Asymptotics in Wasserstein Distance for Empirical Measures of Markov Processes*

Feng-Yu Wang

Center for Applied Mathematics, Tianjin University, Tianjin 300072, China wangfy@tju.edu.cn

October 17, 2025

Abstract

In this paper we introduce some recent progresses on the convergence rate in Wasserstein distance for empirical measures of Markov processes. For diffusion processes on compact manifolds possibly with reflecting or killing boundary conditions, the sharp convergence rate as well as renormalization limits are presented in terms of the dimension of the manifold and the spectrum of the generator. For general ergodic Markov processes, explicit estimates are presented for the convergence rate by using a nice reference diffusion process, which are illustrated by some typical examples. Finally, some techniques are introduced to estimate the Wasserstein distance between empirical measures.

AMS subject Classification: 60B05, 60B10.
Keywords: Empirical measure, Wasserstein distance, Markov process, Riemannian Manifold, eigenvalue.

1 Introduction

The empirical measure is a fundamental statistic to estimate the stationary distribution of a Markov process. In this paper, we study the long time behavior of empirical measures for Markov processes under the Wasserstein distance. For two nonnegative functions f, g on a space E, we write $f \lesssim g$ if there exists a constant $c > 0$ such that $f \leq cg$ holds on E, and write $f \asymp g$ if $f \lesssim g$ and $g \lesssim f$.

Let (M, ρ) be a Polish space, let $\mathscr{P}$ be the space of all probability measures on M. For a Markov process X_t on M, the empirical measure is defined as

$$\mu_t := \frac{1}{t}\int_0^t \delta_{X_s} \mathrm{d}s$$

*Supported in part by NNSFC (11921001) and the National Key R&D Program of China (No. 2020YFA0712900).

for $t > 0$, where δ_{X_s} is the Dirac measure at X_s. We intend to study the long time behavior of μ_t under the p-Wasserstein distance for any $p \in [1, \infty)$:

$$\mathbb{W}_p(\mu_1, \mu_2) := \inf_{\pi \in \mathscr{C}(\mu_1, \mu_2)} \left(\int_{M \times M} \rho^p \mathrm{d}\pi \right)^{\frac{1}{p}}, \quad \mu_1, \mu_2 \in \mathscr{P},$$

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

To this end, we will consider the ergodicity and quasi-ergodicity cases respectively.

1.1 Ergodicity case

For any $\nu \in \mathscr{P}$, let $\mathbb{E}^\nu$ denote the expectation for the Markov process X_t with initial distribution ν, and denote by $P_t^* \nu$ the law of X_t with initial distribution ν. We have

$$\int_M f \mathrm{d}(P_t^* \nu) = \mathbb{E}^\nu [f(X_t)] = \int_M P_t f(x) \nu(\mathrm{d}x), \quad f \in \mathscr{B}_b(M)$$

where $P_t f(x) := \mathbb{E}^x[f(X_t)]$, and $\mathscr{B}_b(M)$ the class of all bounded measurable functions on M.

If $\mu \in \mathscr{P}$ satisfies $P_t^* \mu = \mu$ for all $t \geq 0$, we call μ an invariant probability measure of the Markov process. If furthermore

$$\lim_{t \to \infty} P_t^* \nu = \mu \text{ weakly}$$

holds for any $\nu \in \mathscr{P}$, we call the Markov process ergodic.

In particular, when the Markov process is exponentially ergodic in $L^2(\mu)$, i.e.

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

holds for some constant $c, \lambda > 0$, where $\| \cdot \|_{L^2(\mu)}$ is the operator norm in $L^2(\mu)$ and $\mu(f) := \int_E f \mathrm{d}\mu$ for $f \in L^2(\mu)$, we have

$$\mathbf{V}_f := \int_0^\infty \mu(\hat{f} P_t \hat{f}) \mathrm{d}t \in (0, \infty), \quad 0 \neq \hat{f} := f - \mu(f), \ f \in L^2(\mu).$$

Moreover, according to [39], for any $f \in L^2(\mu)$, $\mathbb{P}$-a.s.

$$\lim_{t \to \infty} \mu_t(f) = \mu(f),$$

and

$$\lim_{t \to \infty} \sqrt{t}\big(\mu_t(f) - \mu(f)\big) = N(0, 2\mathbf{V}_f) \text{ in law},$$

where $N(0, 2\mathbf{V}_f)$ is the normal distribution with mean 0 and variance $2\mathbf{V}_f$.

Noting that for any $p \geq 1$

$$\mathbb{W}_p(\mu_t, \mu) \geq \mathbb{W}_1(\mu_t, \mu) = \sup_{\|f\|_{Lip} \leq 1} |\mu_t(f) - \mu(f)|,$$

where $\| \cdot \|_{Lip}$ is the Lipschitz constant, the above result implies that the convergence of $\mathbb{E}[\mathbb{W}_p(\mu_t, \mu)]$ could not be faster than $t^{-\frac{1}{2}}$ as $t \to \infty$.

We will show that for elliptic diffusions on compact manifolds, this best possible rate is reached if and only if the dimension of manifold is not larger than 3, and in this case we also derive the exact limit for $t\mathbb{E}[\mathbb{W}_2(\mu_t, \mu)^2]$ as $t \to \infty$, which is given explicitly by the spectrum of the generator.

1.2 Quasi-ergodicity case

Consider Markov process X_t with finite life time τ. Typical examples are the diffusion processes on a domain with killing boundary, for which the life time is the first hitting time to the boundary. In this case, we study the long time behavior of the empirical measure μ_t under the condition $\{t < \tau\}$, i.e. the process is not yet killed at time t.

A probability measure μ is called quasi-invariant for X_t, if

$$\mathbb{E}^{\mu}(f(X_t)|t<\tau) = \mu(f), \quad t \geq 0, f \in \mathscr{B}_b(M),$$

i.e. with initial distribution μ the conditional distribution of X_t under $\{\tau > t\}$ remains to be μ for any $t > 0$. Moreover, we call the Markov process X_t quasi-ergodic with quasi-invariant probability measure μ, if

$$\lim_{t\to\infty} \mathbb{E}^{\nu}(f(X_t)|\tau > t) = \mu(f), \quad \nu \in \mathscr{P}_0,$$

where $\nu \in \mathscr{P}_0$ means that $\nu \in \mathscr{P}$ with $\mathbb{E}^{\nu}[1_{\{\tau>t\}}] > 0$ for all $t > 0$. In this case, we study the long time behaviors of $\mathbb{E}^{\nu}(\mathbb{W}_p(\mu_t, \mu)|\tau > t)$ and $\mathbb{W}_p(\mathbb{E}^{\nu}(\mu_t|\tau > t), \mu)$ for $\nu \in \mathscr{P}_0$.

In the remainder of the paper, we first consider killed diffusion processes on compact manifolds with boundary in Section 2, then study the (reflecting) diffusion processes on compact manifolds in Section 3, and moreover investigate general ergodic Markov processes in Sections 4 and 5 where some typical examples are presented. The corresponding study on weighted empirical measures, time-changed Markov processes, fractional Brownian motions and non-linear Markov processes can be found in [11, 14, 18, 19, 20, 21, 35, 38, 40]. Finally, we summarize some general results for upper and lower bound estimates on Wasserstein distance.

2 Killed Diffusions on Manifolds

In this part, we consider the killed diffusion process on a Riemannian manifold with boundary.

Let M be a d-dimensional compact connected Riemannian manifold with boundary ∂M, let $\mathrm{d}x$ denote the volume measure, and let N be the unit normal vector field of ∂M. We call ∂M convex if its second fundamental form is nonnegative, i.e.

$$\mathbb{I}(U,U) := -\langle \nabla_U N, U\rangle \geq 0, \quad U \in T\partial M.$$

Let $V \in C^1(M)$ such that

$$\mu(\mathrm{d}x) := \mathrm{e}^{V(x)}\mathrm{d}x$$

is a probability measure. We consider the diffusion process X_t generated by

$$L := \Delta + \nabla V$$

which is killed on the boundary ∂M, where Δ and ∇ stand for the Laplacian and gradient operators respectively.

It is classical that under the Dirichlet boundary condition, the operator L is a self-adjoint in $L^2(\mu)$ with discrete spectrum:

$$L\phi_i = -\lambda_i \phi_i, \quad \phi_i|_{\partial M} = 0,$$

where $\lambda_i \asymp i^{\frac{2}{d}}$ for $i \geq 1$ and $\{\phi_i\}_{i\geq 0}$ is an orthonormal basis of $L^2(\mu)$, see for instance [8].

Since X_t is the diffusion process on M generated by L with killing boundary condition, its life time is

$$\tau := \inf\big\{t \geq 0 :\ X_t \in \partial M\big\},$$

and

$$\mathscr{P}_0 = \big\{\nu \in \mathscr{P} :\ \nu(\partial M) < 1\big\}.$$

Moreover, the diffusion process is quasi-ergodic with unique quasi-invariant probability measure $\mu_0(\mathrm{d}x) = \phi_0^2 \mathrm{d}\mu$. The following result shows that in most cases we have $\mathbb{W}_2(\mathbb{E}^\nu(\mu_t|\tau > t), \mu_0) \asymp t^{-1}$.

