L}(M)$ and $\phi\in C_b^1(\mathbb{R})$,

$$\sqrt{\Gamma(f,f)(x)}=|\nabla f(x)|:=\limsup_{y\to x}\frac{|f(y)-f(x)|}{\rho(x,y)},\quad x\in M,$$
$$\Gamma(fg,h)=f\Gamma(g,h)+g\Gamma(f,h),\quad \Gamma(\phi(f),h)=\phi'(f)\Gamma(f,h).$$

Moreover, the generator $(\hat L,\mathscr{D}(\hat L))$ satisfies the chain rule

$$\hat L\phi(f)=\phi'(f)\hat Lf+\phi''(f)|\nabla f|^2,\quad f\in\mathscr{D}(\hat L)\cap C_{b,L}(M),\phi\in C^2(\mathbb{R}).$$

2.3 Main result

Note that for $B(r) = r$, we have $X_t^B = X_t$ so that μ_t^B reduces to the empirical measure $\mu_t := \frac{1}{t}\int_0^t \delta_{X_s} \mathrm{d}s$. In this paper, we aim to estimate $\mathbb{W}_p(\mu_t^B, \mu)$ for a general Bernstein function B, which includes $\mathbb{W}_p(\mu_t, \mu)$ as special case. To this end, we make the following assumption.

For any $p \geq q \geq 1$, let $\|\cdot\|_{L^q(\mu)\to L^p(\mu)}$ be the operator norm from $L^q(\mu)$ to $L^p(\mu)$. Let $(\hat{P}_t)_{t\geq 0}$ be the semigroup of the reversible Markov process $\hat{X}_t$, i.e.

$$\hat{P}_t f(x) := \mathbb{E}\big(f(\hat{X}_t)\big|\hat{X}_0 = x\big), \quad t \geq 0, f \in \mathscr{B}_b(M),$$

where $\mathscr{B}_b(M)$ is the set of all bounded measurable functions on M,

(A_1) *Let $p \in [2, \infty)$. $\hat{P}_t$ has heat kernel $\hat{p}_t$ with respect to μ, and there exist constants $\beta, \lambda, d, k \in (0, \infty)$ such that*

$$\|\nabla \hat{P}_t\|_{L^2(\mu)\to L^p(\mu)} \leq k\mathrm{e}^{-\lambda t} t^{-\beta}, \quad t > 0, \tag{2.3}$$

$$\int_M \big(\hat{P}_t \rho(x, \cdot)^p\big)^{\frac{2}{p}}(x)\mu(\mathrm{d}x) \leq kt, \quad t \in (0, 1], x \in M, \tag{2.4}$$

$$\int_M \hat{p}_t(x, x)\mu(\mathrm{d}x) \leq k(1 \wedge t)^{-\frac{d}{2}}, \quad t > 0, \tag{2.5}$$

$$\|P_t - \mu\|_{L^2(\mu)} \leq k\mathrm{e}^{-\lambda t}, \quad t \geq 0. \tag{2.6}$$

Note that in (A_1) the only condition we need for the Markov process is (2.6), while other conditions (2.3)-(2.5) are made for the reference semigroup $\hat{P}_t$ which is flexible in applications. Indeed, for smaller distance ρ, the energy form $\hat{\mathscr{E}}$ is bigger, so that $\hat{P}_t$ has better properties. For instance, let μ be a probability measure on a connected Riemannian manifold comparable with the volume measure, when the Riemmanian distance ρ is small enough we have large enough Dirichlet form $\hat{\mathscr{E}}(f, f) := \mu(|\nabla f|^2)$ such that $\mathrm{gap}(\hat{L}) > 0$, see [5] where the stronger log-Sobolev inequality is considered.

Condition (2.4) refers to the $\frac{1}{2}$-Hölder continuity of the symmetric diffusion process $\hat{X}_t$, which is true for a broad class of diffusion processes. Indeed, for a diffusion process $\hat{X}_t$ with generator satisfying

$$\hat{L}\rho(x, \cdot)^p \leq c\big(1 + \rho(x, \cdot)^p\big)$$

for some constant $c > 0$, which is the case when $\hat{X}_t$ solves an SDE on $\mathbb{R}^d$ with linear growth coefficients and $\rho(x, y) := |x - y|$, we have

$$\hat{P}_t \rho(x, \cdot)^p(x) = \mathbb{E}^x[\rho(x, \hat{X}_t)^p] \leq c \int_0^t \mathrm{e}^{cs} \mathrm{d}s \leq c\mathrm{e}^c t, \quad t \in [0, 1],$$

which implies (2.4) for some constant $k > 0$ and any probability measure μ. Condition (2.5) is a standard upper bound estimate on the heat kernel for d-dimensional elliptic diffusions, see for instance [9, Theorem 2.3.6]. Moreover, according to Proposition 2.1 below, when $p = 2$, condition (2.3) with $\beta = \frac{1}{2}$ follows from the existence of spectral gap, i.e.

$$\mathrm{gap}(\hat{L}) := \inf\big\{\hat{\mathscr{E}}(f, f) : \ f \in \mathscr{D}(\hat{\mathscr{E}}), \mu(f) = 0, \mu(f^2) = 1\big\} > 0.$$

In this case, (2.3) follows from (2.5), since the later implies that $\hat{L}$ has discrete spectrum and hence has a spectral gap, see [27, Theorem 3.3.19]. When $p > 2$, we have

$$\|\nabla \hat{P}_t\|_{L^2(\mu)\to L^p(\mu)} = \|\nabla \hat{P}_{\frac{t}{2}} \hat{P}_{\frac{t}{2}}\|_{L^2(\mu)\to L^p(\mu)} \le \|\nabla \hat{P}_{\frac{t}{2}}\|_{L^2(\mu)\to L^2(\mu)} \|\hat{P}_{\frac{t}{2}}\|_{L^2(\mu)\to L^p(\mu)},$$

so that (2.3) follows from $\mathrm{gap}(\hat{L}) > 0$ together with a suitable upper bound of $\|\hat{P}_t\|_{L^2(\mu)\to L^p(\mu)}$, which is available for elliptic diffusions on compact manifolds, see the proof of Example 2.2 for details.

Before moving on, let us compare the above conditions with those in [34, (A_1)]: there exist constants $c, \lambda > 0, d \ge d' \ge 1$ and a map $k : (1,\infty) \to (0,\infty)$ such that

$$\|\hat{P}_t - \mu\|_{1\to\infty} \le c t^{-\frac{d}{2}} \mathrm{e}^{-\lambda t}, \quad t > 0, \tag{2.7}$$

$$\lambda_i \ge c i^{\frac{2}{d'}}, \quad i \in \mathbb{Z}_+, \tag{2.8}$$

$$|\nabla \hat{P}_t f| \le k(p) (\hat{P}_t |\nabla f|^p)^{\frac{1}{p}}, \quad t \in [0,1], p \in (1,\infty), f \in C_{b,L}(M). \tag{2.9}$$

The first essential difference is that $\hat{P}_t$ in [34, (A_1)] is associated with the symmetric part of the Dirichlet form for the underlying Markov process, while $\hat{P}_t$ in the present framework is essentially different, the only link between the present $\hat{P}_t$ and the underlying Markov semigroup P_t is that they share the invariant probability measure μ.

Since the generator $\hat{L}$ of $\hat{P}_t$ is not the symmetric part of the generator L for the studied Markov process, the eigenvalues of $\hat{L}$ has nothing to do with the behavior of the underlying Markov process, so the condition (2.8) is dropped from the present assumption (A_1).

As we will use $\hat{P}_t$ to regularize the empirical measures, we adopt the conditions (2.3) and (2.5) for the gradient and heat kernel estimates on $\hat{P}_t$, where (2.3) is comparable with (2.7) for small time. Again, because the eigenvalues of $\hat{L}$ has nothing to do with the underlying Markov generator, the spectral representation of $\hat{P}_t$ is no longer useful for the study, we need the gradient estimate (2.3) rather than (2.9), where the later is easier to verify in applications.

Proposition 2.1. *If* $\mathrm{gap}(\hat{L}) > 0$, *then for any* $\lambda \in (0, \mathrm{gap}(\hat{L}))$ *there exists a constant* $k > 0$ *such that*

$$\|\nabla \hat{P}_t f\|_{L^2(\mu)} \le k t^{-\frac{1}{2}} \mathrm{e}^{-\lambda t} \|f\|_{L^2(\mu)}, \quad t > 0, f \in L^2(\mu).$$

Proof. Denote $\lambda_1 := \mathrm{gap}(\hat{L})$, let $(E_s)_{s\ge0}$ be the spectral family of $-\hat{L}$. We find a constant $k > 0$ such that

$$\begin{aligned}\|\nabla \hat{P}_t f\|_{L^2(\mu)}^2 &= \int_{\lambda_1}^\infty s \mathrm{e}^{-2st} \mathrm{d}E_s(f) \le \mathrm{e}^{-2\lambda t} \Big(\sup_{s\ge\lambda_1} s \mathrm{e}^{-2(s-\lambda)t} \Big) \int_{\lambda_1}^\infty \mathrm{d}E_s(f) \\ &\le k t^{-1} \mathrm{e}^{-2\lambda t} \int_{\lambda_1}^\infty \mathrm{d}E_s(f) \le k t^{-1} \mathrm{e}^{-2\lambda t} \|f\|_{L^2(\mu)}^2, \quad t > 0, f \in L^2(\mu).\end{aligned}$$

□

We also need the following defined quantity $d' \in (0,\infty]$ induced by P_t.

Definition 2.1. d' is the smallest positive constant such that the heat kernel p_t of P_t with respect to μ satisfies

$$\int_{M\times M} p_t(x,y)^2\mu(\mathrm{d}x)\mu(\mathrm{d}y) \le k(1\wedge t)^{-\frac{d'}{2}},\quad t>0. \tag{2.10}$$

If p_t does not exist, or p_t exists but (2.10) does not hold for any $d' \in (0,\infty)$, we denote $d' = \infty$.

For constants β, d in (A_1), and d' in Definition 2.1, we denote

$$K_{\beta,d,d',\alpha} := \beta + \frac{d}{8}\Big[1+\Big(1-\frac{4\alpha}{d'}\Big)^+\Big],\quad \alpha\in[0,1].$$

Moreover, for any $t>0$, let

$$\xi_t := \begin{cases} t^{-1}, & \text{if } K_{\beta,d,d',\alpha}<1,\\ t^{-1}[\log(2+t)]^2, & \text{if } K_{\beta,d,d',\alpha}=1,\ d'\ne 4\alpha,\\ t^{-1}[\log(2+t)]^3, & \text{if } K_{\beta,d,d',\alpha}=1,\ d'= 4\alpha,\\ t^{-\frac{1}{2K_{\beta,d,d',\alpha}-1}}, & \text{if } K_{\beta,d,d',\alpha}>1,\ d'\ne 4\alpha,\\ t^{-\frac{1}{2K_{\beta,d,d',\alpha}-1}}\log(2+t), & \text{if } K_{\beta,d,d',\alpha}>1,\ d'= 4\alpha. \end{cases} \tag{2.11}$$

Theorem 2.2. *Assume (A_1) for some $p\in[2,\infty)$ and let $B\in\mathbf{B}^\alpha$ for some $\alpha\in[0,1]$. Then there exists a constant $c>0$ such that*

$$\mathbb{E}^\mu[\mathbb{W}_p(\mu_t^B,\mu)^2] \le c\xi_t,\quad t>0. \tag{2.12}$$

If the semigroup P_t^B of X_t^B has heat kernel p_t^B with respect to μ, then for any $q\in[1,2]$ and $x\in M$,

$$\mathbb{E}^x[\mathbb{W}_p(\mu_t^B,\mu)^q] \le \frac{2^{q-1}}{t^q}\int_0^1 \mathbb{E}^x\big[\mu(\rho(X_s^B,\cdot)^p)^{\frac{q}{p}}\big]\mathrm{d}s + 2^{q-1}\|p_1^B(x,\cdot)\|_{L^{\frac{2}{2-q}}(\mu)}(c\xi_{t-1})^{\frac{q}{2}},\quad t>1. \tag{2.13}$$

Remark 2.1. (1) The reference semigroup $\hat{P}_t$ from (A_1) will be used to regularize the empirical measure μ_t^B into $\mu_{t,r}^B := \mu_t^B\hat{P}_r$ for $r\in(0,1)$, which has density with respect to μ so that the estimate (1.3) applies to $\mathbb{W}_p(\mu_{t,r}^B,\mu)$, see the proof of Theorem 2.2 for details. In particular, for the stochastic Hamiltonian system (1.2), conditions (2.3)-(2.5) hold for $\hat{P}_t$ generated by $\hat{L} := \Delta - \nabla H$ on $\mathbb{R}^{2d}$, where $H(x^{(1)},x^{(2)}) := V(x^{(1)}) + \frac{1}{2}|x^{(2)}|^2$ for $x^{(1)},x^{(2)}\in\mathbb{R}^d$, see the proof of Theorem 3.1 below with $m=n=d, \kappa=1$ and Q being the identity matrix.

(2) By the standard Markov property, for any $\nu\in\mathscr{P}$ and $q\in[1,2]$ with $h_\nu := \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\in L^{\frac{2}{2-q}}(\mu)$, (2.12) implies

$$\big(\mathbb{E}^\nu[\mathbb{W}_p(\mu_t^B,\mu)^q]\big)^{\frac{2}{q}} = \Big(\int_M h_\nu(x)\mathbb{E}^x[\mathbb{W}_p(\mu_t^B,\mu)^q]\mu(\mathrm{d}x)\Big)^{\frac{2}{q}}$$

$$
\begin{aligned}
&\le \Big(\int_M h_\nu(x)\big(\mathbb{E}^x[\mathbb{W}_p(\mu_t^B,\mu)^2]\big)^{\frac{q}{2}}\mu(\mathrm{d}x)\Big)^{\frac{2}{q}}\\
&\le \|h_\nu\|_{L^{\frac{2}{2-q}}(\mu)}^{\frac{2}{q}}\int_M \mathbb{E}^x[\mathbb{W}_p(\mu_t^B,\mu)^2]\mu(\mathrm{d}x)\\
&= \|h_\nu\|_{L^{\frac{2}{2-q}}(\mu)}^{\frac{2}{q}}\mathbb{E}^\mu[\mathbb{W}_p(\mu_t^B,\mu)^2]\le c\|h_\nu\|_{L^{\frac{2}{2-q}}(\mu)}^{\frac{2}{q}}\xi_t,\quad t>0.
\end{aligned}
$$

(3) It is easy to see that ξ_t is decreasing in d', so estimates in Theorem 2.2 remain true if d' is replaced by ∞, for which $K_{\beta,d,d',\alpha}=K:=\beta+\frac{d}{4}$ and ξ_t reduces to

$$
\xi_t(K):=\begin{cases} t^{-1}, & \text{if } K<1,\\ t^{-1}[\log(2+t)]^2, & \text{if } K=1,\\ t^{-\frac{1}{2K-1}}, & \text{if } K>1.\end{cases} \tag{2.14}
$$

Therefore, when μ is good enough such that the associated symmetric diffusion semigroup $\hat{P}_t$ satisfies conditions (2.3)-(2.5), then for any Markov process satisfying (2.6) and any $B\in\mathbf{B}$, there exists a constant $c>0$ such that

$$
\mathbb{E}^\mu[\mathbb{W}_p(\mu_t^B,\mu)^2]\le c\xi_t(K),\quad t>0.
$$

To illustrate Remark 2.1(2), we present below two examples, which provide a uniform Wasserstein convergence rate for empirical measures of Markov processes with given invariant measure μ, where the uniform rate is sharp in the second example.

Example 2.1. Let $M=\mathbb{R}^n$, let $V\in C^2(\mathbb{R}^n)$ such that $V(x)=\psi(x)+(1+\theta|x|^2)^\tau,\ x\in\mathbb{R}^n$, where $\psi\in C_b^2(\mathbb{R}^n)$, $\theta>0,\tau\in(\frac{1}{2},\infty]$ are constants. Let $\mu(\mathrm{d}x)=\mu_V(\mathrm{d}x):=\frac{\mathrm{e}^{-V(x)}\mathrm{d}x}{\int_{\mathbb{R}^n}\mathrm{e}^{-V(x)}\mathrm{d}x}$. Then for any Markov process on $\mathbb{R}^n$ satisfying (2.6) and any $B\in\mathbf{B}$, there exists a constant $c>0$ such that

$$
\mathbb{E}^\mu[\mathbb{W}_2(\mu_t^B,\mu)^2]\le c\begin{cases} t^{-1}, & \text{if } n=1,\tau>1,\\ t^{-1}[\log(2+t)]^2, & \text{if } n=1,\tau=1,\\ t^{-\frac{2\tau-1}{\tau n}}, & \text{otherwise.}\end{cases} \tag{2.15}
$$

Proof. Let $\hat{L}=\Delta-\nabla V$. By Remark 2.1(2), it suffices to verify (2.3)-(2.5) for $p=2,\beta=\frac{1}{2}$, and $d=\frac{2\tau n}{2\tau-1}$. Since $\lim_{|x|\to\infty}\hat{L}|\cdot|(x)=-\infty<0$, [26, Corollary 1.4] ensures $\mathrm{gap}(\hat{L})>0$, so that by Proposition 2.1, (2.3) holds for $p=2$ and $\beta=\frac{1}{2}$.

Next, by [28, Theorem 2.4.4] and $\nabla^2 V\ge -c_1 I_n$, we find a constant $c_2>0$ such that

$$
\hat{p}_r(x,x)\le\frac{c_2}{\mu(B(x,\sqrt{r}))},\quad x\in\mathbb{R}^n, r\in(0,1], \tag{2.16}
$$

where $B(x,r):=\{y\in\mathbb{R}^n:|x-y|<r\}, r>0$. Then (2.5) with $d=\frac{2\tau n}{2\tau-1}$ follows provided

$$
\int_{\mathbb{R}^n}\frac{\mu(\mathrm{d}x)}{\mu(B(x,r))}\le cr^{-\frac{2\tau n}{2\tau-1}},\quad r\in(0,1] \tag{2.17}
$$

holds for some constant $c > 0$. Below we prove this estimate.

Since ψ is bounded, there exists a constant $C > 1$ such that

$$C^{-1}\mathrm{e}^{-(1+\theta|x|^2)^\tau}\mathrm{d}x \le \mu(\mathrm{d}x) \le C\mathrm{e}^{-(1+\theta|x|^2)^\tau}\mathrm{d}x. \tag{2.18}$$

On the other hand, by

$$\frac{|x|}{2} \le |x| - \frac{r}{4} \le |x|, \quad r \in [0,1],\ |x| \ge 1,$$

we find a constant $c_3 > 0$ such that

$$\begin{aligned}
&\Big(1+\theta\Big|x-\frac{rx}{4|x|}\Big|^2\Big)^\tau = (1+\theta|x|^2)^\tau + \int_0^r \frac{\mathrm{d}}{\mathrm{d}s}\Big(1+\theta\big(|x|-\frac{s}{4}\big)^2\Big)^\tau \mathrm{d}s\\
&= (1+\theta|x|^2)^\tau - \frac{\tau\theta}{2}\int_0^r \Big(1+\theta\big(|x|-\frac{s}{4}\big)^2\Big)^{\tau-1}\Big(|x|-\frac{s}{4}\Big)\mathrm{d}s\\
&\le (1+\theta|x|^2)^\tau - c_3 r|x|^{2\tau-1}, \quad r\in[0,1],\ |x|\ge 1.
\end{aligned}$$

Hence, there exist constants $c_4, c_5 > 0$ such that for $|x| \ge 1$ and $r \in (0,1]$,

$$\begin{aligned}
&\mu(B(x,r)) \ge c_4 \int_{B\left(x-\frac{rx}{2|x|},\frac{r}{4}\right)} \mathrm{e}^{-(1+\theta|y|^2)^\tau}\mathrm{d}y\\
&\ge c_5 r^n \mathrm{e}^{-(1+\theta|x-\frac{rx}{4|x|}|^2)^\tau} \ge c_5 r^n \mathrm{e}^{-(1+\theta|x|^2)^\tau + c_3 r|x|^{2\tau-1}}, \quad r\in[0,1],\ |x|\ge 1.
\end{aligned} \tag{2.19}$$

Combining this with (2.18), we find constants $c_6, c_7 > 0$ such that

$$\begin{aligned}
&\int_{\mathbb{R}^n} \frac{\mu(\mathrm{d}x)}{\mu(B(x,r))} = \int_{B(0,1)} \frac{\mu(\mathrm{d}x)}{\mu(B(x,r))} + \int_{B(0,1)^c} \frac{\mu(\mathrm{d}x)}{\mu(B(x,r))}\\
&\le C^2 r^{-n} + c_6 r^{-n}\int_{\mathbb{R}^n} \mathrm{e}^{-c_3 r|x|^{2\tau-1}}\mathrm{d}x = C^2 r^{-n} + c_7 r^{-n}\int_0^\infty s^{n-1}\mathrm{e}^{-c_3 r s^{2\tau-1}}\mathrm{d}s\\
&= C^2 r^{-n} + c_7 r^{-\frac{2\tau n}{2\tau-1}}\int_0^\infty s^{n-1}\mathrm{e}^{-c_3 s^{2\tau-1}}\mathrm{d}s, \quad r\in(0,1].
\end{aligned}$$

This implies (2.17) for some constant $c > 0$.

Finally, it is easy to see that $\nabla^2 V \ge -cI_n$ and $|\nabla V(x)|^2 \le c(1+|x|^{4\tau})$ hold for some constant $c > 0$. So, we find a constant $c_8 > 0$ such that

$$\begin{aligned}
&\hat{L}|x-\cdot|^2 = 2n + 2\langle \nabla V, x-\cdot\rangle = 2n + 2\langle \nabla V(x), x-\cdot\rangle - 2\langle \nabla V(x) - \nabla V, x-\cdot\rangle\\
&\le 2n + |\nabla V(x)|^2 + |x-\cdot|^2 + 2c_1|x-\cdot|^2 \le c_8(1+|x|^{4\tau} + |x-\cdot|^2), \quad x\in\mathbb{R}^n.
\end{aligned} \tag{2.20}$$

This implies

$$\hat{P}_t|x-\cdot|^2(x) = \mathbb{E}^x|x-\hat{X}_t|^2 \le c_8\big(1+|x|^{4\tau}|\big)t\mathrm{e}^{c_8 t}, \quad x\in\mathbb{R}^n,\ t>0. \tag{2.21}$$

Noting that $\mu(|\cdot|^{4\tau}) < \infty$, we verify condition (2.4) for $p = 2$ and some constant $k > 0$. □

In the next example, the upper bound (2.22) is sharp. Indeed, according to [36, Corollary 1.3(2)], for $n \ge 3$ and $B \in \mathbf{B}^\alpha$, we have $\inf_{x\in M}\mathbb{E}[\mathbb{W}_2(\mu_t^B,\mu)^2] \ge ct^{-\frac{2}{n-2\alpha}}$. With $\alpha \to 0$ this lower bound reduces to the upper bound in (2.22).

Example 2.2. Let M be an n-dimensional compact connected Riemanian manifold possibly with a boundary ∂M. Let $V \in C^2(M)$ such that $\mu(\mathrm{d}x) = \mathrm{e}^{-V(x)}\mathrm{d}x$ is a probability measure on M. Then for any Markov process on M satisfying (2.6) and any $B \in \mathbf{B}$, there exists a constant $c > 0$ such that

$$\mathbb{E}^\mu[\mathbb{W}_2(\mu_t^B, \mu)^2] \le c \begin{cases} t^{-1}, & \text{if } n = 1, \\ t^{-1}[\log(2+t)]^2, & \text{if } n = 2, \\ t^{-\frac{2}{n}}, & \text{if } n \ge 3. \end{cases} \tag{2.22}$$

In general, for any $p \ge 2$, there exists a constant $c > 0$ such that

$$\mathbb{E}^\mu[\mathbb{W}_p(\mu_t^B, \mu)^2] \le c \begin{cases} t^{-1}, & \text{if } n = 1, p \in [2,3), \\ t^{-1}[\log(2+t)]^2, & \text{if } n(p-1) = 2, \\ t^{-\frac{2}{n(p-1)}}, & \text{otherwise.} \end{cases} \tag{2.23}$$

Proof. Let $\hat{X}_t$ be the diffusion process generated by $\hat{L} = \Delta - \nabla V$ with reflecting boundary if it exists. Since M is a compact connected Riemannian manifold, (2.3)-(2.5) are well known for $p = 2, \beta = \frac{1}{2}$ and $d = n$, hence (2.22) holds according to Remark 2.1(2). Moreover, it is classical that $\|\hat{P}_t\|_{L^2(\mu)\to L^p(\mu)} \le c(1 \wedge t)^{-\frac{n(p-2)}{4p}}$ holds for some constant $c > 0$ and all $t > 0$. Then (2.3) holds for $p \ge 2$ and $\beta = \frac{1}{2} + \frac{n(p-2)}{4p}$. By Remark 2.1(2), (2.23) holds. □

2.4 Proof of Theorem 2.2

We will apply the estimate (1.3) for $p \in [2, \infty)$. This inequality is proved in [14] by using the Kantorovich dual formula and Hamilton-Jacobi equations, which are available when (M, ρ) is a length space as we assumed, see [25]. In the following, we prove estimates (2.12) and (2.13) by five steps.

(a) For any $r > 0$ and $t > 0$, consider the regularized empirical measure

$$\mu_{t,r}^B := \mu_t^B \hat{P}_r = f_{t,r}\mu, \quad f_{t,r} := \frac{1}{t}\int_0^t \hat{p}_r(X_s^B, \cdot)\mathrm{d}s. \tag{2.24}$$

Note that $\pi := \frac{1}{t}\int_0^t \delta_{X_s^B} \times (\delta_{X_s^B}\hat{P}_r)\mathrm{d}s \in \mathscr{C}(\mu_t^B, \mu_{t,r}^B)$, so that by the definition of $\mathbb{W}_p$ and Jensen's inequality as $p \ge 2$, we obtain

$$\begin{aligned} \mathbb{W}_p(\mu_{t,r}^B, \mu_t^B)^2 &\le \bigg(\int_{M\times M} \rho(x,y)^p \pi(\mathrm{d}x, \mathrm{d}y)\bigg)^{\frac{2}{p}} = \bigg(\frac{1}{t}\int_0^t \hat{P}_r\rho(X_s^B, \cdot)^p(X_s^B)\mathrm{d}s\bigg)^{\frac{2}{p}} \\ &\le \frac{1}{t}\int_0^t \big(\hat{P}_r\rho(X_s^B, \cdot)^p\big)^{\frac{2}{p}}(X_s^B)\mathrm{d}s, \quad t > 0, \ r \in (0,1]. \end{aligned}$$

Combining this with (2.4) and that μ is an invariant measure of X_s^B, we obtain

$$\mathbb{E}^\mu\big[\mathbb{W}_p(\mu_{t,r}^B, \mu_t^B)^2\big] \le \int_M \big(\hat{P}_r\rho(x, \cdot)^p\big)^{\frac{2}{p}}(x)\mu(\mathrm{d}x) \le kr, \quad t > 0, r \in (0,1]. \tag{2.25}$$

(b) Since $\hat{P}_s$ has symmetric heat kernel $\hat{p}_s$, (2.24) implies $\hat{P}_s f_{t,r} = \hat{P}_{\frac{s+r}{2}} f_{t,\frac{s+r}{2}}$. Combining this with (1.3), (2.3), Jensen's inequality, and that μ is $\hat{P}_s$-invariant, we derive

$$
\begin{aligned}
\mathbb{W}_p(\mu_{t,r}^B,\mu)^2 &\le p^2 \bigg\| \nabla \int_0^\infty \hat{P}_{\frac{r+s}{2}}(f_{t,\frac{r+s}{2}}-1)\mathrm{d}s \bigg\|_{L^p(\mu)}^2 \\
&\le 4^\beta p^2 k^2 \bigg(\int_0^\infty \frac{\mathrm{e}^{-\lambda(s+r)/2}}{(s+r)^\beta} \|f_{t,\frac{r+s}{2}}-1\|_{L^2(\mu)}\mathrm{d}s\bigg)^2.
\end{aligned}
\tag{2.26}
$$

By (2.24), we obtain

$$
\begin{aligned}
&\|f_{t,\frac{r+s}{2}}-1\|_{L^2(\mu)}^2 = \frac{2}{t^2}\int_0^t \mathrm{d}t_1 \int_{t_1}^t \mu\Big(\big(\hat{p}_{\frac{r+s}{2}}(X_{t_1}^B,\cdot)-1\big)\big(\hat{p}_{\frac{r+s}{2}}(X_{t_2}^B,\cdot)-1\big)\Big)\mathrm{d}t_2 \\
&= \frac{2}{t^2}\int_0^t \mathrm{d}t_1 \int_{t_1}^t \big[\hat{p}_{r+s}(X_{t_1}^B,X_{t_2}^B)-1\big]\mathrm{d}t_2.
\end{aligned}
$$

Combining this with (2.26), we find a constant $c_1>0$ such that for any $t>0, r\in(0,1]$ and measurable function $h:(0,\infty)\to(0,\infty)$,

$$
\begin{aligned}
&\mathbb{W}_p(\mu_{t,r}^B,\mu)^2 \le \frac{c_1}{t^2}\bigg\{ \int_0^\infty \frac{\mathrm{e}^{-\frac{\lambda}{2}(s+r)}}{(s+r)^\beta}\bigg(\int_0^t \mathrm{d}t_1 \int_{t_1}^t \big[\hat{p}_{r+s}(X_{t_1}^B,X_{t_2}^B)-1\big]\mathrm{d}t_2\bigg)^{\frac{1}{2}}\mathrm{d}s\bigg\}^2 \\
&\le \frac{c_1}{t^2}\bigg(\int_0^\infty \frac{\mathrm{e}^{-\frac{\lambda}{2}(s+r)}}{h(s+r)}\mathrm{d}s\bigg)\int_0^\infty \frac{\mathrm{e}^{-\frac{\lambda}{2}(s+r)}h(s+r)}{(s+r)^{2\beta}}\mathrm{d}s \int_0^t\mathrm{d}t_1\int_{t_1}^t \big[\hat{p}_{r+s}(X_{t_1}^B,X_{t_2}^B)-1\big]\mathrm{d}t_2.
\end{aligned}
\tag{2.27}
$$

(c) We claim that there exist constants $c',\lambda'>0$, such that for all $r>0$ and $t_2>t_1\ge 0$,

$$
\mathbb{E}^\mu\big[\hat{p}_r(X_{t_1}^B,X_{t_2}^B)-1\big] \le c'(1\wedge r)^{-\frac{d}{4}}\Big[r\wedge 1 + 1_{\{\alpha>0,d'<\infty\}}\{1\wedge(t_2-t_1)\}^{\frac{d'}{\alpha d}}\Big]^{-\frac{d}{4}}\mathrm{e}^{-\lambda'(t_2-t_1)}.
\tag{2.28}
$$

Indeed, since μ is P_t^B-invariant, μ is also P_t^{B*}-invariant, where P_t^{B*} is the adjoint operator of P_t^B in $L^2(\mu)$. By (2.2) and (2.6), we obtain that for all $t>0$,

$$
\|P_t^{B*}-\mu\|_{L^2(\mu)} = \|P_t^B-\mu\|_{L^2(\mu)} = \big\|\mathbb{E}P_{S_t^B}-\mu\big\|_{L^2(\mu)} \le k\mathbb{E}\big[\mathrm{e}^{-\lambda S_t^B}\big] = k\mathrm{e}^{-B(\lambda)t}.
\tag{2.29}
$$