Theorem 2.1 ([29])**.** *Let $\mu_0 = \phi_0^2 \mu$. Then for any $\nu \in \mathscr{P}_0$,*

$$\begin{aligned} &\lim_{t\to\infty} \big\{t^2 \mathbb{W}_2(\mathbb{E}^\nu[\mu_t|\tau > t], \mu_0)^2\big\} \\ &= \frac{1}{\{\mu(\phi_0)\nu(\phi_0)\}^2} \sum_{i=1}^{\infty} \frac{\{\nu(\phi_0)\mu(\phi_i) + \mu(\phi_0)\nu(\phi_i)\}^2}{(\lambda_i - \lambda_0)^3} > 0, \end{aligned}$$

and the limit is finite if and only if $\nu \in \mathscr{D}\big((-\hat{L})^{-\frac{3}{2}}\big)$, i.e.

$$\sum_{i=1}^{\infty} \nu(\phi_i)^2 \lambda_i^{-3} < \infty.$$

By the Sobolev embedding theorem, we have $\mathscr{P} \subset \mathscr{D}((-\hat{L})^{-\frac{3}{2}})$ for $d \leq 6$, and $\mathscr{D}((-\hat{L})^{-\frac{3}{2}}) \supset L^{\frac{2d}{d+6}}(\mu)$ for $d > 6$. Hence, by Theorem 2.1, $\mathbb{W}_2(\mathbb{E}^\nu[\mu_t|\tau > t], \mu_0) \asymp t^{-1}$ holds if either $d \leq 6$ or $d \geq 7$ and $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^{\frac{2d}{d+6}}(\mu)$.

When $d \geq 7$ but $\nu \notin \mathscr{D}((-\hat{L})^{-\frac{3}{2}})$, the limit in the above theorem becomes ∞, so that the convergence of $W_2(\mathbb{E}^\nu[\mu_t|\tau > t], \mu_0)$ is slower than t^{-1}, for which the exact convergence rate remains open.

Next, we consider the long time behavior of $\mathbb{E}^\nu(\mathbb{W}_2(\mu_t, \mu_0)|\tau > t)$.

Theorem 2.2 ([32])**.** *Let $\nu \in \mathscr{P}_0$. The the following assertions hold.*

(1) *When ∂M is convex, we have*

$$\liminf_{t\to\infty} \Big\{t\mathbb{E}^\nu\big[\mathbb{W}_2(\mu_t, \mu_0)^2\big|t < \tau\big]\Big\} = \sum_{i=1}^{\infty} \frac{2}{(\lambda_i - \lambda_0)^2}.$$

(2) *If ∂M is non-convex, then there exists a constant $c \in (0,1]$ such that*

$$c\sum_{i=1}^{\infty}\frac{2}{(\lambda_i-\lambda_0)^2} \le \liminf_{t\to\infty}\Big\{t\mathbb{E}^{\nu}\big[\mathbb{W}_2(\mu_t,\mu_0)^2\big|t<\tau\big]\Big\}$$
$$\le \limsup_{t\to\infty}\Big\{t\mathbb{E}^{\nu}\big[\mathbb{W}_2(\mu_t,\mu_0)^2\big|t<\tau\big]\Big\} \le \sum_{i=1}^{\infty}\frac{2}{(\lambda_i-\lambda_0)^2}.$$

Since $\lambda_i-\lambda_0 \asymp i^{\frac{2}{d}}$, it is easy to see that

$$\sum_{i=1}^{\infty}\frac{2}{(\lambda_i-\lambda_0)^2}<\infty$$

if and only if $d\le 3$. So, Theorem 2.2 shows that $\mathbb{E}^{\nu}\big[\mathbb{W}_2(\mu_t,\mu_0)^2\big|t<\tau\big]\asymp t^{-1}$ for $d\le 3$, but the convergence is slower than than t^{-1} for $d\ge 4$. The next result gives the exact convergence rate for $d\ge 4$ where $d=4$ is critical, but for completeness we also include the case for $d\le 3$. The lower bound estimate for $d=4$ is due to [33] where more general Markov processes are considered, other estimates are taken from [32].

Theorem 2.3 ([32, 33])**.** *Let $\nu\in\mathscr{P}_0$. Then for larger $t>1$*

$$\mathbb{E}^{\nu}\big[\mathbb{W}_2(\mu_t,\mu_0)^2\big|t<\tau\big]\asymp\begin{cases}t^{-1}, & \text{if } d\le 3,\\ t^{-1}\log t, & \text{if } d=4,\\ t^{-\frac{2}{d-2}}, & \text{if } d\ge 5.\end{cases}$$

3 (Reflected) Diffusions on Manifolds

Let M be a d-dimensional connected compact Riemannian manifold without boundary or with a boundary ∂M. In this part, we consider the (reflected, if ∂M exists) diffusion process X_t generated by

$$L:=\Delta+\nabla V+Z,$$

where $V\in C^1(M)$ such that $\mu(\mathrm{d}x):=\mathrm{e}^{V(x)}\mathrm{d}x$ is a probability measure, and Z is a C^1-vector field with

$$\mathrm{div}_{\mu}(Z):=ZV+\mathrm{div}Z=0,$$

where $ZV:=\langle Z,\nabla V\rangle$. Then X_t is ergodic with unique invariant probability measure μ. We will study the convergence rate of $\mathbb{W}_p(\mu_t,\mu)$ when $t\to\infty$.

Corresponding to the killed case where the Dirichlet eigenvalue problem is involved, in the present case we will use the Neumann eigenvalue problem if ∂M exists, and the closed eigenvalue problem otherwise.

Let N be the inward unit normal vector field on ∂M if the boundary exists. Consider the Neumann/closed eigenvalue problem:

$$L\phi_i=-\lambda_i\phi_i,\quad \mu(\phi_i^2)=1,\quad N\phi_i|_{\partial M}=0,$$

where $\phi_0\equiv 1, \lambda_0=0$, and the Neumann condition $N\phi_i|_{\partial M}=0$ applies only when ∂M exists. It is well known that $\lambda_i\asymp i^{\frac{2}{d}}$ for $i\ge 0$ and $\{\phi_i\}_{i\ge 0}$ is an orthonormal basis of $L^2(\mu)$, see for instance [8].

3.1 Long time behavior of $\mathbb{E}[\mathbb{W}_2(\mu_t,\mu)^2]$

Recall that for any $0\neq f\in L^2(\mu)$ with $\mu(f)=0$,

$$\mathbf{V}_f:=\int_0^\infty \mu(fP_tf)\mathrm{d}t\in(0,\infty),$$

where P_t is the diffusion semigroup of X_t. By $\mathrm{div}_\mu(Z)=0$, $Z\phi_i:=\langle Z,\nabla\phi_i\rangle$ satisfies $\mu(Z\phi_i)=0$. The following result is due to [37] for $Z=0$ and [33] otherwise.

Theorem 3.1 ([33, 37])**.** *Let X_t be the (reflected if ∂M exists) diffusion generated by $L=\Delta+\nabla V+Z$ on M.*

(1) *When ∂M is either empty or convex,*

$$\lim_{t\to\infty} t\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2]=\sum_{i=1}^\infty \frac{2}{\lambda_i^2}\Big(1-\frac{1}{\lambda_i}\mathbf{V}_{Z\phi_i}\Big)$$

holds uniformly in $\nu\in\mathscr{P}$.

(2) *If ∂M is non-convex, then there exists a constant $c\in(0,1]$ such that*

$$c\sum_{i=1}^\infty \frac{2}{\lambda_i^2}\Big(1-\frac{1}{\lambda_i}\mathbf{V}_{Z\phi_i}\Big)\le\liminf_{t\to\infty}\inf_{\nu\in\mathscr{P}} t\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2]$$
$$\le\limsup_{t\to\infty}\sup_{\nu\in\mathscr{P}} t\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2]\le\sum_{i=1}^\infty \frac{2}{\lambda_i^2}\Big(1-\frac{1}{\lambda_i}\mathbf{V}_{Z\phi_i}\Big).$$

Theorem 3.1 shows that a divergence-free perturbation Z accelerates the convergence of $\mathbb{E}[\mathbb{W}_2(\mu_t,\mu)^2]$ with the exact factor $-\frac{1}{\lambda_i}\mathbf{V}_{Z\phi_i}$. This fits well to the observation in [12] that divergence-free perturbations accelerate the convergence in the algorithm of Gibbs measure.

On the other hand, since $\lambda_i\asymp i^{\frac{2}{d}}$, Theorem 3.1 shows that $E^\nu[\mathbb{W}_2(\mu_t,\mu)^2]\asymp t^{-1}$ if and only if $d\le 3$. Correspondingly to Theorem 2.3, the following result also present exact convergence rates for $d\ge 4$, where the lower bound estimate for $d=4$ is due to [33] and other estimates are taken from [37].