Denoting $\lambda'=B(\lambda)$, noting that $\hat{p}_r(x,\cdot)-1=\int_M \hat{p}_{\frac{r}{2}}(x,y)\big[\hat{p}_{\frac{r}{2}}(y,\cdot)-1\big]\mu(\mathrm{d}y)$, by the Markov property of X_t^B, (2.5) and (2.29), we derive

$$
\begin{aligned}
&\mathbb{E}^\mu\big[\hat{p}_r(X_{t_1}^B,X_{t_2}^B)-1\big] = \int_M P_{t_2-t_1}^B\big[\hat{p}_r(x,\cdot)-1\big](x)\mu(\mathrm{d}x) \\
&= \int_{M\times M} \hat{p}_{\frac{r}{2}}(x,y)P_{t_2-t_1}^B\big[\hat{p}_{\frac{r}{2}}(y,\cdot)-1\big](x)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \\
&\le \bigg(\int_{M\times M}\hat{p}_{\frac{r}{2}}(x,y)^2\mu(\mathrm{d}x)\mu(\mathrm{d}y)\bigg)^{\frac{1}{2}}\bigg(\int_M \big\|P_{t_2-t_1}^B[\hat{p}_{\frac{r}{2}}(y,\cdot)-1]\|_{L^2(\mu)}^2\mu(\mathrm{d}y)\bigg)^{\frac{1}{2}} \\
&\le k\mathrm{e}^{-\lambda'(t_2-t_1)}\int_M \hat{p}_r(x,x)\mu(\mathrm{d}x) \le k^2(1\wedge r)^{-\frac{d}{2}}\mathrm{e}^{-\lambda'(t_2-t_1)}.
\end{aligned}
\tag{2.30}
$$

Next, let $\alpha > 0$ and $d' < \infty$. By the Markov property we obtain

$$\begin{aligned}\mathbb{E}^\mu\big[\hat{p}_r(X^B_{t_1}, X^B_{t_2}) - 1\big] &= \int_M \mathbb{E}^x\big[\hat{p}_r(x, X^B_{t_2-t_1}) - 1\big]\mu(\mathrm{d}x)\\ &= \int_{M\times M} \big\{\hat{p}_r(x,y) - 1\big\} p^B_{t_2-t_1}(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &= \int_{M\times M} \hat{p}_r(x,y)\big\{p^B_{t_2-t_1}(x,y) - 1\big\}\mu(\mathrm{d}x)\mu(\mathrm{d}y).\end{aligned} \tag{2.31}$$

Noting that

$$p^B_{t_2-t_1}(x,y) - 1 = P^{B*}_{\frac{t_2-t_1}{2}}\Big\{p^B_{\frac{t_2-t_1}{2}}(x,\cdot) - 1\Big\}(y), \quad t_2 > t_1 \geq 0, \tag{2.32}$$

by the display after [36, (3.12)], (2.10) implies

$$\int_{M\times M} p^B_t(x,y)^2\mu(\mathrm{d}x)\mu(\mathrm{d}y) \leq k\mathbb{E}\big[(1\wedge S^B_t)^{-\frac{d'}{2}}\big] \leq k'(1\wedge t)^{-\frac{d'}{2\alpha}}, \quad t > 0 \tag{2.33}$$

for some constants $k' > 0$. Combining (2.29) with (2.32), (2.5) and (2.10), we find constants $c_2, c_3 > 0$ such that

$$\begin{aligned}&\int_{M\times M} \hat{p}_r(x,y)\big\{p^B_{t_2-t_1}(x,y) - 1\big\}\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &\leq \Big(\int_M \hat{p}_{2r}(x,x)\mu(\mathrm{d}x)\Big)^{\frac{1}{2}}\Big(\int_M \Big\|P^{B*}_{\frac{t_2-t_1}{2}}\Big\{p^B_{\frac{t_2-t_1}{2}}(x,\cdot) - 1\Big\}\Big\|^2_{L^2(\mu)}\mu(\mathrm{d}x)\Big)^{\frac{1}{2}}\\ &\leq c_2(1\wedge r)^{-\frac{d}{4}}\mathrm{e}^{-(t_2-t_1)B(\lambda)/2}\Big(\int_M \Big\|p^B_{\frac{t_2-t_1}{2}}(x,\cdot) - 1\Big\|^2_{L^2(\mu)}\mu(\mathrm{d}x)\Big)^{\frac{1}{2}}\\ &\leq c_3(1\wedge r)^{-\frac{d}{4}}\{1\wedge(t_2-t_1)\}^{-\frac{d'}{4\alpha}}\mathrm{e}^{-(t_2-t_1)\lambda'}, \quad t_2 > t_1 \geq 0.\end{aligned}$$

This together with (2.30) and (2.31) implies (2.28) for some constant $c' > 0$.

(d) When $\alpha > 0$ and $d' < \infty$, we find constants $c_4, c_5 > 0$ such that

$$\begin{aligned}&\frac{1}{t}\int_0^t \mathrm{d}t_1 \int_{t_1}^t \mathrm{e}^{-\lambda'(t_2-t_1)}\big[(r+s)\wedge 1 + \{1\wedge(t_2-t_1)\}^{\frac{d'}{\alpha d}}\big]^{-\frac{d}{4}}\mathrm{d}t_2\\ &\leq \frac{c_4}{t}\int_0^t \mathrm{d}t_1\int_0^\infty \mathrm{e}^{-\lambda'\theta}\big([(r+s)\wedge 1]^{\frac{\alpha d}{d'}} + 1\wedge\theta\big)^{-\frac{d'}{4\alpha}}\mathrm{d}\theta \leq c_5 I(r+s),\end{aligned}$$

where, since $\frac{d\alpha}{d'}(\frac{d'}{4\alpha} - 1)^+ = \frac{d}{4}(1 - \frac{4\alpha}{d'})^+$,

$$I(r+s) := [(1\wedge(r+s)]^{-\frac{d}{4}(1-\frac{4\alpha}{d'})^+}\Big(1 + 1_{\{d'=4\alpha\}}\log\big[2 + (r+s)^{-1}\big]\Big). \tag{2.34}$$

When $d' = \infty$ or $\alpha = 0$, we have $I(r+s) = [(1\wedge(r+s)]^{-\frac{d}{4}}$ and

$$\frac{1}{t}\int_0^t \mathrm{d}t_1\int_{t_1}^t \mathrm{e}^{-\lambda'(t_2-t_1)}\big[(r+s)\wedge 1\big]^{-\frac{d}{4}}\mathrm{d}t_2 \leq c_5 I(r+s)$$

holds for some constant $c_5>0$. So, in any case,

$$\frac{1}{t}\int_0^t \mathrm{d}t_1\int_{t_1}^t \mathrm{e}^{-\lambda'(t_2-t_1)}\big[(r+s)\wedge 1+1_{\{\alpha>0,d'<\infty\}}\{1\wedge(t_2-t_1)\}^{\frac{d'}{\alpha d}}\big]^{-\frac{d}{4}}\mathrm{d}t_2\le c_5 I(r+s).$$

Combining this with (2.27) and (2.28), and choosing $h(r+s)=\frac{(r+s)^{\beta+\frac{d}{8}}}{\sqrt{I(s+r)}}$, we find a constant $c_6>0$ such that

$$\mathbb{E}^\mu[\mathbb{W}_p(\mu_{t,r}^B,\mu)^2]\le\frac{c_6}{t}\bigg(\int_0^\infty\frac{\mathrm{e}^{-\lambda' s}\sqrt{I(r+s)}}{[1\wedge(r+s)]^{\beta+\frac{d}{8}}}\mathrm{d}s\bigg)^2. \tag{2.35}$$

By (2.34) and the definition of $K_{\beta,d,d',\alpha}$, we find a constant $c_7>0$ such that

$$\begin{aligned}\int_0^\infty\frac{\mathrm{e}^{-\lambda' s}\sqrt{I(r+s)}}{[1\wedge(r+s)]^{\beta+\frac{d}{8}}}\mathrm{d}s&\le\int_0^\infty\frac{\mathrm{e}^{-\lambda' s}}{[1\wedge(r+s)]^{K_{\beta,d,d',\alpha}}}\Big(1+1_{\{d'=4\alpha\}}\sqrt{\log[2+(r+s)^{-1}]}\Big)\mathrm{d}s\\&\le c_7\eta(r),\qquad r\in(0,1],\end{aligned}$$

where

$$\eta(r):=\begin{cases}1, & \text{if } K_{\beta,d,d',\alpha}<1,\\ \log(2+r^{-1}), & \text{if } K_{\beta,d,d',\alpha}=1, d'\ne 4\alpha,\\ [\log(2+r^{-1})]^{\frac{3}{2}}, & \text{if } K_{\beta,d,d',\alpha}=1, d'=4\alpha,\\ r^{1-K_{\beta,d,d',\alpha}}, & \text{if } K_{\beta,d,d',\alpha}>1, d'\ne 4\alpha,\\ r^{1-K_{\beta,d,d',\alpha}}\sqrt{\log(2+r^{-1})}, & \text{if } K_{\beta,d,d',\alpha}>1, d'=4\alpha.\end{cases}$$

This together with (2.35) and (2.25) implies

$$\mathbb{E}^\mu[\mathbb{W}_p(\mu_t,\mu)^2]\le 2\inf_{r\in(0,1]}\big\{\mathbb{E}^\mu[\mathbb{W}_p(\mu_{t,r}^B,\mu)^2]+\mathbb{E}^\mu[\mathbb{W}_p(\mu_{t,r}^B,\mu_t^B)^2]\big\}\le c_8\inf_{r\in(0,1]}\big\{t^{-1}\eta(r)^2+r\big\}$$

for some constant $c_8>0$ and all $t>0$. Therefore, (2.12) holds for some constant $c>0$.

(e) To prove (2.13), let $t>1$ and $\bar\mu_{t-1}^B:=\frac{1}{t-1}\int_1^t\delta_{X_s^B}\mathrm{d}s$, so that $\mu_t^B=\frac{1}{t}\int_0^1\delta_{X_s^B}\mathrm{d}s+\frac{t-1}{t}\bar\mu_{t-1}^B$. Then

$$\mathbb{W}_p(\mu_t^B,\mu)\le\frac{1}{t}\int_0^1\big[\mu\big(\rho(X_s^B,\cdot)^p\big)\big]^{\frac{1}{p}}\mathrm{d}s+\frac{t-1}{t}\mathbb{W}_p(\bar\mu_{t-1}^B,\mu).$$

By Jensen's inequality and the Markov property, this implies

$$\begin{aligned}&\mathbb{E}^x\big[\mathbb{W}_p(\mu_t^B,\mu)^q\big]\le\frac{2^{q-1}}{t^q}\mathbb{E}^x\bigg(\int_0^1\mu\big(\rho(X_s^B,\cdot)^p\big)^{\frac{1}{p}}\mathrm{d}s\bigg)^q+2^{q-1}\mathbb{E}^x\big[\mathbb{W}_p(\bar\mu_{t-1}^B,\mu)^q\big]\\&\le\frac{2^{q-1}}{t^q}\int_0^1\mathbb{E}^x\Big[\mu\big(\rho(X_s^B,\cdot)^p\big)^{\frac{q}{p}}\Big]\mathrm{d}s+2^{q-1}\mathbb{E}^{\nu_x}\big[\mathbb{W}_p(\mu_{t-1}^B,\mu)^q\big],\end{aligned}\tag{2.36}$$

where $\nu_x:=p_1^B(x,\cdot)\mu$ is the distribution of X_1^B for $X_0^B=x$. By Hölder's inequality we obtain

$$\begin{aligned}&\mathbb{E}^{\nu_x}\big[\mathbb{W}_p(\mu_{t-1}^B,\mu)^q\big]=\int_M p_1^B(x,y)\mathbb{E}^y\big[\mathbb{W}_p(\mu_{t-1}^B,\mu)^q\big]\mu(\mathrm{d}y)\\&\le\|p_1^B(x,\cdot)\|_{L^{\frac{2}{2-q}}(\mu)}\bigg(\int_M\mathbb{E}^y\big[\mathbb{W}_p(\mu_{t-1}^B,\mu)^2\big]\mu(\mathrm{d}y)\bigg)^{\frac{q}{2}}=\|p_1^B(x,\cdot)\|_{L^{\frac{2}{2-q}}(\mu)}\big(\mathbb{E}^\mu[\mathbb{W}_p(\mu_{t-1}^B,\mu)^2]\big)^{\frac{q}{2}}.\end{aligned}$$

Combining this with (2.36), we deduce (2.13).

2.5 An extension

For some infinite-dimensional models, see for instance [32], (2.5) fails for any $d \in (0,\infty)$, but there may be a decreasing function $\gamma : (0,\infty) \to (0,\infty)$ such that

$$\int_M \hat{p}_t(x,x)\mu(\mathrm{d}x) \le \gamma(t), \quad t > 0.$$

Moreover, in case that P_t is not L^2-exponential ergodic, by the weak Poincaré inequality which holds for a broad class of ergodic Markov processes, see [21], we have

$$\lim_{t\to\infty} \|P_t - \mu\|_{L^\infty(\mu)\to L^2(\mu)} = 0.$$

To cover these two situations for which Theorem 2.2 does not apply, we present the following result for the empirical measure μ_t of the Markov process X_t with semigroup P_t.

Theorem 2.3. *Assume* (2.3), (2.4). *If there exist a constant* $q \in [1,\infty], q' \in [\frac{q}{q-1},\infty]$ *and a decreasing function* $\gamma : (0,\infty) \to (0,\infty)$ *such that*

$$\lim_{t\to\infty} \|P_t - \mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)} = 0, \tag{2.37}$$

$$\int_M \|\hat{p}_{\frac{r}{2}}(y,\cdot)\|_{L^q(\mu)} \|\hat{p}_{\frac{r}{2}}(y,\cdot)\|_{L^{q'}(\mu)} \mu(\mathrm{d}y) \le \gamma(r), \quad r > 0. \tag{2.38}$$

Then there exists a constant $c > 0$ *such that for any* $t > 0$,

$$\mathbb{E}^\mu[\mathbb{W}_p(\mu_t,\mu)^2] \le c \inf_{r\in(0,1]} \left\{ \frac{\int_0^t \|P_s - \mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)} \mathrm{d}s}{t} \left(\int_0^1 \frac{\sqrt{\gamma(r+s)}}{(r+s)^\beta} \mathrm{d}s \right)^2 + r \right\}.$$

Proof. Let $B(\lambda) = \lambda$ so that $X_t^B = X_t, P_t^B = P_t$ and $\mu_t^B = \mu_t$. Noting that $\hat{p}_r(x,\cdot) = \int_M \hat{p}_{\frac{r}{2}}(x,y)\hat{p}_{\frac{r}{2}}(y,\cdot)\mu(\mathrm{d}y)$, by (2.38) we obtain

$$\begin{aligned}
&\mathbb{E}^\mu[\hat{p}_r(X_{t_1},X_{t_2}) - 1] = \int_M (P_{t_2-t_1} - \mu)\hat{p}_r(x,\cdot)(x)\mu(\mathrm{d}x) \\
&= \int_{M\times M} \hat{p}_{\frac{r}{2}}(x,y)(P_{t_2-t_1} - \mu)\hat{p}_{\frac{r}{2}}(y,\cdot)(x)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \\
&\le \int_M \|\hat{p}_{\frac{r}{2}}(\cdot,y)\|_{L^q(\mu)} \|(P_{t_2-t_1} - \mu)\hat{p}_{\frac{r}{2}}(y,\cdot)\|_{L^{\frac{q}{q-1}}(\mu)} \mu(\mathrm{d}y) \\
&\le \|P_{t_2-t_1} - \mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)} \int_M \|\hat{p}_{\frac{r}{2}}(y,\cdot)\|_{L^q(\mu)} \|\hat{p}_{\frac{r}{2}}(y,\cdot)\|_{L^{q'}(\mu)} \mu(\mathrm{d}y) \\
&\le \gamma(r)\|P_{t_2-t_1} - \mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)}, \quad r > 0, t_2 > t_1.
\end{aligned}$$

Combining this with (2.37), we find a constant $c_1 > 0$ such that

$$\int_0^t \mathrm{d}t_1 \int_{t_1}^t \mathbb{E}^\mu[\hat{p}_{r+s}(X_{t_1},X_{t_2}) - 1]\mathrm{d}t_2 \le c_1\gamma(r+s)t \int_0^t \|P_s - \mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)} \mathrm{d}s, \quad r,s > 0.$$

So, by (2.27) with

$$h(r+s) = (s+r)^\beta\bigg(\gamma(r+s)t\int_0^t \|P_s-\mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)}\mathrm{d}s\bigg)^{-\frac{1}{2}},$$

we find constants $c_2, c_3 > 0$ such that

$$\begin{aligned}\mathbb{E}^\mu[\mathbb{W}_p(\mu_{t,r},\mu)^2] &\le \frac{c_2}{t}\bigg(\int_0^\infty \frac{\mathrm{e}^{-\frac{\lambda}{2}(r+s)}\sqrt{\gamma(r+s)}}{(r+s)^\beta}\mathrm{d}s\bigg)^2\int_0^t \|P_s-\mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)}\mathrm{d}s\\ &\le \frac{c_3}{t}\bigg(\int_0^1 \frac{\sqrt{\gamma(r+s)}}{(r+s)^\beta}\mathrm{d}s\bigg)^2\int_0^t \|P_s-\mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)}\mathrm{d}s,\quad t>0, r\in(0,1].\end{aligned}$$

This together with (2.25) and the triangle inequality implies the desired estimate. □

To verify Theorem 2.3, we present below a simple example where P_t only has algebraic convergence in $\|\cdot\|_{L^\infty(\mu)\to L^2(\mu)}$, so Theorem (2.2) does not apply.

Example 2.3. Let $M=[0,1], \rho(x,y)=|x-y|$ and $\mu(\mathrm{d}x)=\mathrm{d}x$. For any $l\in(2,\infty)$, let X_t be the diffusion process on $M\setminus\{0,1\}$ generated by

$$L := \big\{x(1-x)\big\}^l\frac{\mathrm{d}^2}{\mathrm{d}x^2} + l\big\{x(1-x)\big\}^{l-1}(1-2x)\frac{\mathrm{d}}{\mathrm{d}x}.$$

Then Theorem 2.2 does not apply, but by Theorem 2.3 there exists a constant $c>0$ such that for any $t>0$,

$$\mathbb{E}^\mu[\mathbb{W}_p(\mu_t,\mu)^2] \le c\begin{cases} t^{-1}, & \text{if } l\in(2,5), p\in[2,\frac{13-l}{4}),\\ t^{-1}[\log(2+t)]^3, & \text{if } l\in(2,5], p=\frac{13-l}{4},\\ \big[t^{-1}\log(2+t)\big]^{\frac{8}{4p+l-5}}, & \text{if } l\in(2,5], p>\frac{13-l}{4},\\ t^{-\frac{4}{l-1}}[\log(2+t)]^2, & \text{if } l>5, p=2\\ t^{-\frac{8}{p(l-1)}}, & \text{if } l>5, p>2.\end{cases} \tag{2.39}$$

Proof. We first observe that (2.6) fails, so that Theorem 2.2 does not apply. Indeed, the Dirichlet form of L satisfies

$$\mathscr{E}(f,g) = \int_0^1 \big\{x(1-x)\big\}^l (f'g')(x)\mathrm{d}x,\quad f,g\in C_b^1(M)\subset\mathscr{D}(\mathscr{E}). \tag{2.40}$$

Let ρ_L be the intrinsic distance function to the point $\frac{1}{2}\in M$. We find a constant $c_1>0$ such that

$$\rho_L(x) = \bigg|\int_{\frac{1}{2}}^x \big\{s(1-s)\big\}^{-\frac{l}{2}}\mathrm{d}s\bigg| \ge c_1\big(x^{1-\frac{l}{2}}+(1-x)^{1-\frac{l}{2}}\big),\quad x\in M.$$

Then for any $\varepsilon>0$, we have $\mu(\mathrm{e}^{\varepsilon\rho_L})=\infty$, so that by [1], $\mathrm{gap}(L)=0$. On the other hand, since L is symmetric in $L^2(\mu)$, by [21, Lemma 2.2], (2.6) implies the same inequality for $k=1$, so that $\mathrm{gap}(L)\ge\lambda>0$. Hence, (2.6) fails.

To apply Theorem 2.3, let $\hat{P}_t$ be the standard Neumann heat semigroup on M generated by Δ. It is classical that (2.3) and (2.4) hold for

$$\beta = \frac{1}{2} + \frac{p-2}{4}. \tag{2.41}$$

Moreover, there exists a constant $c_2 > 1$ such that

$$\|\hat{P}_{\frac{r}{2}}\|_{L^m(\mu)\to L^n(\mu)} \le c_2(1 + r^{-\frac{n-m}{2nm}}), \quad 1 \le m \le n \le \infty, r > 0,$$

so that for $q' = \infty$ and $q > 1$,

$$\begin{aligned}&\|\hat{p}_{\frac{r}{2}}(y,\cdot)\|_{L^q(\mu)}\|\hat{p}_{\frac{r}{2}}(y,\cdot)\|_{L^{q'}(\mu)} \le c_2\|\hat{P}_{\frac{r}{2}}\|_{L^1(\mu)\to L^q(\mu)}(1 + r^{-\frac{1}{2}})\\ &\le c_2^2(1 + r^{-\frac{q-1}{2q}})(1 + r^{-\frac{1}{2}}), \quad r > 0.\end{aligned}$$

Hence, there exists a constant $c_3 > 0$ such that (2.38) holds for

$$\gamma(r) = c_3(1 + r^{-\frac{2q-1}{2q}}).$$

Combining this with (2.41), we find a constant $k > 0$ such that for any $r \in (0,1)$,

$$\eta(r) := \bigg(\int_0^1 \frac{\sqrt{\gamma(r+s)}}{(r+s)^\beta}\mathrm{d}s\bigg)^2 \le k\cdot\begin{cases}1, & \text{if } 1 < q < \frac{1}{p-2},\\ [\log(1+r^{-1})]^2, & \text{if } 1 < q = \frac{1}{p-2},\\ r^{\frac{1-(p-2)q}{2q}}, & \text{if } q > 1 \vee \frac{1}{p-2}.\end{cases} \tag{2.42}$$

Indeed, (2.41) implies

$$-\frac{2q-1}{4q} - \beta = -\frac{(p+2)q-1}{4q}\begin{cases}> -1, & \text{if } 1 < q < \frac{1}{p-2},\\ = -1, & \text{if } 1 < q = \frac{1}{p-2},\\ < -1, & \text{if } q > 1 \vee \frac{1}{p-2},\end{cases}$$

so that we find constants $k_1, k_2 > 0$ such that for any $r \in (0,1)$,

$$\int_0^1 \frac{\sqrt{\gamma(r+s)}}{(r+s)^\beta}\mathrm{d}s \le k_1\int_0^1 (r+s)^{-\frac{2q-1}{4q}-\beta}\mathrm{d}s \le k_2\cdot\begin{cases}1, & \text{if } 1 < q < \frac{1}{p-2},\\ \log(1+r^{-1}), & \text{if } 1 < q = \frac{1}{p-2},\\ r^{\frac{1-(p-2)q}{4q}}, & \text{if } q > 1 \vee \frac{1}{p-2},\end{cases}$$

which implies (2.41).

To calculate $\|P_t - \mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)}$ for $q' = \infty$, we apply the weak Poincaré inequality studied in [21]. Let

$$M_s = [s, 1-s], \quad s \in (0, 1/2).$$

Noting that $\mu(\mathrm{d}x) = \mathrm{d}x$ and letting $\nu(\mathrm{d}x) = \{x(1-x)\}^l\mathrm{d}x$, we find a constant $c_4 > 0$ such that

$$\sup_{r\in[s,\frac{1}{2}]} \mu([r,1/2])\nu([s,r]) \le 2^l \sup_{r\in[s,\frac{1}{2}]}\Big(\frac{1}{2} - r\Big)\big(s^{1-l} - r^{1-l}\big) \le c_4 s^{1-l}, \quad s \in (0,1/2).$$

By the weighted Hardy inequality [20], see for instance [28, Proposition 1.4.1], we have

$$\mu(f^2 1_{[s,\frac{1}{2}]}) \le 4c_4 s^{1-l}\nu(|f'|^2), \quad f \in C^1([s,1/2]), f(1/2)=0.$$

By symmetry, the same holds for $[\frac{1}{2}, 1-s]$ replacing $[s,\frac{1}{2}]$. So, according to [28, Lemma 1.4.3], see also [4], we derive

$$\mu(f^2 1_{M_s}) \le 4c_4 s^{1-l}\nu(|f'|^2 1_{M_s}) + \mu(f 1_{M_s})^2, \quad f \in C^1([s, s-1]).$$

Combining this with (2.40), for any $f \in C_b^1(M)$ with $\mu(f)=0$, we have $\mu(f1_{M_s}) = -\mu(f1_{M_s^c})$ so that

$$\begin{aligned}\mu(f^2) &= \mu(f^2 1_{M_s^c}) + \mu(f^2 1_{M_s}) \le \mu(f^2 1_{M_s^c}) + 4c_4 s^{1-l}\mathscr{E}(f,f) + \mu(f 1_{M_s^c})^2,\\ &\le 4c_4 s^{1-l}\mathscr{E}(f,f) + 2\|f\|_\infty^2 \mu(M_s^c)^2 \le 4c_4 s^{1-l}\mathscr{E}(f,f) + 8s^2\|f\|_\infty^2, \quad s\in(0,1/2).\end{aligned}$$

For any $r \in (0,1)$, let $s = (r/8)^{\frac{1}{2}}$. We find a constant $c_5>0$ such that

$$\mu(f^2) \le c_5 r^{-\frac{l-1}{2}}\mathscr{E}(f,f) + r\|f\|_\infty^2, \quad r\in(0,1), \mu(f)=0, f\in C_b^1(M).$$

By [21, Corollary 2.4(2)], this implies

$$\|P_t - \mu\|_{L^\infty(\mu)\to L^2(\mu)} = \|P_t-\mu\|_{L^2(\mu)\to L^1(\mu)} \le c_5(1+t)^{-\frac{2}{l-1}}, \quad t>0$$

for some constant $c_5>0$. Since P_t is contractive in $L^n(\mu)$ for any $n \ge 1$, this together with the interpolation theorem implies

$$\|P_t - \mu\|_{L^\infty(\mu)\to L^{\frac{q}{q-1}}(\mu)} \le c_6(1+t)^{-\frac{4(q-1)}{q(l-1)}}, \quad t>0.$$

Noting that $q' = \infty$, we find a constant $k>0$ such that

$$\Gamma(t) := \frac{1}{t}\int_0^t \|P_s-\mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)} \mathrm{d}s \le k \begin{cases} t^{-1}, & \text{if } l\in(2,5), q>\frac{4}{5-l},\\ t^{-1}\log(2+t), & \text{if } l=5, q=\infty,\\ (1+t)^{-\frac{4}{l-1}}, & \text{if } l>5, q=\infty.\end{cases} \tag{2.43}$$

We now prove the desired estimates case by case.

(1) Let $l\in(2,5)$ and $p\in[2,\frac{13-l}{4})$. Taking $q\in(\frac{4}{5-l},\frac{1}{p-2})$ in (2.42) and (2.43), we obtain

$$\inf_{r\in(0,1]}\big\{\eta(r)\Gamma(t)+r\big\} \le k\inf_{r\in(0,1]}\big\{t^{-1}+r\big\} = kt^{-1}.$$

So, the desired estimate follows from Theorem 2.3.

(2) Let $l\in(2,5]$ and $p=\frac{13-l}{4}$. Taking $q=\frac{4}{5-l}=\frac{1}{p-2}$ in (2.42) and (2.43) we find a constant $c>0$ such that

$$\inf_{r\in(0,1]}\big\{\eta(r)\Gamma(t)+r\big\} \le k\inf_{r\in(0,1]}\big\{t^{-1}[\log(2+t)][\log(1+r^{-1})]^2+r\big\} \le ct^{-1}[\log(2+t)]^3.$$

This implies the desired estimate according to Theorem 2.3.

(3) Let $l \in (2,5]$ and $p > \frac{13-l}{4}$. We have $q := \frac{4}{5-1} > \frac{1}{p-2}$, so that (2.42) and (2.43) imply

$$\inf_{r\in(0,1]} \big\{\eta(r)\Gamma(t)+r\big\} \le k \inf_{r\in(0,1]} \big\{t^{-1}[\log(2+t)]r^{-\frac{4p+l-13}{8}}+r\big\} \le c\big[t^{-1}\log(2+t)\big]^{-\frac{4p+l-5}{8}}$$

for some constant $c>0$, which implies the desired estimate by Theorem 2.3.

(4) Let $l>5$ and $p=2$. By taking $q=\infty$ in (2.42) and (2.43), we find a constant $c>0$ such that

$$\inf_{r\in(0,1]} \big\{\eta(r)\Gamma(t)+r\big\} \le k \inf_{r\in(0,1]} \big\{t^{-\frac{4}{l-1}}[\log(1+r^{-1})]^2+r\big\} \le ct^{-\frac{4}{l-1}}[\log(2+t)]^2.$$

By Theorem 2.3, the desired estimate holds.

(5) Let $l>5$ and $p>2$. By taking $q=\infty$ we find a constant $c>0$ such that (2.42) and (2.43) imply

$$\inf_{r\in(0,1]} \big\{\eta(r)\Gamma(t)+r\big\} \le k \inf_{r\in(0,1]} \big\{t^{-\frac{4}{l-1}}r^{-\frac{p-2}{2}}+r\big\} \le ct^{-\frac{8}{p(l-1)}}$$

for some consatnt $c>0$. Hence the desired estimate holds according to Theorem 2.3.

□

3 Subordinate stochastic Hamiltonian systems

Consider the following degenerate SDE for $X_t=(X_t^{(1)},X_t^{(2)})$ on $\mathbb{R}^{n+m}=\mathbb{R}^n\times\mathbb{R}^m$ ($n,m\ge 1$ may be different):

$$\text{(3.1)} \qquad \begin{cases} \mathrm{d}X_t^{(1)} = \kappa Q X_t^{(2)}\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = \sqrt{2}\,\mathrm{d}W_t - \big\{Q^*(\nabla V)(X_t^{(1)})+\kappa X_t^{(2)}\big\}\mathrm{d}t, \end{cases}$$

where W_t is the m-dimensional Brownian motion, $Q\in\mathbb{R}^{n\otimes m}$, $\kappa>0$ is a constant, and $V\in C^2(\mathbb{R}^n)$ satisfies

$$\text{(3.2)} \qquad \sup_{x_1\in\mathbb{R}^n}\frac{\|\nabla^2V(x_1)\|}{1+|\nabla V(x_1)|}<\infty, \quad \int_{\mathbb{R}^n}|\nabla V(x_1)|^2\mathrm{e}^{-V(x_1)}\mathrm{d}x_1<\infty.$$

Let

$$\mu_V(\mathrm{d}x_1) := \frac{\mathrm{e}^{-V(x_1)}\mathrm{d}x_1}{\int_{\mathbb{R}^n}\mathrm{e}^{-V(x_1)}\mathrm{d}x_1}, \quad \mathscr{N}_\kappa(\mathrm{d}x_2) := \Big(\frac{\kappa}{2\pi}\Big)^{\frac{m}{2}}\mathrm{e}^{-\frac{\kappa}{2}|x_2|^2}\mathrm{d}x_2.$$

Then the SDE (3.1) is well-posed, and the solution has invariant probability measure

$$\text{(3.3)} \qquad \mu(\mathrm{d}x_1,\mathrm{d}x_2) := \mu_V(\mathrm{d}x_1)\mathscr{N}_\kappa(\mathrm{d}x_2).$$

Recall that for a metric space (M,ρ),

$$B(x,r) := \big\{y\in M:\ \rho(x,y)<r\big\}, \quad x\in M, r>0.$$

We will verify (A_1) and (2.10) for the present model by using the following assumption.