Theorem 3.2 ([33, 37])**.** *The following holds for large $t>0$ uniformly in $\nu\in\mathscr{P}$:*

$$\mathbb{E}^\nu\big[\mathbb{W}_2(\mu_t,\mu)^2\big]\asymp\begin{cases} t^{-1}, & \text{if } d\le 3,\\ t^{-1}\log(t+1), & \text{if } d=4,\\ t^{-\frac{2}{d-2}}, & \text{if } d\ge 5.\end{cases}$$

Indeed, for $d=4$ we have the following renormalization formula.

Theorem 3.3 ([23])**.** *When $d=4$ and ∂M is empty, there holds*

$$\lim_{t\to\infty}\sup_{\nu\in\mathscr{P}}\left|\frac{t}{\log t}\mathbb{E}^\nu[\mathbb{W}_2^2(\mu_t,\mu)]-\frac{\mathrm{vol}(M)}{8\pi^2}\right|=0,$$

where $\mathrm{vol}(M)$ *is the volume of M.*

When $d\ge 5$, it is not clear whether the the following limit exists or not:

$$\lim_{t\to\infty} t^{\frac{2}{d-2}}\mathbb{E}[\mathbb{W}_2^2(\mu_t,\mu)].$$

3.2 Long time behavior of $t\mathbb{W}_2(\mu_t,\mu)^2$ for $d\le 3$

We first consider the weak convergence of $t\mathbb{W}_2(\mu_t,\mu)^2$.

Theorem 3.4 ([37]). *If $d\le 3, Z=0$, and ∂M is either empty or convex, then*

$$\lim_{t\to\infty} t\mathbb{W}_2(\mu_t,\mu)^2=\sum_{i=1}^\infty \frac{2\xi_i^2}{\lambda_i^2}\quad \textit{in law},$$

where $\{\xi_i\}$ are i.i.d. standard normal random variables.

Next, we consider the convergence of $t\mathbb{W}_2(\mu_t,\mu)^2$ in $L^q(\mathbb{P})$ for any $q\ge 1$ to the following specific process $\Xi(t)$:

$$\Xi(t):=\sum_{i=1}^\infty \frac{\psi_i(t)}{\lambda_i},$$

where

$$\psi_i(t):=\frac{1}{\sqrt{t}}\int_0^t \phi_i(X_s)\mathrm{d}s.$$

Recall that by the central limit theorem of [39], $\psi_i(t)\to N(0,2\mathbf{V}_{\phi_i})$ weakly with

$$\mathbf{V}_{\phi_i}=\int_0^\infty \mu(\phi_i P_t\phi_i)\mathrm{d}t=\frac{1}{\lambda_i}\Big(1-\frac{1}{\lambda_i}\mathbf{V}_{Z\phi_i}\Big).$$

We will consider initial distributions in the classes

$$\mathscr{P}_{k,R}:=\big\{\nu=\rho_\nu\mu\in\mathscr{P}(M):\ \|\rho_\nu\|_{L^k(\mu)}\le R\big\},\quad k,R\ge 1.$$

Theorem 3.5 ([36]). *Let ∂M be empty or convex. If $d\le 2$, then*

$$\lim_{t\to\infty}\sup_{\nu\in\mathscr{P}}\mathbb{E}^\nu\big[\big|t\mathbb{W}_2(\mu_t,\mu)^2-\Xi(t)\big|^q\big]=0,\quad q\ge 1.$$

If $d=3$, then

$$\lim_{t\to\infty}\sup_{\nu\in\mathscr{P}_{k,R}}\mathbb{E}^\nu\big[\big|t\mathbb{W}_2(\mu_t,\mu)^2-\Xi(t)\big|^q\big]=0,\quad k,R,q\ge 1.$$

3.3 Moment estimates on $\mathbb{W}_p(\mu_t,\mu)$

The following result shows that the exact convergence rate of $\big(\mathbb{E}[\mathbb{W}_p(\mu_t,\mu)^q]\big)^{\frac{2}{q}}$ is uniformly in $p\ge 1$ and $q>0$.

Theorem 3.6 ([36]). *Let $p\in[1,\infty)$ and $q\in(0,\infty)$.*

(1) *If $p\le \frac{2d}{(d-2)^+}\vee\frac{d(d-2)}{2}$, where $\frac{2d}{(d-2)^+}=\infty$ when $d\le 2$, then the following asymptotic formula for large $t>1$ holds uniformly in $\nu\in\mathscr{P}$:*

$$\big(\mathbb{E}^\nu[\mathbb{W}_p(\mu_t,\mu)^q]\big)^{\frac{2}{q}}\asymp\begin{cases} t^{-1}, & \textit{if } d\le 3,\\ t^{-1}\log t, & \textit{if } d=4,\\ t^{-\frac{2}{d-2}}, & \textit{if } d\ge 5.\end{cases}$$

(2) *If $p>\frac{2d}{(d-2)^+}\vee\frac{d(d-2)}{2}$, then for any $k,R\ge 1$ the above asymptotic formula for large $t>1$ holds uniformly in $\nu\in\mathscr{P}_{k,R}$.*

4 Exponential Ergodic Markov processes

In this part, we consider a general framework of exponential ergodic Markov process, which will be extended in Section 5 to non-exponentially ergodic setting.

Let (M,ρ) be a length space, which is the basic space for analysis on the Wasserstein space (see [3, 24]), i.e. for any $x,y\in M$, the distance $\rho(x,y)$ can be approximated by the length of curves linking x and y. A typical class of length space is the geodesic space, where the distance $\rho(x,y)$ can be reached by the length of a geodesic curve linking x and y.

Let X_t be an ergodic Markov process with unique invariant probability measure μ. We estimate the upper bound of $\mathbb{E}[\mathbb{W}_p(\mu_t,\mu)^2]$. Since the empirical measure is usually singular with μ, to apply analysis techniques we need to regularize μ_t using the following introduced diffusion process on M.

Let $\hat X_t$ be a reversible diffusion 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 $C_{b,L}(M)$ is 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 $\mathscr{D}(\hat{\mathscr{E}})$ under the $\hat{\mathscr{E}}_1$-norm

$$\|f\|_{\hat{\mathscr{E}}_1}:=\sqrt{\mu(f^2)+\hat{\mathscr{E}}(f,f)},$$

and the Dirichlet form restricted on $C_{b,L}(M)$ is formulated as

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

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 $\phi\in C_b^1(\mathbb{R})$, we have

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

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

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

We make the following assumption.

(*A*) The reference diffusion semigroup $\hat{P}_t$ has heat kernel $\hat{p}_t$ with respect to μ, and there exist constants $\beta, \lambda, d, k \in (0, \infty)$ such that

$$\|\nabla \hat{P}_t f\|_{L^2(\mu)\to L^p(\mu)} \le k\mathrm{e}^{-\lambda t} t^{-\beta},\ t > 0, \tag{4.1}$$

$$\int_M \big(\hat{P}_t \rho(x, \cdot)^p\big)^{\frac{2}{p}}(x)\mu(\mathrm{d}x) \le kt,\ t \in [0, 1], \tag{4.2}$$

$$\int_M \hat{p}_t(x, x)\mu(\mathrm{d}x) \le kt^{-\frac{d}{2}},\ t \in (0, 1]. \tag{4.3}$$

Moreover, there exist constants $\theta_1, \theta_2 \in (0, \infty)$ such that

$$\|P_t - \mu\|_{L^2(\mu)} \le \theta_1 \mathrm{e}^{-\theta_2 t},\ t \ge 0. \tag{4.4}$$

Note that conditions (4.1)-(4.3) can be verified by choosing a suitable symmetric reference distance ρ. Indeed, for small ρ the Dirichlet form $\hat{\mathscr{E}}$ is larger, so that $\hat{P}_t$ has better properties. When $p = 2$, (4.1) holds for $\beta = \frac{1}{2}$ if and only if $\hat{L}$ has a spectral gap, i.e.

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

The only condition on X_t is the exponential ergodicity (4.4), which will be weakened later on by allowing more general ergodic rate.

Let $K := \beta + \frac{d}{4}$. The convergence rate of $\mathbb{E}[\mathbb{W}_p(\mu_t, \mu)^2]$ will be given by

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

Theorem 4.1 ([34])**.** *Assume (A) and let $K := \beta + \frac{d}{4}$. Then there exists a constant $c > 0$ such that for any $t > 0$,*

$$\mathbb{E}^\mu[\mathbb{W}_p(\mu_t, \mu)^2] \le c\xi_K(t), \quad t > 0. \tag{4.5}$$

If P_t has heat kernel p_t with respect to μ, then for any $q \in [1, 2], t > 1$ and $x \in M$,

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

In particular, when p_1 is bounded, there exists a constant $c > 0$ such that

$$\sup_{x\in M} \mathbb{E}^x[\mathbb{W}_p(\mu_t, \mu)^2] \le c\xi_K(t-1), \quad t > 1. \tag{4.6}$$

Comparing with the exact convergence rate for elliptic diffusions on compact manifolds, the present convergence rate is less sharp. However, as a universal convergence rate for arbitrary exponential ergodic Markov processes, $\xi_K(t)$ is almost optimal. To see this, let us consider the following example.