(A_2) *QQ^* is invertible,* (3.2) *holds, and there exist constants $k>0$ and $n'\ge n$ such that*

$$\nabla^2 V \ge -kI_n, \quad \int_{\mathbb{R}^n} \frac{\mu_V(\mathrm{d}x_1)}{\mu_V(B(x_1,r))} \le kr^{-n'}, \quad r\in(0,1], \tag{3.4}$$

$$\mu_V(f^2)\le k\mu_V(|\nabla f|^2), \quad f\in C_b^1(\mathbb{R}^n),\ \mu_V(f)=0. \tag{3.5}$$

We have the following result.

Theorem 3.1. *Assume (A_2), let $B\in \mathbf{B}^\alpha$ for some $\alpha\in[0,1]$, and let $\rho(x,y)=|x-y|$ for $x,y\in\mathbb{R}^{n+m}$. Let $\xi_t(K)$ be in* (2.14) *for*

$$K := \begin{cases} \frac{1}{2}+\frac{n'+2m}{4}-\frac{\alpha(n'+2m)}{2(3n'+2m)}, & \text{if } \|\nabla^2 V\|_\infty<\infty\\ \frac{1}{2}+\frac{n'+2m}{4}, & \text{if } \|\nabla^2 V\|_\infty=\infty. \end{cases} \tag{3.6}$$

Then there exists a constant $c>0$ such that

$$\mathbb{E}^\mu\big[\mathbb{W}_2(\mu_t^B,\mu)^2\big]\le c\xi_t(K), \quad t>0. \tag{3.7}$$

If $\|\nabla^2 V\|_\infty<\infty$, then for any $t\ge 2$ and $x\in\mathbb{R}^{n+m}$,

$$\big[\mathbb{E}^x \mathbb{W}_2(\mu_t^B,\mu)\big]^2 \le c\xi_t(K)\mathbb{E}^x\bigg[\int_0^1 |X_s^B|^2\mathrm{d}s + \frac{1}{\mu\big(B(x_1,(1\wedge S_1^B)^{\frac{3}{2}})\times B(x_2,(1\wedge S_1^B)^{\frac{1}{2}})\big)}\bigg]. \tag{3.8}$$

To prove this result, we first present a dimension-free Harnack inequality for the following more general model:

$$\begin{cases} \mathrm{d}X_t^{(1)} = \big\{AX_t^{(1)}+QX_t^{(2)}\big\}\mathrm{d}t,\\ \mathrm{d}X_t^{(2)} = Z_t(X_t)\mathrm{d}t+\sigma_t\mathrm{d}W_t, \end{cases} \tag{3.9}$$

where Q and W_t are in (3.1), $A\in\mathbb{R}^{n\otimes n}$, and

$$\sigma:[0,\infty)\to\mathbb{R}^{m\otimes m}, \quad Z:[0,\infty)\times\mathbb{R}^{n+m}\to\mathbb{R}^m$$

are measurable such that the following conditions hold:

(A_3) *There exist a constant $k>0$ and an integer $k_0\ge 0$ such that*

$$\sup_{t\ge0}\|\sigma_t^{-1}\|_\infty + \sup_{t\ge 0, x\ne y}\frac{|Z_t(x)-Z_t(y)|}{|x-y|}\le k, \quad \mathrm{Rank}[A^iQ: 0\le i\le k_0]=n.$$

Lemma 3.2. *Assume (A_3), and let P_t be the Markov semigroup associated with* (3.9). *Then for any $p\in(1,\infty)$, there exists a constant $c(p)>0$ such that*

$$|P_tf(y)|^p\le \big(P_t|f|^p(x)\big)\exp\bigg[\frac{c|x_1-y_1|^2}{(1\wedge t)^{4k_0+3}}+\frac{c|x_2-y_2|^2}{(1\wedge t)^{4k_0+1}}\bigg]$$

holds for all $t>0$ and $x=(x_1,x_2), y=(y_1,y_2)\in\mathbb{R}^n\times\mathbb{R}^m$.

Proof. By Jensen's inequality, we only need to prove for $t \in (0, 1]$. The proof is refined from that of [29, Lemma 3.2]. Let X_t solve (3.9) for $X_0 = x$, and for fixed $t_0 \in (0, 1]$, let Y_t solve the following SDE with $Y_0 = y$:

$$\begin{cases} \mathrm{d}Y_t^{(1)} = \big\{AY_t^{(1)} + QY_t^{(2)}\big\}\mathrm{d}t, \\ \mathrm{d}Y_t^{(2)} = \Big\{Z_t(X_t) + \dfrac{x_2 - y_2}{t_0} + \dfrac{\mathrm{d}}{\mathrm{d}t}\big[t(t_0 - t)Q^*\mathrm{e}^{(t_0-t)A^*}b_{t_0}\big]\Big\}\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad t \in [0, t_0], \end{cases}$$

where

$$\begin{aligned} b_{t_0} &:= Q_{t_0}^{-1}\Big\{\mathrm{e}^{t_0 A}(x_1 - y_1) + \int_0^{t_0} \frac{t_0 - s}{t_0}\mathrm{e}^{(t_0-s)A}Q^*(x_2 - y_2)\mathrm{d}s\Big\}, \\ Q_t &:= \int_0^t s(t - s)\mathrm{e}^{(t-s)A}QQ^*\mathrm{e}^{(t-s)A^*}\mathrm{d}s, \quad t > 0. \end{aligned}$$

By [29, (3.2) and (3.3)], we have $X_{t_0} = Y_{t_0}$ and

$$\sup_{t\in[0,t_0]} |X_t - Y_t| \le c|x - y|, \quad x, y \in \mathbb{R}^{n+m} \tag{3.10}$$

holds for some constant $c > 0$.

According to the proof of [38, Theorem 4.2], the rank condition in (A_3) implies

$$\|Q_{t_0}^{-1}\| \le c_1 t_0^{-2k_0-3}, \quad t_0 \in (0, 1]$$

for some constant $c_1 > 0$. Then there exists a constant $c_2 > 0$ such that

$$|b_{t_0}| \le c_2 t_0^{-2k_0-3}\big(|x_1 - y_1| + t_0|x_2 - y_2|\big). \tag{3.11}$$

Combining this with the first condition in (A_3), we see that

$$\psi_t := \sigma_t^{-1}\bigg(Z_t(X_t) - Z_t(Y_t) + \frac{x_2 - y_2}{t_0} + \frac{\mathrm{d}}{\mathrm{d}t}\Big\{t(t_0 - t)Q^*\mathrm{e}^{(t_0-t)A^*}b_{t_0}\Big\}\bigg)$$

satisfies

$$\sup_{t\in[0,t_0]} |\psi_t|_\infty^2 \le c_4\Big(\frac{|x_1 - y_1|^2}{t_0^{4k_0+4}} + \frac{|x_2 - y_2|^2}{t_0^{4k_0+2}}\Big), \quad t_0 \in (0, 1] \tag{3.12}$$

for some constant $c_4 > 0$. So, by Girsanov's theorem, under the probability measure $R\mathrm{d}\mathbb{P}$, where

$$R := \exp\bigg[-\int_0^{t_0}\langle\psi_t, \mathrm{d}W_t\rangle - \frac{1}{2}\int_0^{t_0}|\psi_t|^2\mathrm{d}t\bigg],$$

the process $(Y_t)_{t\in[0,t_0]}$ is a weak solution to (3.9) with initial value y. Combining this together with $X_{t_0} = Y_{t_0}$ as observed above, we find a constant $c_5 > 0$ depending on p such that

$$\begin{aligned} |P_{t_0}f(y)|^p &= |\mathbb{E}[Rf(Y_{t_0})]|^p = |\mathbb{E}[Rf(X_{t_0})]|^p \le \big(\mathbb{E}[R^{\frac{p}{p-1}}]\big)^{p-1}\mathbb{E}[|f|^p(X_{t_0})] \\ &\le \mathrm{e}^{c_5\int_0^{t_0}\|\psi_t\|_\infty^2\mathrm{d}t}P_{t_0}|f|^p(x). \end{aligned}$$

By (3.12), this implies the desired Harnack inequality. □

Proof of Theorem 3.1. Let $p = 2, M := \mathbb{R}^{n+m}$. To apply Theorem 2.2, let $\hat{X}_t$ be the diffusion process generated by

$$\hat{L} := \Delta - (\nabla H) \cdot \nabla, \tag{3.13}$$

where

$$H(x) := V(x_1) + \frac{\kappa}{2}|x_2|^2, \quad x = (x_1, x_2) \in \mathbb{R}^{n+m}. \tag{3.14}$$

In the following, we verify (A_1) and Definition 2.1 for

$$\beta = \frac{1}{2}, \quad d = n' + 2m, \quad d' = \begin{cases} 3n' + 2m, & \text{if } \|\nabla^2 V\|_\infty < \infty, \\ \infty, & \text{otherwise.} \end{cases} \tag{3.15}$$

(a) Verify (2.3). By (3.3), (3.5) and the Poincaré inequality for the Gaussian measure $\mathscr{N}_\kappa$, we find a constant $C > 0$ such that

$$\mu(f^2) \le C\mu(|\nabla f|^2) + \mu(f)^2, \quad f \in C_b^1(\mathbb{R}^{n+m}). \tag{3.16}$$

Consequently, $\mathrm{gap}(\hat{L}) \ge C^{-1} > 0$, so that Proposition 2.1 implies (2.3) for $p = 2$ and $\beta = \frac{1}{2}$.

(b) Verify (2.4) and (2.6). By (3.4) and (3.14), there exists a constant $c_1 > 0$ such that

$$\nabla^2 H \ge -c_1 I_{n+m}. \tag{3.17}$$

Then as in (2.20), we find a constant $c_2 > 0$ such that

$$\hat{L}|x - \cdot|^2 \le c_2(1 + |\nabla V(x_1)|^2 + |x - \cdot|^2), \quad x \in \mathbb{R}^{n+m}.$$

This implies

$$\hat{P}_t|x - \cdot|^2(x) = \mathbb{E}^x|x - \hat{X}_t|^2 \le c_2\big(1 + \nabla V(x_1)|^2\big)t\mathrm{e}^{c_2 t}, \quad x \in \mathbb{R}^{n+m}, \ t > 0. \tag{3.18}$$

Combining this with (3.2) and (3.3), we verify condition (2.4) for $p = 2$ and some constant $k > 0$. Moreover, according to [12], (3.2) and (3.5) imply (2.6) for some constants $k, \lambda > 0$.

(c) Verify (2.5). According to [28, Theorem 2.4.4], by (3.17) we find a constant $c_3 > 0$ such that

$$\hat{p}_r(x, x) \le \frac{c_3}{\mu(B(x, \sqrt{r}))}, \quad x \in \mathbb{R}^{n+m}, r \in (0, 1]. \tag{3.19}$$

Combining this with (3.3), (3.4) and $B(x, \sqrt{r}) \subset B(x_1, \sqrt{r/2}) \times B(x_2, \sqrt{r/2})$, we find a constant $c_4 > 0$ such that

$$\int_{\mathbb{R}^{n+m}} \frac{\mu(\mathrm{d}x)}{\mu(B(x, \sqrt{r}))} \le c_4 r^{-\frac{n'}{2}} \int_{\mathbb{R}^m} \frac{\mathrm{e}^{-\kappa|x_2|^2}\mathrm{d}x_2}{\int_{B(x_2, \sqrt{r/2})} \mathrm{e}^{-\kappa|y_2|^2}\mathrm{d}y_2}, \quad r \in (0, 1]. \tag{3.20}$$

By the same argument leading to (2.17), we find a constant $c_5 > 0$ such that

$$\int_{\mathbb{R}^m} \frac{\mathrm{e}^{-\kappa|x_2|^2}\mathrm{d}x_2}{\int_{B(x_2, r)} \mathrm{e}^{-\kappa|y_2|^2}\mathrm{d}y_2} \le c_5 r^{-2m}, \quad r \in (0, 1]. \tag{3.21}$$

Combining this with (3.20), we find a constant $c > 0$ such that

$$\int_{\mathbb{R}^{n+m}} \frac{\mu(\mathrm{d}x)}{\mu(B(x,\sqrt{r}))} \le c r^{-\frac{n'+2m}{2}}, \quad r \in (0,1]. \tag{3.22}$$

Since $\hat{p}_t(x,x)$ is decreasing in $t > 0$, this together with (3.19) implies (2.5) for $d = n' + 2m$.

(d) To estimate d', we assume $\|\nabla^2 V\|_\infty < \infty$. By Lemma 3.2 for $p = 2$, where $k_0 = 0$ holds for the present model, we find a constant $c_6 > 0$ such that

$$|P_t f(x)|^2 \le (P_t f^2(y)) \mathrm{e}^{c_6(1\wedge t)^{-3}|x_1-y_1|^2 + c_6(1\wedge t)^{-1}|x_2-y_2|^2}, \quad t > 0, x, y \in \mathbb{R}^{n+m}.$$

Choosing $f := p_t(x,\cdot) \wedge l$, we derive

$$\begin{aligned}&\bigg(\int_{\mathbb{R}^{n+m}} (p_t(x,\cdot)\wedge l)^2 \mathrm{d}\mu\bigg)^2 \mathrm{e}^{-c_6(1\wedge t)^{-3}|x_1-y_1|^2 - c_6(1\wedge t)^{-1}|x_2-y_2|^2}\\ &\le P_t(p_t(x,\cdot)\wedge l)^2(y), \quad l \ge 1.\end{aligned}$$

Integrating both sides with respect to $\mu(\mathrm{d}y)$ and noting that μ is P_t-invariant, we obtain

$$\begin{aligned}\int_{\mathbb{R}^{n+m}} (p_t(x,\cdot)\wedge l)^2 \mathrm{d}\mu &\le \frac{1}{\int_{\mathbb{R}^{n+m}} \mathrm{e}^{-c_6(1\wedge t)^{-3}|x_1-y_1|^2 - c_6(1\wedge t)^{-1}|x_2-y_2|^2} \mu(\mathrm{d}y)}\\ &\le \frac{\mathrm{e}^{2c_6}}{\mu\big(B(x_1,(1\wedge t)^{\frac{3}{2}}) \times B(x_2,(1\wedge t)^{-\frac{1}{2}})\big)}.\end{aligned}$$

Letting $l \to \infty$ we arrive at

$$\int_{\mathbb{R}^{n+m}} p_t(x,\cdot)^2 \mathrm{d}\mu \le \frac{\mathrm{e}^{2c_6}}{\mu\big(B(x_1,(1\wedge t)^{\frac{3}{2}}) \times B(x_2,(1\wedge t)^{-\frac{1}{2}})\big)}, \quad t > 0, x \in \mathbb{R}^{n+m}. \tag{3.23}$$

This together with (3.4) and (3.21) yields

$$\begin{aligned}&\int_{\mathbb{R}^{n+m}\times\mathbb{R}^{n+m}} p_t(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &\le \mathrm{e}^{2c_6} \int_{\mathbb{R}^{n+m}} \frac{\mu(\mathrm{d}x)}{\mu\big(B(x_1,(1\wedge t)^{\frac{3}{2}}) \times B(x_2,(1\wedge t)^{-\frac{1}{2}})\big)} \le c(1\wedge t)^{-\frac{3n'+2m}{2}}\end{aligned}$$

for some constant $c > 0$. Therefore, (2.10) holds for $d' = 3n' + 2m$.