Example 4.1 (Markov processes on compact manifolds). Let M be a d-dimensional compact connected Riemannian manifold possibly with a boundary ∂M, let $\mu(\mathrm{d}x) = \mathrm{e}^{V(x)}\mathrm{d}x$ be a probability measure on M for some $V \in C^2(M)$, and let $\hat{L} := \Delta + \nabla V$ (with Neumann boundary condition if ∂M exists). Then (4.1)-(4.3) hold for $\beta = \frac{1}{2}$ so that $K = \frac{1}{2} + \frac{d}{4}$. By Theorem 4.1, for any Markov process on M satisfying (4.4) for some constant $\theta_1, \theta_2 > 0$, there exists a constant $c > 0$ such that (4.5) holds, and (4.6) holds when p_1 is bounded, for

$$\xi_K(t) := \begin{cases} t^{-1}, & \text{if } d < 2, \\ t^{-1}[\log t]^2, & \text{if } d = 2, \\ t^{-\frac{2}{d}}, & \text{if } d > 2. \end{cases}$$

On the other hand, according to [35], for X_t being the α-stable time changed process of $\hat{X}_t$,

$$\mathbb{E}^x[\mathbb{W}_2(\mu_t, \mu)^2] \asymp \begin{cases} t^{-1}, & \text{if } d < 2(1+\alpha), \\ t^{-1}\log t, & \text{if } d = 2(1+\alpha), \\ t^{-\frac{2}{d-2\alpha}}, & \text{if } d > 2(1+\alpha). \end{cases}$$

Since $\xi_K(t)$ is the universal convergence rate for all $\alpha \in (0,1)$, it is reached by the exact rate as $\alpha \to 0$, except $[\log t]^2$ in the critical case.

Next, we consider a class of Markov processes on $\mathbb{R}^n$.

Example 4.2 (Markov processes on $\mathbb{R}^n$). Let $M = \mathbb{R}^n$, let $V \in C^2(\mathbb{R}^n)$ such that

$$V(x) = \psi(x) + (1 + \theta|x|^2)^\tau, \quad x \in \mathbb{R}^n,$$

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 (4.4), there exists a constant $c > 0$ such that

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

Proof. Let $\hat{L} = \Delta - \nabla V$. By Theorem 4.1, it suffices to verify (4.1)-(4.3) for $p = 2, \beta = \frac{1}{2}$, and $d = \frac{2\tau n}{2\tau - 1}$. Since

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

[25, Corollary 1.4] ensures $\mathrm{gap}(\hat{L}) > 0$, so that (4.1) holds for $p = 2$ and $\beta = \frac{1}{2}$.

Next, by [26, Theorem 2.4.4] and $\nabla^2 V \geq -c_1 I_n$, where I_n is the $n\times n$-unit matrix, we find a constant $c_2>0$ such that

$$\hat{p}_r(x,x) \leq \frac{c_2}{\mu(B(x,\sqrt{r}))}, \quad x\in\mathbb{R}^n, r\in(0,1], \tag{4.7}$$

where $B(x,r) := \{y\in\mathbb{R}^n : |x-y|<r\}, r>0$. Then (4.3) with $d=\frac{2\tau n}{2\tau-1}$ follows provided

$$\int_{\mathbb{R}^n} \frac{\mu(\mathrm{d}x)}{\mu(B(x,r))} \leq c r^{-\frac{2\tau n}{2\tau-1}}, \quad r\in(0,1], x\in\mathbb{R}^n \tag{4.8}$$

holds for some constant $c>0$. Since ψ is bounded, there exists a constant $C>1$ such that

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

So, (4.8) is trivial for $|x|\leq 1$. On the other hand, when $|x|\geq 1$ we have

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

so we may find a constant $c_3>0$ such that

$$\begin{aligned}
&\Big(1+\theta\Big|x-\frac{rx}{4|x|}\Big|^2\Big)^\tau = (1+\theta|x|^2)^\tau + \int_0^r \frac{\mathrm{d}}{\mathrm{d}s}\Big(1+\theta\big(|x|-\frac{s}{4}\big)^2\Big)^\tau \mathrm{d}s\\
&= (1+\theta|x|^2)^\tau - \frac{\tau\theta}{2}\int_0^r \Big(1+\theta\big(|x|-\frac{s}{4}\big)^2\Big)^{\tau-1}\Big(|x|-\frac{s}{4}\Big)\mathrm{d}s\\
&\leq (1+\theta|x|^2)^\tau - c_3 r|x|^{2\tau-1}.
\end{aligned}$$

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

$$\begin{aligned}
\mu(B(x,r)) &\geq c_4\int_{B\left(x-\frac{rx}{2|x|},\frac{r}{4}\right)} \mathrm{e}^{-(1+\theta|y|^2)^\tau}\mathrm{d}y\\
&\geq c_5 r^n \mathrm{e}^{-(1+\theta|x-\frac{rx}{4|x|}|^2)^\tau} \geq c_5 r^n \mathrm{e}^{-(1+\theta|x|^2)^\tau + c_3 r|x|^{2\tau-1}}.
\end{aligned} \tag{4.9}$$

Therefore, there exist constants $c_6, c_7>0$ such that

$$\begin{aligned}
&\int_{\mathbb{R}^n}\frac{\mu_V(\mathrm{d}x)}{\mu_V(B(x,r))} \leq c_6 r^{-n}\int_{\mathbb{R}^n}\mathrm{e}^{-c_3 r|x|^{2\tau-1}}\mathrm{d}x\\
&= c_7 r^{-n}\int_0^\infty s^{n-1}\mathrm{e}^{-c_3 r s^{2\tau-1}}\mathrm{d}s = c_7 r^{-\frac{2\tau n}{2\tau-1}}\int_0^\infty s^{n-1}\mathrm{e}^{-c_3 s^{2\tau-1}}\mathrm{d}s.
\end{aligned}$$

Thus, (4.8) holds for some constant $c>0$.

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

$$\begin{aligned}
\hat{L}|x-\cdot|^2 &= 2n + 2\langle\nabla V, x-\cdot\rangle = 2n + 2\langle\nabla V(x), x-\cdot\rangle - 2\langle\nabla V(x)-\nabla V, x-\cdot\rangle\\
&\leq 2n + |\nabla V(x)|^2 + |x-\cdot|^2 + 2c_1|x-\cdot|^2 \leq c_8(1+|x|^{4\tau}+|x-\cdot|^2), \quad x\in\mathbb{R}^n.
\end{aligned} \tag{4.10}$$

This implies

$$\hat{P}_t|x-\cdot|^2(x) = \mathbb{E}^x|x-\hat{X}_t|^2 \leq c_8\big(1+|x|^{4\tau}|\big)t\mathrm{e}^{c_8 t}, \quad x\in\mathbb{R}^n,\ t>0. \tag{4.11}$$

Noting that $\mu(|\cdot|^{4\tau})<\infty$, we verify condition (4.2) for $p=2$ and some constant $k>0$. □

As a special case of Example 4.2, we consider the stochastic Hamiltonian system, a typical degenerate SDE for $X_t = (X_t^{(1)}, X_t^{(2)})$ on $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$ ($n, m \geq 1$ may be different):

$$\begin{cases} \mathrm{d}X_t^{(1)} = \kappa Q X_t^{(2)} \mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = \sqrt{2}\,\mathrm{d}W_t - \big\{\theta Q^* X_t^{(1)} + \kappa X_t^{(2)}\big\}\mathrm{d}t, \end{cases} \tag{4.12}$$

where W_t is the m-dimensional Brownian motion, $Q \in \mathbb{R}^{n\otimes m}$ such that QQ^* is invertible, and $\kappa, \theta > 0$ are constants. Let

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

By [28], where more general degenerate models are considered, the associated Markov semigroup P_t is exponentially ergodic in entropy, hence (4.4) holds. So, as shown in Example 4.2, (4.1)-(4.3) hold for $p = 2, \beta = \frac{1}{2}$, and $d = \frac{2(n+m)}{2\tau - 1}$. Therefore, for any time-changed process of X_t, there exists a constant $c > 0$ such that

$$\mathbb{E}^\mu[\mathbb{W}_2(\mu_t, \mu)^2] \leq c t^{-\frac{1}{m+n}}.$$

5 More general ergodic Markov processes

For some infinite-dimensional models, see for instance [31], (4.3) fails for any $d \in (0, \infty)$, but there may be a decreasing function $\gamma : (0, \infty) \to (0, \infty)$ such that