(e) For K in (3.6), β, d, d' in (3.15) and $\alpha \in [0,1]$, we have

$$d' \ge 3n' + 2m \ge 5 > 4\alpha, \quad K_{\beta,d,d',\alpha} = K.$$

Then (3.7) follows from (2.12).

Next, by (3.16) we have $\mu(\mathrm{e}^{c|\cdot|}) < \infty$ for some constant $c > 0$, see for instance [1]. Combining this with (2.21), we find a constant $c_7 > 0$ such that

$$\mathbb{E}^x[\mu(\rho(X_s^B,\cdot)^2)] = \mathbb{E}^x[\mu(|X_s^B - \cdot|^2)] \le c_7(1 + \mathbb{E}^x|X_s^B|^2).$$

Moreover, (3.23) implies

$$\|p_1^B(x,\cdot)\|_{L^2(\mu)}^2 \le \mathbb{E}\bigg[\frac{\mathrm{e}^{2c_6}}{\mu\big(B(x_1,(1\wedge S_1^B)^{\frac{3}{2}}) \times B(x_2,(1\wedge S_1^B)^{-\frac{1}{2}})\big)}\bigg].$$

Then (3.8) follows from (2.13) □

Example 3.1. Consider (3.1) with invertible QQ^*, and let V be in Example 2.1. Then for any $\alpha\in[0,1]$ and $B\in\mathbf{B}^\alpha$, there exists a constant $c>0$ such that for any $t\ge 1$,

$$\mathbb{E}^\mu[\mathbb{W}_2(\mu_t^B,\mu)^2]\le c\begin{cases} t^{-\frac{(2\tau-1)(3\tau n+2\tau m-m)}{(\tau n+2\tau m-m)][3\tau n+(m-\alpha)(2\tau-1)]}}, & \text{if } \frac{1}{2}<\tau\le 1,\\ t^{-\frac{2\tau-1}{\tau n+m(2\tau-1)}}, & \text{if } \tau>1.\end{cases} \tag{3.24}$$

When $\alpha>0$ and $\tau\in(\frac{1}{2},1]$, there exists a constant $\varepsilon>0$ such that for any $t\ge 1$ and $x\in\mathbb{R}^{n+m}$,

$$\big[\mathbb{E}^x\mathbb{W}_2(\mu_t^B,\mu)\big]^2\le ct^{-\frac{(2\tau-1)(3\tau n+2\tau m-m)}{(\tau n+2\tau m-m)][3\tau n+(m-\alpha)(2\tau-1)]}}\mathrm{e}^{\kappa|x_2|^2+(1+\theta|x_1|^2)^\tau-\varepsilon|x_2|-\varepsilon|x_1|^\tau}. \tag{3.25}$$

Proof. (1) As explained in the proof of Example 2.1 that (A_2) holds for $n'=\frac{2\tau n}{2\tau-1}$ and some constant $k>0$. So, K defined in (3.6) satisfies $K>1$. It is easy to see that $\|\nabla^2V\|_\infty=\infty$ for $\tau>1$ while$\|\nabla^2V\|_\infty<\infty$ for $\tau\in(\frac{1}{2},1]$. Then estimate (3.24) follows from Theorem 3.1.

(2) Let $\alpha>0$ and $\tau\le 1$. We find a constant $c_1>0$ such that

$$\sup_{t\in[0,1]}\mathbb{E}^x[|X_t|^2]\le c_1(1+|x|^2). \tag{3.26}$$

Next, similarly to (2.19), there exists a constant $c_2>0$ such that

$$\mu\big(B(x_1,r^{\frac{3}{2}})\times B(x_2,r^{\frac{1}{2}})\big)\ge c_2r^{\frac{3n+m}{2}}\mathrm{e}^{c_2r^{\frac{3}{2}}|x_2|+c_2r^{\frac{\tau}{2}}|x_1|^\tau-\kappa|x_2|^2-(1+\theta|x_1|^2)^\tau},\quad r\in(0,1],x\in\mathbb{R}^{n+m}.$$

Combining this with (3.23), we find a constant $c_3>0$ such that

$$\begin{aligned}\int_{\mathbb{R}^{n+m}}p_t(x,\cdot)^2\mathrm{d}\mu&\le\frac{\mathrm{e}^{2c_6}}{\mu\big(B(x_1,(1\wedge t)^{\frac{3}{2}})\times B(x_2,(1\wedge t)^{-\frac{1}{2}}\big)}\\&\le c_3^2\frac{\exp[\kappa|x_2|^2+(1+\theta|x_1|^2)^\tau-c_2r^{\frac{3}{2}}|x_2|-c_2r^{\frac{\tau}{2}}|x_1|^\tau]}{(1\wedge t)^{\frac{3n+m}{2}}},\quad t>0,x\in\mathbb{R}^{n+m}.\end{aligned} \tag{3.27}$$

Consequently,

$$\begin{aligned}\mathbb{E}^x[|X_t|^2]&=\int_{\mathbb{R}^{n+m}}|y|^2p_t(x,y)\mu(\mathrm{d}y)\le\bigg(\mu(|\cdot|^4)\int_{\mathbb{R}^{n+m}}p_t(x,y)^2\mu(\mathrm{d}y)\bigg)^{\frac{1}{2}}\\&\le c_3\frac{\exp[\frac{\kappa}{2}|x_2|^2+\frac{1}{2}(1+\theta|x_1|^2)^\tau-\frac{1}{2}c_2r^{\frac{3}{2}}|x_2|-\frac{1}{2}c_2r^{\frac{\tau}{2}}|x_1|^\tau]}{(1\wedge t)^{\frac{3n+m}{4}}},\quad t>0,x\in\mathbb{R}^{n+m}.\end{aligned}$$

This together with (3.26) yields

$$\sup_{s\in[0,1]}\mathbb{E}^x[|X_s^B|^2]\le\sup_{t\ge0}\mathbb{E}^x[|X_t|^2]\le c\mathrm{e}^{\frac{\kappa}{2}|x_2|^2+\frac{1}{2}(1+\theta|x_1|^2)^\tau},\quad x\in\mathbb{R}^{n+m}. \tag{3.28}$$

Moreover, when $\alpha>0$, the second inequality in (2.33) implies

$$\mathbb{E}\big[(1\wedge S_1^B)^{-\frac{3n+m}{2}}\big]<\infty,$$

which together with (3.27) yields

$$\mathbb{E}^x\Bigg[\frac{1}{\mu\big(B(x_1,(1\wedge S_1^B)^{\frac{3}{2}})\times B(x_2,(1\wedge S_1^B)^{-\frac{1}{2}}\big)}\Bigg]\le c\mathrm{e}^{\kappa|x_2|^2+(1+\theta|x_1|^2)^\tau-\varepsilon|x_2|-\varepsilon|x_1|^\alpha}$$

for some constant $\varepsilon>0$. Combining this with (3.28), we deduce the (3.25) from that in Theorem 3.1.

□

4 Subordinate spherical velocity Langevin diffusions

In this section, we consider the following degenerate SDE on $M := \mathbb{R}^n \times \mathbb{S}^{n-1} (n \ge 2)$:

$$\begin{cases} \mathrm{d}X_t^{(1)} = X_t^{(2)}\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = -\frac{1}{n-1}\big(I_n - X_t^{(2)} \otimes X_t^{(2)}\big)\nabla V(X_t^{(1)})\mathrm{d}t + \sigma\big(I_n - X_t^{(2)} \otimes X_t^{(2)}\big) \circ \mathrm{d}W_t, \end{cases} \tag{4.1}$$

where $V \in C^2(\mathbb{R}^n)$, W_t is an n-dimensional Brownian motion, $\sigma > 0$ is a constant, and $\circ\mathrm{d}$ is the Stratonovich differential. The solution of (4.1) is called the spherical velocity Langevin diffusion process generated by

$$L := \frac{\sigma^2}{2}\Delta^{(2)} + x_2 \cdot \nabla^{(1)} - (\nabla^{(2)}\Phi) \cdot \nabla^{(2)},$$

where $\Delta^{(2)}$ and $\nabla^{(2)}$ are the Laplacian and gradient on $\mathbb{S}^{n-1}$ respectively, $\nabla^{(1)}$ is the gradient on $\mathbb{R}^n$, and

$$\Phi(x) := \frac{1}{n-1}\big(\nabla^{(1)}V(x_1)\big) \cdot x_2, \quad x = (x_1, x_2) \in M.$$

See [12] and references therein for the background of this model.

Let ρ be the Riemannian distance on $M := \mathbb{R}^n \times \mathbb{S}^{n-1}$, let V satisfy (A_2), and let

$$\mu(\mathrm{d}x) := \mu_V(\mathrm{d}x_1)\Lambda(\mathrm{d}x_2),$$

where Λ is the normalized volume measure on $\mathbb{S}^{n-1}$. We have the following result.

Theorem 4.1. *Let V satisfy (A_2) and let $B \in \mathbf{B}$. Then there exists a constant $c > 0$ such that for any $t \ge 2$,*

$$\mathbb{E}^\mu\big[\mathbb{W}_2(\mu_t^B, \mu)^2\big] \le ct^{-\frac{2}{n'+n-1}},$$

$$\big[\mathbb{E}^x\mathbb{W}_2(\mu_t^B, \mu)\big]^2 \le ct^{-\frac{2}{n'+n-1}}\mathbb{E}^x\bigg[\int_0^1 |X_s^B|^2\mathrm{d}s + \frac{1}{\mu(B(x, (1 \wedge S_1^B)^{\frac{3}{2}}))}\bigg], \quad x \in \mathbb{R}^n \times \mathbb{S}^{n-1}.$$

In particular, for V given in Example 2.1, these estimates hold for $n' = 2n$.

Proof. According to [12, Theorem 1.1], (A_2) implies (2.6) for some constants $k, \lambda > 0$. Let Δ be the Laplacian on $M := \mathbb{R}^n \times \mathbb{S}^{n-1}$. To apply Theorem 2.2, we choose the reference symmetric diffusion process generated by

$$\hat{L} := \Delta - \{\nabla^{(1)}V(x_1)\} \cdot \nabla^{(1)},$$

which is symmetric in $L^2(\mu)$. As shown in the proof of Example 3.1, (3.4) holds for $n' = 2n$. So, by Theorem 2.2 for $d' = \infty$, it suffices to verify (2.3), (2.4) and (2.5) for $p = 2$, $\beta = \frac{1}{2}$, $d = n' + n - 1$.

By (A_2) and the compactness of $\mathbb{S}^{n-1}$, the Bakry-Emery curvature of $\hat{L}$ is bounded below by a constant, and there exists a constant $\lambda > 0$ such that

$$\mu(f^2) \le \frac{1}{2\lambda}\mu(|\nabla f|^2) + \mu(f)^2, \quad f \in C_b^1(M).$$

Then as explained in steps (a)-(b) in the proof of Theorem 3.1, (2.3) and (2.4) hold for $p = 2, \beta = \frac{1}{2}$ and some constants $k > 0$.

By [28, Theorem 2.4.4] and that the Bakry-Emery curvature of $\hat{L}$ is bounded from below, we find a constant $c_1 > 0$ such that

$$\hat{p}_r(x,x) \le \frac{c_1}{\mu(B(x,\sqrt{r}))}, \quad x \in M, r \in (0,1].$$

Noting that

$$\Lambda(B(x_2,r)) \ge c_2 r^{n-1}, \quad r \in (0,1]$$

holds for some constant $c_2 > 0$, as explained in step (c) in the proof of Theorem 3.1, we derive (2.5) for $d = n' + n - 1$.

□

5 Subordinate Wright-Fisher type diffusion processes

For 1-dimensional Wright-Fisher diffusions, the convergence rate has been derived for the empirical measure with respect to the Wasserstein distance induced by the intrinsic distance. However, for higher dimensional Wright-Fisher type diffusion processes, the intrinsic distance is less explicit, so that it is hard to apply the framework introduced in [34]. Below, we consider higher dimensional Wright-Fisher type diffusion processes using Wasserstein distance induced by the Euclidean distance rather the intrinsic distance, so that the framework introduced in the present paper works well.

Let $2 \le N \in \mathbb{N}$, consider

$$\Delta^{(N)} := \Big\{ x \in [0,1]^N : |x|_1 := \sum_{i=1}^N x_1 \le 1 \Big\}.$$

Given $q = (q_i)_{1 \le i \le N+1} \in [1,\infty)^{N+1}$, the Dirichlet distribution μ with parameter q is a probability measure on $\Delta^{(n)}$ defined as follows:

$$\begin{aligned} &\mu(\mathrm{d}x) := 1_{\Delta^{(N)}}(x) h(x) \mathrm{d}x, \quad h(x) := \frac{\Gamma(|q|_1)}{\prod_{i=1}^{N+1} \Gamma(q_i)} \prod_{i=1}^{N+1} x_i^{q_i - 1}, \\ &|q|_1 := \sum_{i=1}^{N+1} q_i, \quad x_{N+1} := 1 - |x|_1, \quad x \in \Delta^{(N)}. \end{aligned} \tag{5.1}$$

The Dirichlet distribution arises naturally in Bayesian inference as conjugate priors for categorical distribution, and also arises in population genetics describing the distribution of allelic frequencies, see for instance [8] and references within.

Let X_t be either the Wright-Fisher diffusion with mutation generated by

$$\sum_{i,j=1}^N \big(x_i \delta_{ij} - x_i x_j\big) \partial_{x_i} \partial_{x_j} + \sum_{i=1}^N (q_i - |q|_1 x_i) \partial_{x_i},$$

or the diffusion process generated by

$$\sum_{i=1}^N \Big\{x_i(1-|x|_1)\partial_{x_i}^2 + \big(q_i(1-|x|_1) - q_{N+1}x_i\big)\partial_{x_i}\Big\}.$$

Then the associated Dirichlet form is determined by

$$\mathscr{E}(f,g) := \begin{cases} \mu\big(\sum_{i,j=1}^N (x_i\delta_{ij} - x_ix_j)(\partial_{x_i}f)(\partial_{x_j}g)\big), & \text{the first case,}\\ \mu\big(\sum_{i,j=1}^N x_i(1-|x|_1)(\partial_{x_i}f)(\partial_{x_j}g)\big), & \text{the second case,}\end{cases} \tag{5.2}$$

for any $f,g \in C_b^1(\Delta^{(N)})$. In both cases, the Poincaré inequality

$$\mu(f^2) \le \frac{1}{\lambda}\mathscr{E}(f,f), \quad f \in C_b^1(\Delta^{(N)}), \mu(f) = 0 \tag{5.3}$$

holds for some constant $\lambda > 0$, see [23] and [11] for the value of the largest constant λ (i.e. the spectral gap).

Theorem 5.1. *Let $p \in [2,\infty)$, $B \in \mathbf{B}^\alpha$ for some $\alpha \in [0,1]$, and $\rho(x,y) = |x-y|, x,y \in \Delta^{(N)}$. Let $\xi_t(K)$ be in (2.14) for*

$$K := \frac{1}{2} + \frac{d}{4} - \frac{\alpha d}{2d'}, \quad d := |q|_1 - 1, \quad d' := 4\sum_{i=1}^N q_i + 2q_{N+1} - 2.$$

Then there exists a constant $c > 0$ such that

$$\sup_{x\in\Delta^{(N)}} \mathbb{E}^x[\mathbb{W}_p(\mu_t^B,\mu)^2] \le c\xi_t(K), \quad t > 0.$$

To apply Theorem 2.2, let $\hat{P}_t$ be the Neumann semigroup on $\Delta^{(N)}$ generated by

$$\hat{L} := \Delta + (\nabla \log h)\cdot\nabla,$$

where h is in (5.1). The associated Dirichlet form is determined by

$$\hat{\mathscr{E}}(f,f) = \mu(|\nabla f|^2), \quad f \in C_b^1(\Delta^{(N)}).$$

To verify (2.3), we first present the following lemma on the super Poincaré inequality of $\hat{\mathscr{E}}$.

Lemma 5.2. *For the above $\hat{\mathscr{E}}$ and μ, there exists a constant $c > 0$ such that*

$$\mu(f^2) \le r\hat{\mathscr{E}}(f,f) + c\big(1 + r^{-\frac{|q|_1-1}{2}}\big)\mu(|f|)^2, \quad f \in \mathscr{D}(\hat{\mathscr{E}}), r > 0. \tag{5.4}$$

Proof. (a) We follow the idea of [37]. For any $s > 1$ and $r > 0$, let

$$D_s := \big\{x \in \Delta^{(N)}:\ \phi(x) := (1-|x|_1)^{-1} \le s\big\},$$
$$\lambda(s) := \inf\big\{\hat{\mathscr{E}}(f,f):\ f \in C^1(\Delta^{(N)}), f|_{D_s} = 0, \mu(f^2) = 1\big\},$$

$$s_r := \inf\big\{s>1:\lambda(s)\ge 8r^{-1}\big\},$$

so that $\phi(x):=(1-|x|_1)^{-1}$ satisfies

$$h(s):=\sup_{D_s}\hat{\mathscr{E}}(\phi,\phi)=Ns^4,\quad s>1.$$

According to [37, Theorem 2.1], if

$$\mu(f^2)\le r\hat{\mathscr{E}}(f,f)+\beta_s(r)\mu(|f|)^2,\quad r>0, f\in C^1(\Delta^{(N)}), f|_{D_s^c}=0 \tag{5.5}$$

holds for some $\beta_s:(0,\infty)\to(0,\infty)$, then there exists a constant $c_1>0$ such that

$$\mu(f^2)\le r\hat{\mathscr{E}}(f,f)+\beta(r)\mu(|f|)^2,\quad f\in\mathscr{D}(\hat{\mathscr{E}}), r>0 \tag{5.6}$$

holds for

$$\beta(r):=c_1+\big(2+8Nrs_r^2\big)\beta_{3s_r}\Big(\frac{r}{8+4Nrs_r^2}\Big),\quad r>0. \tag{5.7}$$

(b) Let $\mathscr{E}$ be in (5.2). By [37, Lemma 3.2], there exist constants $c_2, s_0>0$ such that

$$\inf\big\{\mathscr{E}(f,f):\ f\in C^1(\Delta^{(N)}), f|_{D_s}=0, \mu(f^2)=1\big\}\ge c_2 s,\quad s\ge s_0.$$

Noting that $\hat{\mathscr{E}}(f,f)\ge s\mathscr{E}(f,f)$ holds for $f\in C^1(\Delta^{(N)}), f|_{D_s}=0$, this implies

$$\lambda(s)\ge c_2 s^2,\quad s\ge s_0.$$

Hence, we find a constant $c_3>0$ such that

$$s_r\ge c_3(1+r^{-\frac{1}{2}}),\quad r>0. \tag{5.8}$$

(c) To estimate $\beta_s(r)$ in (5.5), we first consider the following product probability measure $\tilde{\mu}$ on $[0,1]^N$:

$$\tilde{\mu}(\mathrm{d}x):=\prod_{i=1}^N\mu_i(\mathrm{d}x_i),\quad \mu_i(\mathrm{d}x_i):=q_i x_i^{q_i-1}\mathrm{d}x_i,\quad 1\le i\le N.$$

For any $r\in(0,\frac{1}{2}]$ and $I:=[a,b]\subset[0,1]$ with $\mu_i(I)=r$, we intend to prove

$$\mu_i^\partial((\partial I)\setminus\{0,1\})=\mu_i^\partial(\{a,b\}\setminus\{0,1\})\ge q_i r^{1-q_i^{-1}}, \tag{5.9}$$

where $\mu_i^\partial(\{s\}):=q_i s^{q_i-1}$ for $s\in[0,1]$ is the boundary measure induced by μ_i. We have

$$b^{q_i}=\mu([0,b])\ge\mu(I)=r,$$

so that $b\ge r^{q_i}$. If $b<1$, then

$$\mu_i^\partial((\partial I)\setminus\{0,1\})\ge\mu_i^\partial(\{b\})=q_i b^{q_i-1}\ge q_i r^{1-q_i^{-1}}.$$

When $b = 1$, we have

$$1 - a^{q_i} = \mu_i([a,1]) = \mu_i(I) = r,$$

which implies $a \geq (1-r)^{q_i^{-1}}$ and hence, for $r \in (0, \frac{1}{2}]$,

$$\mu_i^\partial((\partial I) \setminus \{0,1\}) \geq \mu_i^\partial(\{a\}) = q_i a^{q_i - 1} \geq q_i (1-r)^{1-q_i^{-1}} \geq q_i r^{1-q_i^{-1}}.$$

In conclusion, (5.9) holds for any $r \in (0, \frac{1}{2}]$ and interval $I \subset [0,1]$ with $\mu_i(I) = r$, so that

$$\kappa(r) := \inf_{\mu_i(I) \leq r} \frac{\mu_i^\partial((\partial I) \setminus \{0,1\})}{\mu_i(I)} \geq q_i r^{-q_i^{-1}}, \quad r \in (0, 1/2].$$

This implies

$$\kappa^{-1}(2r^{-\frac{1}{2}}) := \sup\left\{r' \in (0,1/2] : \ \kappa(r') \geq 2r^{-\frac{1}{2}}\right\} \geq \frac{1}{2} \wedge \left\{q_i^{q_i} r^{\frac{q_i}{2}}\right\}, \quad r > 0.$$

According to [27, Theorem 3.4.16(1)], we find a constant $c_4 > 0$ such that this implies

$$\mu_i(f^2) \leq r\mu_i(|f'|^2) + c_4(1 + r^{-\frac{q_i}{2}})\mu_i(|f|)^2, \quad f \in C^1([0,1]), r > 0, 1 \leq i \leq N.$$

By [37, Proposition 2.2], we derive

$$\tilde{\mu}(f^2) \leq r\tilde{\mu}(|\nabla f|^2) + c_5\Big(1 + r^{-\frac{1}{2}\sum_{i=1}^N q_i}\Big)\tilde{\mu}(|f|)^2, \quad f \in C^1([0,1]^N), r > 0. \tag{5.10}$$

Now, given $f \in C^1(\Delta^{(N)})$ with $f|_{D_s^c} = 0$, take

$$g(x) := f(x)(1 - |x|_1)^{\frac{q_{N+1}-1}{2}}, \quad x \in \Delta^{(N)}.$$

By (5.10), we find constants $c_6, c_7 > 0$ such that

$$\begin{aligned}\mu(f^2) = c_6\tilde{\mu}(g^2) &\leq c_6 r_1 \tilde{\mu}(|\nabla g|^2) + c_6 c_5\Big(1 + r_1^{-\frac{1}{2}\sum_{i=1}^N q_i}\Big)\tilde{\mu}(|g|)^2 \\ &\leq c_7 r_1 \mu(|\nabla f|^2) + c_7 s^2 r_1 \mu(f^2) + c_7\Big(1 + r_1^{-\frac{1}{2}\sum_{i=1}^N q_i}\Big)s^{q_{N+1}-1}\mu(|f|)^2, \quad r_1 > 0.\end{aligned}$$

For any $r > 0$, by choosing

$$r_1 = \frac{1}{2c_7}\big(r \wedge s^{-2}\big)$$

in the above inequality, we find a constant $c_8 > 0$ such that (5.5) holds for

$$\beta_s(r) := c_8\Big(1 + (r \wedge s^{-2})^{-\frac{1}{2}\sum_{i=1}^N q_i}\Big)s^{q_{N+1}-1}, \quad s > 1, r > 0.$$

Combining this with (5.8), we find a constant $c > 0$ such that $\beta(r)$ in (5.7) satisfies

$$\beta(r) \leq c(1 + r^{-\frac{|q|_1 - 1}{2}}), \quad r > 0,$$

and hence (5.4) follows from (5.6). □

Proof of Theorem 5.1. Note that ρ is bounded. Moreover, by (5.14) below, $p_1^B = \mathbb{E}[p_{S_1^B}]$, and the second inequality in (2.33), we see that p_1^B is also bounded. Then the desired estimate follows from (2.13) for $q = 2$. Noting that $q_i \geq 1$ implies

$$d' = 4\sum_{i=1}^N q_i + 2q_{N+1} - 2 \geq 4\sum_{i=1}^N q_i \geq 8 > 4\alpha,$$

by Theorem 2.2, it suffices to verify (A_1) and (2.10) for the given constants d, d' and

$$\beta := \frac{1}{2} + \frac{d(p-2)}{4p},$$

for which we have $K_{\beta,d,d',\alpha} = K$.

(1) Let h be in (2.12). Since $q_i \geq 1$, we have $\nabla^2 \log h \leq 0$, so that the Bakry-Emery curvature of $\hat{L}$ is nonnegative. Since $\Delta^{(N)}$ is convex, according to [28, Theorem 3.3.2(11)], we have

$$|\nabla \hat{P}_t f| \leq \frac{1}{\sqrt{t}}(\hat{P}_t|f|^2)^{\frac{1}{2}}, \quad t > 0. \tag{5.11}$$

Since the Dirichlet form $\hat{\mathscr{E}}$ of $\hat{P}_t$ is larger than $\mathscr{E}$, (5.3) implies the same inequality for $\hat{\mathscr{E}}$, so that

$$\|\hat{P}_t - \mu\|_{L^2(\mu)} \vee \|P_t - \mu\|_{L^2(\mu)} \leq \mathrm{e}^{-\lambda t}, \quad t \geq 0. \tag{5.12}$$

Then (2.6) holds. Moreover, it is easy to see that (5.4) implies the Nash inequality

$$\mu(f^2) \leq c_1\big(1 + \hat{\mathscr{E}}(f,f)\big)^{\frac{d}{d+2}}, \quad f \in \mathscr{D}(\hat{\mathscr{E}}), \mu(|f|) = 1$$

for some constants $c_1 > 0$, so that by [9, Corollary 2.4.7],

$$\begin{aligned} \|\hat{P}_t\|_{L^1(\mu)\to L^\infty(\mu)} &\leq \|\hat{P}_{\frac{t}{2}}\|_{L^1(\mu)\to L^2(\mu)} \|\hat{P}_{\frac{t}{2}}\|_{L^2(\mu)\to L^\infty(\mu)} \\ &= \|\hat{P}_{\frac{t}{2}}\|^2_{L^1(\mu)\to L^2(\mu)} \leq c(1\wedge t)^{-\frac{d}{2}}, \quad t > 0 \end{aligned}$$

holds for some constant $c > 0$. Combining this with the interpolation theorem, we find a constant $c_2 > 0$ such that

$$\|\hat{P}_t\|_{L^{q_1}(\mu)\to L^{q_2}(\mu)} \leq c_2\Big(1 + t^{-\frac{d(q_2-q_1)}{2q_1q_2}}\Big), \quad t > 0, 1 \leq q_1 \leq q_2 \leq \infty. \tag{5.13}$$

Taking $q_1 = 1$ and $q_2 = \infty$ we derived (2.5), while choosing $q_1 = 2, q = p$ we deduce (2.3) with $\beta = \frac{1}{2} + \frac{d(p-2)}{4p}$ from (5.11) and (5.12).

On the other hand, by [37, Proof of (3.1)], the super Poinaré inequality

$$\mu(f^2) \leq r\mathscr{E}(f,f) + c_3(1 + r^{1-2\sum_{i=1}^N q_i - q_{N+1}})\mu(|f|)^2, \quad r > 0, f \in \mathscr{D}(\mathscr{E})$$

holds for some constant $c_3 > 0$. Noting that $2\sum_{i=1}^N q_i + q_{N+1} - 1 = \frac{d'}{2}$, as explained above, we find a constant $c > 0$ such that

$$\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} \le c(1\wedge t)^{-\frac{d'}{2}}, \quad t > 0, \tag{5.14}$$

so that (2.10) holds.

(2) It remains to verify (2.4). Since $q_i - 1 \ge 0$, for any $x, y \in \Delta^{(N)}$, we have

$$\begin{aligned}\sum_{i=1}^N (y_i - x_i)\Big(\frac{q_i - 1}{x_i} - \frac{q_{N+1} - 1}{1 - |y|_1}\Big) &= \sum_{i=1}^N \frac{(q_i - 1)(y_i - x_i)}{y_i} + \frac{(q_{N+1} - 1)(|x|_1 - |y|_1)}{1 - |y|_1} \\ &\le \sum_{i=1}^{N+1} (q_i - 1) < \infty.\end{aligned}$$

So, for any $2 \le n \in \mathbb{N}$, we find constants $c_1(n), c_2(n) > 0$ such that

$$\begin{aligned}\hat{L}|x - \cdot|^n(y) &\le c_1(n)|x - y|^{n-2} + n|x - y|^{n-2}\sum_{i=1}^N (y_i - x_i)\Big(\frac{q_i - 1}{x_i} - \frac{q_{N+1} - 1}{1 - |y|_1}\Big) \\ &\le c_2(n)|x - y|^{n-2}, \quad x, y \in \Delta^{(N)}.\end{aligned}$$

This implies

$$\mathbb{E}^x|x - \hat{X}_t^x|^n \le c_2(n)\int_0^t \mathbb{E}^x|x - \hat{X}_s^x|^{n-2}\mathrm{d}s, \quad t \ge 0.$$

By induction in $n \ge 2$, we find constants $\{c(n) > 0\}_{n\ge 2}$ such that

$$\mathbb{E}^x|x - \hat{X}_t^x|^n \le c(n)t^{\frac{n}{2}}, \quad t \ge 0, x \in \Delta^{(N)}, n \ge 2.$$

In particular, (2.4) holds for some constant $k > 0$.

□

6 Subordinate stable like processes

In this section, we consider the subordinate stable like process in a connected closed smooth domain $M \subset \mathbb{R}^n, n \in \mathbb{N}$.

Let $0 \le V \in C^2(M)$ such that

$$\mu = \mu_V := \frac{\mathrm{e}^{-V(x)}\mathrm{d}x}{\int_M \mathrm{e}^{-V(x)}\mathrm{d}x}$$

is a probability measure on M. Let $\delta' \in (0, 2)$, and let X_t be the Markov process on M associated with the α'-stable like Dirichlet form

$$\begin{aligned}\mathscr{E}(f,g) &:= \int_{M\times M} \frac{(f(x) - f(y))(g(x) - g(y))}{|x - y|^{n+\alpha'}}\mu(\mathrm{d}x)\mu(\mathrm{d}y), \\ &\quad f, g \in \mathscr{D}(\mathscr{E}) := \{f \in L^2(\mu) : \mathscr{E}(f,f) < \infty\}.\end{aligned} \tag{6.1}$$

Let μ_t^B be the empirical measure of $X_t^B := X_{S_t^B}$, where $B \in \mathbf{B}$.

6.1 Bounded M

When M is bounded, we have the following result for the above defined μ_t^B.