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

Next, in infinite-dimensional case the condition (4.2) may be invalid for small time, see Corollary 5.2 below. 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 [22], we have

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

To cover these situations for which Theorem 4.1 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 5.1 ([34])**.** *Assume* (4.1), (4.2). *If there exist a constant* $q \in [1, \infty], q' \in [\frac{q}{q-1}, \infty]$, *a decreasing function* $\gamma : (0, \infty) \to (0, \infty)$ *and an increasing continuous function* $h : [0, 1] \to [0, \infty)$ *with* $h(0 = 0$ *such that*

$$\int_M \big(\hat{P}_t \rho(x, \cdot)^p\big)^{\frac{2}{p}}(x)\mu(\mathrm{d}x) \leq h(t), \quad t \in (0, 1], x \in M, \tag{5.1}$$

$$\lim_{t\to\infty} \|P_t - \mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)} = 0, \tag{5.2}$$

$$\int_M \|\hat{p}_{\frac{r}{2}}(y, \cdot)\|_{L^q(\mu)} \|\hat{p}_{\frac{r}{2}}(y, \cdot)\|_{L^{q'}(\mu)} \mu(\mathrm{d}y) \leq \gamma(r), \quad r > 0. \tag{5.3}$$

For any $t>0$, let

$$\xi(t):=\inf_{r\in(0,1]}\bigg\{\frac{\int_0^t\|P_s-\mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)}\mathrm{d}s}{t}\bigg(\int_0^1\frac{\sqrt{\gamma(r+s)}}{(r+s)^\beta}\mathrm{d}s\bigg)^2+h(r)\bigg\}.$$

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

$$\mathbb{E}^\mu[\mathbb{W}_p(\mu_t,\mu)^2]\le c\xi(t). \tag{5.4}$$

If P_t has heat kernel p_t with respect to μ, then for any $q\in[1,2], t>1$ and $x\in M$,

$$\mathbb{E}^x[\mathbb{W}_p(\mu_t,\mu)^q]\le ct^{-q}\int_0^1\mathbb{E}^x\big[\mu(\rho(X_s,\cdot)^p)^{\frac{q}{p}}\big]\mathrm{d}s+c\big(\|p_1(x,\cdot)\|_{L^{\frac{2}{2-q}}(\mu)}\xi(t-1)\big)^{\frac{q}{2}}. \tag{5.5}$$

To verify Theorem 5.1, we present below a simple example where P_t only has algebraic convergence in $\|\cdot\|_{L^\infty(\mu)\to L^2(\mu)}$, so Theorem 4.1 does not apply.

Example 5.1. 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 4.1 does not apply, but by Theorem 5.1 there exists a constant $c>0$ such that for any $t>0$,

$$\mathbb{E}^\mu[\mathbb{W}_p(\mu_t,\mu)^2]\le c\begin{cases} t^{-1}, & \text{if } l\in(2,5), p\in[2,\frac{13-l}{4}),\\ t^{-1}[\log(2+t)]^3, & \text{if } l\in(2,5], p=\frac{13-l}{4},\\ \big[t^{-1}\log(2+t)\big]^{\frac{8}{4p+l-5}}, & \text{if } l\in(2,5], p>\frac{13-l}{4},\\ t^{-\frac{4}{l-1}}[\log(2+t)]^2, & \text{if } l>5, p=2\\ t^{-\frac{8}{p(l-1)}}, & \text{if } l>5, p>2.\end{cases} \tag{5.6}$$

Proof. We first observe that (4.4) fails, so that Theorem 4.1 does not apply. Indeed, the Dirichlet form of L satisfies

$$\mathscr{E}(f,g)=\int_0^1\big\{x(1-x)\big\}^l(f'g')(x)\mathrm{d}x,\quad f,g\in C_b^1(M)\subset\mathscr{D}(\mathscr{E}). \tag{5.7}$$

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

$$\rho_L(x)=\bigg|\int_{\frac12}^x\big\{s(1-s)\big\}^{-\frac{l}{2}}\mathrm{d}s\bigg|\ge c_1\big(x^{1-\frac{l}{2}}+(1-x)^{1-\frac{l}{2}}\big),\quad x\in M.$$

Then for any $\varepsilon>0$, we have $\mu(\mathrm{e}^{\varepsilon\rho_L})=\infty$, so that by [2], $\mathrm{gap}(L)=0$. On the other hand, since L is symmetric in $L^2(\mu)$, by [22, Lemma 2.2], (4.4) implies the same inequality for $k=1$, so that $\mathrm{gap}(L)\ge\lambda>0$. Hence, (4.4) fails.

To apply Theorem 5.1, let $\hat{P}_t$ be the standard Neumann heat semigroup on M generated by Δ. It is classical that (4.1) and (5.1) with $h(t) = kt$ hold for some constant $k > 0$ and

$$\beta = \frac{1}{2} + \frac{p-2}{4}. \tag{5.8}$$

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 $q > 1$,

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

Hence, there exists a constant $c_3 > 0$ such that (5.3) holds for

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

Combining this with (5.8), we find a constant $k > 0$ such that for any $r \in (0,1)$,

$$\eta(r) := \left(\int_0^1 \frac{\sqrt{\gamma(r+s)}}{(r+s)^\beta}\mathrm{d}s\right)^2 \le k\begin{cases}1, & \text{if } 1 < q < \frac{1}{p-2},\\ [\log(1+r^{-1})]^2, & \text{if } 1 < q = \frac{1}{p-2},\\ r^{\frac{1-(p-2)q}{2q}}, & \text{if } q > 1 \vee \frac{1}{p-2}.\end{cases} \tag{5.9}$$

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 [22]. 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

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

By the weighted Hardy inequality, see for instance [26, 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 [26, Lemma 1.4.3], 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 (5.7), 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

$$\mu(f^2) = \mu(f^2 1_{M_s^c}) + \mu(f^2 1_{M_s}) \le \mu(f^2 1_{M_s^c}) + 4c_4 s^{1-l}\mathscr{E}(f,f) + \mu(f1_{M_s^c})^2,$$

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

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

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

By [22, Corollary 2.4(2)], this implies

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

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

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

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

$$\Gamma(t):=\frac{1}{t}\int_0^t \|P_s-\mu\|_{L^{q'}(\mu)\to L^{\frac{q}{q-1}}(\mu)}\mathrm{d}s \leq k\begin{cases} t^{-1}, & \text{if } l\in(2,5), q>\frac{4}{5-l},\\ t^{-1}\log(2+t), & \text{if } l=5, q=\infty,\\ (1+t)^{-\frac{4}{l-1}}, & \text{if } l>5, q=\infty.\end{cases} \tag{5.10}$$

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 (5.9) and (5.10), we obtain

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

So, the desired estimate follows from Theorem 5.1.

(2) Let $l\in(2,5]$ and $p=\frac{13-l}{4}$. Taking $q=\frac{4}{5-l}=\frac{1}{p-2}$ in (5.9) and (5.10) we find a constant $c>0$ such that

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

This implies the desired estimate according to Theorem 5.1.

(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 (5.9) and (5.10) imply

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

for some constant $c>0$, which implies the desired estimate by Theorem 5.1.

(4) Let $l>5$ and $p=2$. By taking $q=\infty$ in (5.9) and (5.10), we find a constant $c>0$ such that

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

By Theorem 5.1, the desired estimate holds.

(5) Let $l>5$ and $p>2$. By taking $q=\infty$ we find a constant $c>0$ such that (5.9) and (5.10) imply

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

for some constant $c>0$. Hence the desired estimate holds according to Theorem 5.1.

□

Finally, we consider the following semilinear SPDE on a separable Hilbert space $\mathbf{H}$: Consider the following SDE on a separable Hilbert space $\mathbf{H}$:

$$\mathrm{d}\hat{X}_t = \big\{\nabla V(\hat{X}_t) - A\hat{X}_t\big\}\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \tag{5.11}$$

where W_t is the cylindrical Brownian motion on $\mathbf{H}$, i.e.

$$W_t = \sum_{i=1}^{\infty} B_t^i e_i, \quad t \geq 0$$

for an orthonormal basis $\{e_i\}_{i\geq 1}$ of $\mathbf{H}$ and a sequence of independent one-dimensional Brownian motions $\{B_t^i\}_{i\geq 1}$, $(A, \mathscr{D}(A))$ is a positive definite self-adjoint operator and $V \in C^1(\mathbf{H})$ satisfying the following assumption.

(H_1) A has discrete spectrum with eigenvalues $\{\lambda_i > 0\}_{i\geq 1}$ listed in the increasing order counting multiplicities satisfying $\sum_{i=1}^{d} \lambda_i^{-\delta} < \infty$ for some constant $\delta \in (0,1)$, and $V \in C^1(\mathbf{H})$, ∇V is Lipschitz continuous in $\mathbf{H}$ such that

$$\langle \nabla V(x) - \nabla V(y), x - y\rangle \leq (K + \lambda_1)|x - y|^2, \quad x, y \in \mathbf{H} \tag{5.12}$$

holds for some constant $K \in \mathbb{R}$. Moreover, $Z_V := \mu_0(\mathrm{e}^V) < \infty$, where μ_0 is the centered Gaussian measure on $\mathbf{H}$ with covariance operator A^{-1}.

(H_2) There exists an increasing function $\psi : (0,\infty) \to [0,\infty)$ such that

$$|V(x)| \leq \frac{1}{2}\big(\psi(\varepsilon^{-1}) + \varepsilon|x|^2\big), \quad x \in \mathbf{H}, \varepsilon > 0.$$

(H_3) There exist constants $c > 0$ and $\theta \in [0, \lambda_1)$

$$|\nabla V(x)| \leq c + \theta|x|, \quad x \in \mathbf{H}.$$

Under (H_1), for any $\mathscr{F}_0$-measurable random variable X_0 on $\mathbf{H}$, (5.11) has a unique mild solution, and there exists an increasing function $\psi : [0,\infty) \to (0,\infty)$ such that

$$\mathbb{E}[\|\hat{X}_t\|_{\mathbf{H}}^2] \leq \psi(t)\big(1 + \mathbb{E}[|X_0|^2]\big), \quad t \geq 0,$$

see for instance [27, Theorem 3.1.1].

Let $\hat{P}_t$ be the associated Markov semigroup, i.e.

$$\hat{P}_t f(x) := \mathbb{E}^x[f(\hat{X}_t)], \quad t \geq 0, f \in \mathscr{B}_b(\mathbf{H}), \quad x \in \mathbf{H},$$

where $\mathscr{B}_b(\mathbf{H})$ is the class of all bounded measurable functions on $\mathbf{H}$, and $\mathbb{E}^x$ is the expectation for the solution X_t of (5.11) with $X_0 = x$. In general, for a probability measure ν on $\mathbf{H}$, let $\mathbb{E}^\nu$ be the expectation for X_t with initial distribution ν.

By (H_1), we define the probability measure

$$\mu(\mathrm{d}x) := Z_V^{-1}\mathrm{e}^{V(x)}\mu_0(\mathrm{d}x).$$

Then $\hat{P}_t$ is symmetric in $L^2(\mu)$.

Corollary 5.2. *Assume* (H_1) *and* (H_2)*. Let*

$$\xi(t) := \inf_{r\in(0,1)}\left(\frac{1}{t}r^{-1}e^{kr^{-1}+\psi(kr^{-1})}+r^{1-\delta}\right),\quad t>0.$$

Let $X_t=\hat{X}_{S_t}$ *for an increasing stable process* S_t *with* $S_0=0$ *which is independent of* X*.*

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

$$\mathbb{E}^\mu\big[\mathbb{W}_2(\mu_t,\mu)^2\big]\le c\xi(t),\quad t>0. \tag{5.13}$$

(2) *If* (H_3) *holds, then there exists a constant* $k>0$ *such that*

$$\big(\mathbb{E}^x[\mathbb{W}_2(\mu_t,\mu)]\big)^2\le k e^{k|x|^2}\xi(t-1),\quad t>1, x\in\mathbf{H}. \tag{5.14}$$

Proof. By inequality (25) in [31], (H_1) and (H_2) imply (5.3) for

$$\gamma(r) := e^{\psi(k\varepsilon^{-1})+k\varepsilon^{-1}}$$

for some constant $k>0$. By (5) in [31],

$$\|P_t-\mu\|_{L^2(\mu)}\le e^{-\lambda_0 t},\quad t>0$$

for some constant $\lambda_0>0$, which implies (4.1) for $p=2$ and $\beta=\frac{1}{2}$ according to the spectral representation. Moreover, since $k:=\sum_{i=1}^\infty\lambda_i^{-\delta}<\infty$,

$$\sum_{i=1}^\infty\int_0^t e^{-2\lambda_i(t-s)}ds\le\sum_{i=1}^\infty\lambda_i^{-\delta}t^{1-\delta}=kt^{1-\delta},$$

so that (5.1) holds for $p=2$ and $h(t)=kt^{1-\delta}$. Therefore, (5.13) follows form (5.4). If moreover (H_3) holds, then as shown in the proof of [31, Corollary 2.2(2)], we have $p_1(x,x)\le ce^{c|x|^2}$ for some constant $c>0$, so that (5.14) is implied by (5.5). □

The following example extends [31, Example 2.1].

Example 5.2. Assume (H_1) and that such that V is Lipschitz continuous, and let $X_t=\hat{X}_{S_t}$ for an increasing stable process S_t with $S_0=0$ which is independent of X. Then (H_2) holds for $\psi(s)=c_0 s$ for some constant $c_0>0$. So, by taking $r=N(\log t)^{-1}$ for a large enough constant $N>0$, we find a constant $c>0$ such that (5.13) and (5.14) imply that for large $t>0$

$$\mathbb{E}^\mu\big[\mathbb{W}_2(\mu_t,\mu)^2\big]\le c_1(\log t)^{\delta-1},$$
$$\big(\mathbb{E}^x[\mathbb{W}_2(\mu_t,\mu)]\big)^2\le ce^{c|x|^2}(\log t)^{\delta-1},\quad x\in\mathbf{H}.$$

6 Some General Estimates on Wasserstein Distance

In this part, we introduce some useful techniques in estimating the Wasserstein distance for empirical measures of diffusion processes.

Let (M, ρ) be a length space, and recall that for a Lipschitz continuous function f

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

Let $(\hat{L}, \mathscr{D}(\hat{L}))$ be the self-adjoint Dirichlet operator in $L^2(\mu)$ with Dirichlet form $(\hat{\mathscr{E}}, \mathscr{D}(\hat{\mathscr{E}}))$ satisfying $C_{b,L}(M) \subset \mathscr{D}(\hat{\mathscr{E}})$ and

$$\hat{\mathscr{E}}(f, f) = \mu(|\nabla f|^2), \quad f \in C_{b,L}(M).$$

Let X_t be the diffusion process generated by

$$L := \hat{L} + Z,$$

where Z is a bounded vector field with $\mathrm{div}_\mu(Z) = 0$, i.e.

$$\|Z\|_\infty := \sup_{\|f\|_{Lip} \le 1} |Zf| < \infty,$$

$$\int_M Zf \mathrm{d}\mu = 0, \quad f \in C_{b,L}(M).$$

6.1 Upper bound estimate

According to [17, Theorem 2], for any probability density f of μ, we have

$$\mathbb{W}_p(f\mu, \mu)^p \le p^p \mu\big(|\nabla(-\hat{L})^{-1}(f - 1)|^p\big), \quad p \in [1, \infty). \tag{6.1}$$

The idea of the proof goes back to [4], in which the following estimate is presented for probability density functions f_1, f_2:

$$\mathbb{W}_2(f_1\mu_1, f_2\mu_2)^2 \le \int_M \frac{|\nabla(-\hat{L})^{-1}(f_2 - f_1)|^2}{\mathscr{M}(f_1, f_2)} \mathrm{d}\mu, \tag{6.2}$$

where $\mathscr{M}(a, b) := 1_{\{a \wedge b > 0\}} \frac{\log a - \log b}{a - b}$ for $a \ne b$, and $\mathscr{M}(a, a) = 1_{\{a > 0\}} a^{-1}$. In general, for $p \ge 1$, denote $\mathscr{M}_p = \mathscr{M}$ if $p = 2$, and when $p \ne 2$ let

$$\mathscr{M}_p(a, b) = 1_{\{a \wedge b > 0\}} \frac{a^{2-p} - b^{2-p}}{(2 - p)(a - b)} \text{ for } a \ne b, \quad \mathscr{M}_p(a, a) = 1_{\{a > 0\}} a^{1-p}.$$

The following result extends (6.1) and (6.2).