Theorem 6.1. *Assume that M is bounded. Let $\rho(x,y) = |x-y|$ for $x,y \in M$, let $B \in \mathbf{B}^\alpha$ for some $\alpha \in (0,1]$, and let ξ_t be in (2.11) for*

$$\beta := \frac{1}{2} + \frac{n(p-2)}{4p}, \quad d := n, \quad d' := \frac{2n}{\alpha'}. \tag{6.2}$$

Then for any $p \in [2,\infty)$, there exists a constant $c > 0$ such that

$$\sup_{x\in M} \mathbb{E}^x[\mathbb{W}_p(\mu_t^B,\mu)^2] \le c\xi_t, \quad t \ge 2.$$

Proof. Let $\hat{L} = \Delta + \nabla V$ be equipped with the Neumann boundary condition. It is well-known that (2.3)-(2.5) hold for some constant $\lambda > 0$ and constants β and d' in (6.2). According to [7, Theorem 1.1], there exists a constant $k > 0$ such that

$$\sup_{x,y\in M} p_t(x,y) \le k(1\wedge t)^{-\frac{n}{\alpha}'}, \quad t > 0. \tag{6.3}$$

Although only $V = 0$ is considered in [7], this estimate also holds for $V \in C^2(M)$. Indeed, according to [9, Corollary 2.4.7], (6.3) is equivalent to the Nash inequality

$$\mu(f^2) \le c\big(1 + \hat{\mathscr{E}}(f,f)^{\frac{n}{n+\alpha'}}\big), \quad f \in \mathscr{D}(\hat{\mathscr{E}}), \mu(|f|) = 1$$

for some constant $c > 0$. If this inequality holds for $V = 0$, then it also holds for bounded V, which is the case when $V \in C^2(M)$ for compact M.

By (6.3), (2.10) holds for $d' = \frac{2n}{\alpha}'$, and the generator L of P_t has discrete spectrum. Since the Dirichlet form $\mathscr{E}$ is irreducible, the discreteness of the spectrum implies the existence of a spectral gap, so that (2.6) holds for some constant $\lambda > 0$. Since ρ is bounded, and by the second inequality in (2.33), (6.3) implies

$$\|p_1^B\|_\infty \le k\mathbb{E}[(1\wedge S_1^B)^{-\frac{n}{\alpha}'}] =: c_0 < \infty.$$

Combing this with $\|\rho\|_\infty := \sup_{x,y\in M}|x-y| < \infty$ since M is bounded, and applying (2.13) for $q = 2$, we obtain

$$\mathbb{E}^x[\mathbb{W}_p(\mu_t^B,\mu)^2] \le \frac{2}{t^2}\|\rho\|_\infty^2 + 2c_0c\xi_{t-1},$$

which implies the desired estimate, since there exists a constant $c' > 0$ such that $t^{-1}\vee\xi_{t-1} \le c'\xi_t$ holds for $t \ge 2$. □

Remark 6.1. To show the optimality of Theorem 6.1, we simply consider $p = 2, \alpha = 1, \alpha' = 2$, and denote $\mu_t^B = \mu_t$. Then Theorem 6.1 implies

$$\sup_{x\in M} \mathbb{E}^x[\mathbb{W}_2(\mu_t,\mu)^2] \le c\begin{cases} t^{-1}, & \text{if } n \le 3, \\ t^{-1}\{\log(2+t)\}^3, & \text{if } n = 4, \\ t^{-\frac{2}{n-2}}, & \text{if } n \ge 5, \end{cases}$$

which is sharp except for $n = 4$ where the exact order is $t^{-1}\log(2+t)$ according to [39]. This minor loss is caused by the fact that in general $P_t \ne \hat{P}_t$, so that in proof of Theorem 2.2 we can not combining them together by using the semigroup property as in [39].

6.2 $M = \mathbb{R}^n$.

Assume that V satisfies the following conditions for some constants $k > 0, d \geq n$:

$$\nabla^2 V \geq -kI_n, \quad \mathrm{e}^{-V} \in C_b^2(\mathbb{R}^n), \tag{6.4}$$

$$\liminf_{|x|\to\infty} \frac{1}{|x|}\langle \nabla V(x), x\rangle > 0, \tag{6.5}$$

$$\int_{\mathbb{R}^n} \frac{\mu(\mathrm{d}x)}{\mu(B(x,r))} \leq kr^{-d}, \quad r \in (0,1], \tag{6.6}$$

$$\inf_{x\in\mathbb{R}^n} \frac{\mathrm{e}^{V(x)}}{(1+|x|)^{n+\alpha'}} > 0, \tag{6.7}$$

$$\lim_{r\to\infty} \inf_{|x|\geq r-1} r^{n+\alpha'-1}\big\{|\nabla \mathrm{e}^{-V}(x)| + r^{-1}\mathrm{e}^{-V(x)}\big\} = 0. \tag{6.8}$$

We have the following result for μ_t^B defined for the subordinate process of the jump process with Dirichlet form (2.40).

Theorem 6.2. *Assume* (6.4)-(6.8), *and let* $B \in \mathbf{B}$. *Then the following assertions hold.*

(1) *There exists a constant* $c > 0$ *such that*

$$\mathbb{E}^\mu[\mathbb{W}_2(\mu_t^B, \mu)^2] \leq ct^{-\frac{2}{d}}, \quad t \geq 1. \tag{6.9}$$

(2) *Let* $B \in \mathbf{B}^\alpha$ *for some* $\alpha \in (0,1]$, *and there exist constants* $k, d' > 1$ *such that*

$$\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} \leq k(1\wedge t)^{-\frac{d'}{2}}, \quad t \in (0,1], \tag{6.10}$$

then there exists a constant $c > 0$ *such that*

$$\sup_{x\in\mathbb{R}^n} \mathbb{E}^x[\mathbb{W}_2(\mu_t^B, \mu)^2] \leq c\xi_t, \quad t \geq 2, \tag{6.11}$$

where ξ_t *is in* (2.11) *for* $\beta = \frac{1}{2}$, d *in* (6.6) *and* d' *in* (6.10).

Proof. (1) Let $\hat{L} = \Delta - \nabla V$. Then (6.5) implies

$$\limsup_{|x|\to\infty} \hat{L}|\cdot|(x) < 0,$$

so that [26, Corollary 1.4] implies $\mathrm{gap}(\hat{L}) > 0$, so that by Proposition 2.1, (2.3) holds for $p = 2$ and $\beta = \frac{1}{2}$. Moreover, (2.4) and (2.5) have been verified in steps (b) and (c) in the proof of Theorem 3.1, respectively. Finally, by [35, Theorem 1.1(1)], (6.7) and (6.8) imply the Poincaré inequality for $\mathscr{E}$, so that (2.6) holds for $k = 1$ and some constant $\lambda > 0$. Therefore, the estimate (6.9) follows from (2.12) with $\beta = \frac{1}{2}, \alpha = 1$ and $d' = \infty$.

(2) When (6.10) holds, by the second inequality in (2.33), we have

$$\|p_1^B\|_\infty \leq k\mathbb{E}[(1\wedge S_1^B)^{-\frac{d'}{2}}] < \infty.$$

Moroever, (6.5) implies $\mu(|\cdot|^l)<\infty$ for all $l\in(1,\infty)$. Combining this with (6.10), for large enough l we find constants $c_1,c_2>0$ such that

$$\begin{aligned}&\sup_{x\in\mathbb{R}^n}\int_0^1\mathbb{E}^x|X_s^B|^2\mathrm{d}s\le\mathbb{E}\int_0^1\big\|P_{S_s^B}|\cdot|^2\big\|_\infty\mathrm{d}s\\&\le\mu(|\cdot|^{2l})^{\frac{1}{l}}\mathbb{E}\int_0^1\|P_{S_s^B}\|_{L^l(\mu)\to L^\infty(\mu)}\mathrm{d}s\\&\le c_1\int_0^1\mathbb{E}\big[(1\wedge S_s^B)^{-\frac{d'}{2l}}\big]\mathrm{d}s\le c_2\int_0^1 s^{-\frac{d'}{2l\alpha}}\mathrm{d}s<\infty.\end{aligned}$$

Therefore, (6.11) follows from (2.13) with $q=2$.

□

Example 6.1. Let $0\le V\in C^2(\mathbb{R}^n)$ be in Example 2.1 for some $\tau>\frac{1}{2}$. Let $B\in\mathbf{B}^\alpha$ for some $\alpha\in(0,1]$, and let $\xi_t(K)$ be in (2.14) for

$$K_\delta:=\frac{1}{2}+\frac{\tau n(\delta n+\delta\alpha'-\alpha\alpha')}{2\delta(2\tau-1)(n+\alpha')},\quad \delta>2.$$

Then for any $\delta>2$ there exists a constant $c>0$ such that

$$\sup_{x\in\mathbb{R}^n}\mathbb{E}^x[\mathbb{W}_2(\mu_t^B,\mu)^2]\le c\xi_t(K_\delta),\quad t\ge1. \tag{6.12}$$

Proof. Conditions (6.4), (6.5), (6.7), and (6.8) trivially hold. Next, as shown in the proof of (2.17), (6.6) holds for $d=\frac{2\tau n}{2\tau-1}$. Moreover, by [35, Corollary 1.5], for any $\delta>2$ there exists a constant $k>0$ such that

$$\|P_t\|_{L^1(\mu)\to L^\infty(\mu)}\le k(1\wedge t)^{-\frac{\delta(n+\alpha')}{\alpha'}},\quad t\in(0,1],$$

so that (6.10) holds for $d'=\frac{2\delta(n+\alpha')}{\alpha'}$, which satisfies $d'>2\alpha$. It is easy to see that for $\beta=\frac{1}{2}$ and d,d' given above we have $K_\delta=K_{\beta,d,d',\alpha}$, so that (6.12) follows from (6.11).

□

Acknowledgement. The author would like to thank Dr. Jie-Xiang Zhu for corrections, and the referees for helpful comments.

Data availability statements. The manuscript has no associated data.

Declarations. The work is supported in part by the National Key R&D Program of China (No. 2022YFA1006000, 2020YFA0712900) and NNSFC (11921001). The authors have no relevant financial or non-financial interests to disclose.

References

[1] S. Aida, T. Masuda, I. Shigekawa, *Logarithmic Sobolev inequalities and exponential integrability,* J. Funct. Anal. 126(1994), 83–101.

[2] L. Ambrosio, F. Stra, D. Trevisan, *A PDE approach to a 2-dimensional matching problem,* Probab. Theory Relat. Fields 173(2019), 433–477.

[3] F. Baudoin, *Bakry-Emery meet Villani,* J. Funct. Anal. 273(2017), 2275–2291.

[4] Mu-Fa Chen, *Eigenvalues, Inequalities and Ergodic Theory,* Springer-Verlag, 2004.

[5] X. Chen, F.-Y. Wang, *Construction of larger Riemannian metrics with bounded sectional curvatures and applications,* Bull. London Math. Soc. 40(2008), 659–663.

[6] Z.-Q. Chen, M. Fukushima, *Symmetric Markov Processes, Time Change, and Boundary Theory,* London Math. Soc. Monogr. Ser. 35, Princeton University Press, Princeton.

[7] Z.-Q. Chen, T. Kumagai, *Heat kernel estimates for stable-like processes on d-sets,* Stoch. Proc. Appl. 108(2003), 27–62.

[8] R. J. Connor, J. E. Mosimann, *Concepts of independence for proportions with a generalization of the Dirichlet distribution,* J. Amer. Statist. Assoc. 64(1969), 194–206.

[9] E. B. Davies, *Heat Kernels and Spectral Theory,* Cambridge: Cambridge Univ. Press, 1989.

[10] J. Dolbeault, J. A. Klar, C. Mouhot, C. Schmeiser, *Exponential rate of convergence to equilibrium for a model describing fiber Lay-Down processes,* Appl. Math. Research eXpress 2013(2013), 165–175.

[11] S. Feng, L. Miclo, F.-Y. Wang, *Poincaré inequality for Dirichlet distributions and infinite-dimensional generalizations,* Lat. Am. J. Probab. Math. Stat. 14(2017), 361–380.

[12] M. Grothaus, P. Stilgenbauer, *Hypocoercivity for Kolmogorov backward evolution equations and applications,* J. Funct. Anal. 267(2014), 3515–3556.

[13] M. Huesmann, F. Mattesini, D. Trevisan, *Wasserstein asympototics for the empirical measure of fractional Brownian motion on a flat torus,* Stoch. Proc. Appl. 155(2023), 1–26.

[14] M. Ledoux, *On optimal matching of Gaussian samples,* in: Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) vol. 457, in: Veroyatnost Statistika., vol. 25, 2017, pp. 226–264.

[15] H. Li, B. Wu, *Wasserstein convergence rates for empirical measures of subordinated processes on noncompact manifolds,* J. Theoret. Probab. 36(2023), 1243–1268.

[16] H. Li, B. Wu, *Wasserstein convergence for empirical measures of subordinated Dirichlet diffusions on Riemannian manifolds,* Commun. Pure Appl. Anal. 23(2024), 546–576.

[17] H. Li, B. Wu, *Wasserstein convergence for conditional empirical measures of subordinated Dirichlet diffusions on Riemannian manifolds,* arXiv:2204.13559.

[18] H. Li, B. Wu, *Wasserstein convergence for Empirical measures of subordinated fractional Brownian motions on the flat torus,* arXiv:2305.01228.

[19] M. Mariani, D. Trevisan, *Wasserstein asymptotics for Brownian motion on the flat torus and Brownian interlacements,* arXiv:2307.10325.

[20] B. Muckenhoupt, *Hardy's inequality with weights*, XLIV(1972), 31–38.

[21] M. Röckner, F.-Y. Wang, *Weak Poincaré inequalities and convergence rates of Markov semigroups,* J. Funct. Anal. 185(2001), 564–603.

[22] R. Schilling, R. Song, Z. Vondracek,*Bernstein Functions,* Walter de Gruyter, Berlin 2010.

[23] W. Stannat, *On validity of the log-Sobolev inequality for symmetric Fleming-Viot operators,* Ann. Probab. 28(2000), 667–684.

[24] C. Villani, *Hypocoercivity,* Mem. Amer. Math. Soc. 202(2009), iv+141.

[25] C. Villani, *Optimal transport: Old and New Part 1,2,* Springer, 2009.

[26] F.-Y. Wang, *Existence of the spectral gap for elliptic operators,* Arkiv för Math. 37(1999), 395–407.

[27] F.-Y. Wang, *Functional Inequalities, Markov Semigroups and Spectral Theory,* Science Press/Elsevier, 2005.

[28] F.-Y. Wang, *Analysis of Diffusion Processes on Riemannian Manifolds,* World Scientific, 2014.

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

[30] F.-Y. Wang, *Precise limit in Wasserstein distance for conditional empirical measures of Dirichlet diffusion processes,* J. Funct. Anal. 280(2021), 108998, 23pp.

[31] F.-Y. Wang, *Wasserstein convergence rate for empirical measures on noncompact manifolds,* Stoch. Proc. Appl. 144(2022), 271–287.

[32] F.-Y. Wang, *Convergence in Wasserstein distance for empirical measures of semilinear SPDEs,* Ann. Appl. Probab. 33(2023), 70–84.

[33] F.-Y. Wang, *Convergence in Wasserstein distance for empirical measures of Dirichlet diffusion processes on manifolds,* J. Eur. Math. Soc. 25(2023), 3695–3725.

[34] F.-Y. Wang, *Convergence in Wasserstein distance for empirical measures of non-symmetric subordinated diffusion processes,* arXiv:2301.08420.

[35] F.-Y. Wang, J. Wang, *Functional inequalities for stable-like Dirichlet forms,* J. Theor. Probab. 28(2015), 423–448.

[36] F.-Y. Wang, B. Wu, *Wasserstein convergence for empirical measures of subordinated diffusion processes on Riemannian manifolds,* Potential Anal. 59(2023), 933–954.

[37] F.-Y. Wang, W. Zhang, *Nash inequality for Diffusion processes associated with Dirichlet distributions,* Front. Math. China 14(2019), 1317–1338.

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

[39] F.-Y. Wang, J.-X. Zhu, *Limit theorems in Wasserstein distance for empirical Measures of diffusion processes on Riemannian manifolds,* Ann. Inst. Henri Poincar. Probab. Stat. 59(2023), 437–475.

[40] Jie-Xiang Zhu, *Asymptotic behavior of Wasserstein distance for weighted empirical measures of diffusion processes on compact Riemannian manifolds,* arXiv: 2310.01670.