Theorem 6.1. *For any probability density functions f_1 and f_2 with respect to μ such that $f_1 \vee f_2 > 0$,*

$$\mathbb{W}_p(f_1\mu, f_2\mu)^p \le \min\bigg\{ p^p 2^{p-1} \int_M \frac{|\nabla(-\hat{L})^{-1}(f_2 - f_1)|^p}{(f_1 + f_2)^{p-1}} \mathrm{d}\mu, \ p^p \int_M \frac{|\nabla(-\hat{L})^{-1}(f_2 - f_1)|^p}{f_1^{p-1}} \mathrm{d}\mu,$$
$$\int_M \frac{|\nabla(-\hat{L})^{-1}(f_2 - f_1)|^p}{\mathscr{M}_p(f_1, f_2)} \mathrm{d}\mu \bigg\}.$$

Proof. It suffices to prove for $p > 1$. Consider the Hamilton-Jacobi semigroup $(Q_t)_{t>0}$ on $C_{b,L}(M)$:

$$Q_t\phi := \inf_{x\in M}\Big\{\phi(x) + \frac{1}{pt^{p-1}}\rho(x,\cdot)^p\Big\},\quad t>0, \phi\in C_{b,L}(M).$$

Then for any $\phi \in C_{b,L}(M)$, $Q_0\phi := \lim_{t\downarrow 0} Q_t\phi = \phi$, $\|\nabla Q_t\phi\|_\infty$ is locally bounded in $t \geq 0$, and $Q_t\phi$ solves the Hamilton-Jacobi equation

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_t\phi = -\frac{p-1}{p}|\nabla Q_t\phi|^{\frac{p}{p-1}},\quad t>0. \tag{6.3}$$

Let $q = \frac{p}{p-1}$. For any $f \in C_b^1(M)$, and any increasing function $\theta \in C^1((0,1))$ such that $\theta_0 := \lim_{s\to 0}\theta_s = 0, \theta_1 := \lim_{s\to 1}\theta_s = 1$, by (6.3) and the integration by parts formula, we obtain

$$\begin{aligned}
&\mu_1(Q_1 f) - \mu_2(f) = \int_0^1 \Big\{\frac{\mathrm{d}}{\mathrm{d}s}\mu\big([f_1+\theta_s(f_2-f_1)]Q_s f\big)\Big\}\mathrm{d}s\\
&= \int_0^1 \mathrm{d}s\int_M \Big\{\theta_s'(f_2-f_1)Q_s f - \frac{f_1+\theta_s(f_2-f_1)}{q}|\nabla Q_s f|^q\Big\}\mathrm{d}\mu\\
&= \int_0^1 \Big[\theta_s'\hat{\mathscr{E}}\big((-\hat{L})^{-1}(f_2-f_1), Q_s f\big) - \mu\Big(\frac{f_1+\theta_s(f_2-f_1)}{q}|\nabla Q_s f|^q\Big)\Big]\mathrm{d}s\\
&\leq \int_0^1 \mu\Big(\theta_s'\big|\nabla(-\hat{L})^{-1}(f_2-f_1)\big|\cdot\big|\nabla Q_s f\big| - \frac{f_1+\theta_s(f_2-f_1)}{q}|\nabla Q_s f|^q\Big)\mathrm{d}s.
\end{aligned}$$

Combining this with Young's inequality $ab \leq a^p/p + b^q/q$ for $a, b \geq 0$, we arrive at

$$\mu_1(Q_1 f) - \mu_2(f) \leq \frac{1}{p}\int_M |\nabla(-\hat{L})^{-1}(f_2-f_1)|^p\mathrm{d}\mu\int_0^1 \frac{|\theta_s'|^p}{[f_1+\theta_s(f_2-f_1)]^{p-1}}\mathrm{d}s. \tag{6.4}$$

By Kantorovich duality formula

$$\frac{1}{p}\mathbb{W}_p(\mu_1,\mu_2)^p = \sup_{f\in C_b^1(M)}\big\{\mu_1(Q_1 f) - \mu_2(f)\big\},$$

and noting that

$$\begin{aligned}
&f_1+\theta_s(f_2-f_1) = f_1+f_2-\theta_s f_1 - (1-\theta_s)f_2\\
&= (f_1+f_2)\Big(1-\frac{\theta_s f_1}{f_1+f_2} - \frac{(1-\theta_s)f_2}{f_1+f_2}\Big)\\
&\geq (f_1+f_2)\min\{1-\theta_s,\theta_s\},
\end{aligned}$$

we deduce from (6.4) that

$$\mathbb{W}_p(\mu_1,\mu_2)^p \leq \int_0^1 \frac{|\theta_s'|^p}{\min\{\theta_s, 1-\theta_s\}^{p-1}}\mathrm{d}s\int_M \frac{|\nabla(-\hat{L})^{-1}(f_1-f_2)|^p}{(f_1+f_2)^{p-1}}\mathrm{d}\mu. \tag{6.5}$$

By taking

$$\theta_s = 1_{[0,\frac{1}{2}]}(s)2^{p-1}s^p + 1_{(\frac{1}{2},1]}(s)\big\{1-2^{p-1}(1-s)^p\big\},$$

which satisfies

$$\theta_s' = p2^{p-1}\min\{s,1-s\}^{p-1},\quad \min\{\theta_s,1-\theta_s\} = 2^{p-1}\min\{s,1-s\}^p,$$

we deduce from (6.5) that

$$\mathbb{W}_p(f_1\mu,f_2\mu)^p \le p^p 2^{p-1}\int_M \frac{|(-\hat{L})^{-\frac{1}{2}}(f_2-f_1)|^p}{(f_1+f_2)^{p-1}}\mathrm{d}\mu.$$

Next, (6.5) with $\theta_s = 1-(1-s)^p$ implies

$$\mathbb{W}_p(f_1\mu,f_2\mu)^p \le p^p\int_M \frac{|(-\hat{L})^{-\frac{1}{2}}(f_2-f_1)|^p}{f_1^{p-1}}\mathrm{d}\mu.$$

Finally, with $\theta_s = s$ we deduce from (6.4) that

$$\mathbb{W}_p(f_1\mu,f_2\mu)^p \le \int_M \frac{|(-\hat{L})^{-\frac{1}{2}}(f_2-f_1)|^p}{\mathscr{M}_p(f_1,f_2)}\mathrm{d}\mu.$$

Then the proof is finished. □

We now apply Theorem 6.1 to the regularized empirical measure

$$\mu_{t,\varepsilon} := \hat{P}_\varepsilon^*\mu_t$$

for suitable choice of $\varepsilon=\varepsilon_t\downarrow 0$ as $t\uparrow\infty$. To this end, we make the following assumption on the reference diffusion process $\hat{X}_t$ introduced in Section 4.

(A) The following conditions hold for some $d\in[1,\infty)$ and an increasing function $K:[2,\infty)\to(0,\infty)$.

- **Nash inequality.** There exists a constant $C>0$ such that

$$\mu(f^2)\le C\hat{\mathscr{E}}(f,f)^{\frac{d}{d+2}}\mu(|f|)^{\frac{4}{d+2}},\quad f\in\mathscr{D}_0 := \big\{f\in\mathscr{D}(\hat{\mathscr{E}}):\ \mu(f)=0\big\}. \tag{6.6}$$

- **Continuity of symmetric diffusion.** For any $p\in[2,\infty)$,

$$\mathbb{E}^\mu[\rho(\hat{X}_0,\hat{X}_t)^p] = \int_{M\times M}\rho(x,y)^p\hat{p}_t(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\le K(p)t^{\frac{p}{2}},\quad t\in[0,1], \tag{6.7}$$

  where $\hat{p}_t$ is the heat kernel of $\hat{P}_t$ with respect to μ.

- **Boundedness of Riesz transform.** For any $p\in[2,\infty)$,

$$\|\nabla(-\hat{L})^{-\frac{1}{2}}f\|_{L^p(\mu)}\le K(p)\|f\|_{L^p(\mu)},\quad f\in L^p(\mu)\text{ with }\mu(f)=0. \tag{6.8}$$

It is well known that **(A)** holds for the (reflecting) diffusion process generated by $\hat{L} := \Delta + \nabla V$ considered in Introduction.

Besides the elliptic diffusion process on compact manifolds, some criteria on the Nash inequality (6.6) are available in [26, Section 3.4]. In general, (6.6) implies that for some constant $c_0 > 0$,

$$\|\hat{P}_t - \mu\|_{L^p(\mu)\to L^q(\mu)} \le c_0(1\wedge t)^{-\frac{d(q-p)}{2pq}} \mathrm{e}^{-\lambda_1 t}, \quad t>0,\ 1\le p\le q\le\infty, \tag{6.9}$$

and that $-\hat{L}$ has purely discrete spectrum with all eigenvalues $\{\lambda_i\}_{i\ge 0}$, which are listed in the increasing order counting multiplicities, satisfy

$$\lambda_i \ge c_1 i^{\frac{2}{d}}, \quad i\ge 0, \tag{6.10}$$

for some constant $c_1>0$. The Markov semigroup $\hat{P}_t$ generated by $\hat{L}$ has symmetric heat kernel $\hat{p}_t$ with respect to μ formulated as

$$\hat{p}_t(x,y) = 1 + \sum_{i=1}^{\infty} \mathrm{e}^{-\lambda_i t}\phi_i(x)\phi_i(y), \quad t>0,\ x,y\in M. \tag{6.11}$$

All these assertions can be found for instance in [10].

The condition (6.7) is natural for diffusion processes due to the growth property $\mathbb{E}|B_t - B_0|^p \le ct^{\frac{p}{2}}$ for the Brownian motion B_t. There are plentiful results on the boundedness condition (6.8) for the Riesz transform, see [5, 7, 9] and references therein.

The following result shows that under assumption **(A)**, the convergence rate of $(\mathbb{E}\mathbb{W}_p^q(\mu_T,\mu))^{\frac{1}{q}}$ is given by

$$\gamma_d(t) := \begin{cases} t^{-\frac{1}{2}}, & \text{if } d\in[1,4),\\ t^{-\frac{1}{2}}\sqrt{\log t}, & \text{if } d=4,\\ t^{-\frac{1}{d-2}}, & \text{if } d\in(4,\infty). \end{cases}$$

Theorem 6.2 ([36]). *Assume* **(A)**. *Then for any* $(k,p,q)\in(1,\infty]\times[1,\infty)\times(0,\infty)$, *there exists a constant* $c\in(0,\infty)$ *such that*

$$\big(\mathbb{E}^\nu[\mathbb{W}_p^q(\mu_t,\mu)]\big)^{\frac{1}{q}} \le c\|h_\nu\|_{L^k(\mu)}^{\frac{1}{q}}\gamma_d(t), \quad t\ge 2,\ \nu = h_\nu\mu\in\mathscr{P} \text{ with } h_\nu\in L^k(\mu), \tag{6.12}$$

where $\nu = h_\nu\mu$ *means* $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = h_\nu$.

6.2 Lower bound estimate

To derive sharp lower bound estimates, we make the following assumption **(B)** which holds in particular for the (reflecting) diffusion operator $\hat{L} := \Delta + \nabla V$ on a d-dimensional compact connected Riemannian manifold, since in this case conditions (6.13) and (6.14) are well known, and the other conditions have been verified by [37, Lemma 5.2]. For M being a smooth domain in $\mathbb{R}^d$, (6.15) is known as Sard's lemma (see [13, p130, Excercise 5.5]) and has been discussed in [6, Section 3.1.6]. The function f_ξ in (6.16) is called Lusin's approximation of h (see [1, 15]).

(B) Let $\{\lambda_i\}_{i\ge 0}$ be all eigenvalues of $-\hat{L}$ listed in the increasing order with multiplicities. There exist constants $\kappa > 0$ and $d \in [1,\infty)$ such that

$$\lambda_i \le \kappa\, i^{\frac{2}{d}}, \quad i \ge 0, \tag{6.13}$$

$$\mathbb{W}_1(\nu \hat{P}_t, \mu) \le \kappa\, \mathbb{W}_1(\nu,\mu), \quad t \in [0,1],\, \nu \in \mathscr{P}. \tag{6.14}$$

Moreover, for any $f \in \mathscr{D}(\hat{\mathscr{E}})$,

$$\mu\big(\{|\nabla f| > 0, f = 0\}\big) = 0, \tag{6.15}$$

and there exists a constant $c > 0$ independent of f such that

$$\mu\big(f \ne f_\xi\big) \le \frac{c}{\xi^2}\int_M |\nabla f|^2 \mathrm{d}\mu, \quad \xi > 0 \tag{6.16}$$

holds for a family of functions $\{f_\xi : \xi > 0\}$ on M with $\|\nabla f_\xi\|_\infty \le \xi$.

Theorem 6.3 ([36])**.** *Assume* **(A)** *and* **(B)**. *Then for any $q \in (0,\infty)$, we have for large $t > 0$,*

$$\inf_{\nu\in\mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_1^q(\mu_t,\mu)] \succeq \gamma_d(t)^q.$$

Finally, we present a lower bound estimate which also applies to infinite-dimensions and generalizes [16, Proposition 4.2] for the finite-dimensional setting.

Theorem 6.4. *Let $\mu \in \mathscr{P}(E)$ such that*

$$\sup_{x\in E} \mu(B(x,r)) \le \psi(r), \quad r \ge 0 \tag{6.17}$$

holds for some increasing function ψ, where $B(x,r) := \{y \in E : \rho(x,y) < r\}$. Then for any $N \ge 1$ and any probability measure μ_N supported on a set of N points in E,

$$\mathbb{W}_p(\mu_N,\mu) \ge 2^{-\frac{1}{p}}\psi^{-1}\Big(\frac{1}{2N}\Big), \tag{6.18}$$

where $\psi^{-1}(s) := \sup\{r \ge 0 : \psi(r) \le s\}, s \ge 0$.

Proof. Let $D = \mathrm{supp}\mu_N$ which contains N many points, so that from (6.17) we conclude that $D_r := \cup_{x\in D} B(x,r)$ satisfies

$$\mu(D_r) \le \sum_{x\in D} \mu(B(x,r)) \le N\psi(r), \quad r \ge 0.$$

Therefore, for any $\pi \in \mathscr{C}(\mu_N,\mu)$, we get

$$\int_{E\times E} \rho(x,y)^p \pi(\mathrm{d}x,\mathrm{d}y) \ge \int_{D\times D_r^c} r^p \pi(\mathrm{d}x,\mathrm{d}y) = r^p\mu(D_r^c) \ge r^p\{1 - N\psi(r)\}, \quad r \ge 0.$$

Taking $r = \psi^{-1}(1/(2N))$ we derive

$$\mathbb{W}_p(\mu,\nu)^p \ge \sup_{r\ge 0} r^p[1 - N\psi(r)] \ge \frac{1}{2}\big\{\psi^{-1}(1/(2N))\big\}^p.$$

□

To apply Theorem 6.4 to the empirical measure μ_t, we only need to compare μ_t with the discretize empirical measure

$$\mu_{t,N} := \frac{1}{N} \sum_{i=1}^{N} \delta_{X_{it/N}},$$

with suitable choice of $N = N_t \to \infty$ as $t \to \infty$.

Acknowledgement. The author would like to thanks the referee for helpful comments and corrections.

References

[1] E. Acerbi, N. Fusco, *Semicontinuity problems in the calculus of variations,* Arch. Rational Mech. Anal. 86(1984), 125–145.

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

[3] L. Ambrosio, N. Gigli, G. Savaré, *Gradient Flows in Metric Spaces and in the Spaces of Probability Measures,* Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.

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

[5] D. Bakry, *Étude des transformations de Riesz dans les variétés à courbure de Ricci minorée,* Séminaire de Probabilités XXI, Lecture Notes in Math. 1247(1987), 137–172.

[6] D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators,* Grundlehren der mathematischen Wissenschaften 348. Springer, 2014.

[7] A. Carbonaro, O. Dragičević, *Bellman function and dimension-free estimates in a theorem of Bakry,* J. Funct. Anal. 265(2013), 1085–1104.

[8] I. Chavel, *Eigenvalues in Riemannian Geometry,* Acad. Press, 1984.

[9] L.-J. Cheng, A. Thalmaier, F.-Y. Wang, *Covariant Riesz transform on differential forms for* $1 < p \leq 2$*,* Cal. Var. Part. Diff. Equat. 62(2023), No. 245.

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

[11] K. Du, Y. Jiang, J. Li, *Empirical approximation to invariant measures for McKean-Vlasov processes: mean-field interaction vs self-interaction,* Bernoulli 29(2023), 2492–2518.

[12] S. German, D. German, *Stochastic relaxation, Gibbs distribution, and the Baysian restoration of images,* IEEE Trans. Pattern Anal. Machine Intelligence, 6(1984), 721–41.

[13] A. Grigory'an, *Heat Kernel and Analysis on Manifolds,* American Mathematical Society, Providence, RI; International Press, Boston, MA, 2009.

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

[15] F.-C. Liu, *A Luzin type property of Sobolev functions,* Indiana Univ. Math. J. 26(1977), 645–651.

[16] B. Kloeckner, *Approximation by finitely supported measures,* ESAIM Control Optim. Calc. Var. 18(2012), 343–359.

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

[18] H. Li, B. Wu, *Wasserstein convergence for empirical measures of subordinated Dirichlet diffusions on Riemannian manifolds,* arXiv:2206.03901.

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

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

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

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

[23] D. Trevisan, F.-Y. Wang, J.-X. Zhu, *Wasserstein asymptotics for empirical measures of diffusions on four dimensional closed manifolds,* arXiv:2410.21981.

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

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

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

[27] F.-Y. Wang, *Harnack Inequality for Stochastic Partial Differential Equations,* Math. Brief. Springer, 2013

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

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

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

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

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

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

[34] F.-Y. Wang, *Wasserstein convergence rate for empirical measures of Markov processes,* App. Math. Opt. 92(2025), https://doi.org/10.1007/s00245-025-10275-1.

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

[36] W., B.-Y. Wu, J.-X. Zhu, *Sharp L^q-convergence rate in p-Wasserstein distance for empirical measures of diffusion processes,* arXiv:2408.09116.

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

[38] B. Wu, J.-X. Zhu, *Wasserstein convergence rates for empirical measures of random subsequence of* $\{n\alpha\}$, Stochastic Process. Appl. 181(2025), Paper No. 104534, 20 pp.

[39] L. Wu, *Moderate deviations of dependent random variables related to CLT,* Ann. Probab. 23(1995), 420–445.

[40] J.-X. Zhu, *Asymptotic behavior of Wasserstein distance for weighted empirical measures of diffusion processes on compact Riemannian manifolds,* Markov Process. Related Fields 30 (2024), 357–397.