Wasserstein Convergence Rate for Empirical Measures on Noncompact Manifolds *

Feng-Yu Wang[a),b)]

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

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

November 20, 2021

Abstract

Let X_t be the (reflecting) diffusion process generated by $L := \Delta + \nabla V$ on a complete connected Riemannian manifold M possibly with a boundary ∂M, where $V \in C^1(M)$ such that $\mu(\mathrm{d}x) := \mathrm{e}^{V(x)}\mathrm{d}x$ is a probability measure. We estimate the convergence rate for the empirical measure $\mu_t := \frac{1}{t}\int_0^t \delta_{X_s}\mathrm{d}s$ under the Wasserstein distance. As a typical example, when $M = \mathbb{R}^d$ and $V(x) = c_1 - c_2|x|^p$ for some constants $c_1 \in \mathbb{R}, c_2 > 0$ and $p > 1$, the explicit upper and lower bounds are present for the convergence rate, which are of sharp order when either $d < \frac{4(p-1)}{p}$ or $d \geq 4$ and $p \to \infty$.

AMS subject Classification: 60D05, 58J65.
Keywords: Empirical measure, diffusion process, Wasserstein distance, Riemannian manifold.

1 Introduction

Let M be a d-dimensional complete connected Riemannian manifold, possibly with a boundary ∂M. Let $V \in C^1(M)$ such that $Z_V := \int_M \mathrm{e}^{V(x)}\mathrm{d}s < \infty$, where $\mathrm{d}x := \mathrm{vol}(\mathrm{d}x)$ stands for the Riemannian volume measure. Then $\mu(\mathrm{d}x) := Z_V^{-1}\mathrm{e}^{V(x)}\mathrm{d}x$ is a probability measure, and the (reflecting if ∂M exists) diffusion process X_t generated by $L := \Delta + \nabla V$ is reversible with stationary distribution μ. When M is compact, the convergence rate of the empirical measure

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

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

under the Wasserstein distance is investigated in [19]. More precisely, let ρ be the Riemannian distance on M, and let

$$\mathbb{W}_2(\mu_1,\mu_2) := \inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)} \|\rho\|_{L^2(\pi)}$$

be the associated L^2-Warsserstein distance for probability measures on M, where $\mathscr{C}(\mu_1,\mu_2)$ is the class of all couplings of μ_1 and μ_2. For two positive functions ξ,η of t, we denote $\xi(t)\sim\eta(t)$ if $c^{-1}\le\frac{\xi(t)}{\eta(t)}\le c$ holds for some constant $c>1$ and large $t>0$. According to [19], for large $t>0$ we have

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

where the lower bound estimate on $\mathbb{E}[\mathbb{W}_2(\mu_t,\mu)^2]$ for $d=4$ is only derived for a typical example that M is the 4-dimensional torus and $V=0$. Moreover, when ∂M is either convex or empty, we have

$$\lim_{t\to\infty} t\mathbb{E}[\mathbb{W}_2(\mu_t,\mu)^2]=\sum_{i=1}^{\infty}\frac{2}{\lambda_i^2}, \tag{1.1}$$

where $\{\lambda_i\}_{i\ge1}$ are all non-trivial eigenvalues of $-L$ (with Neumann boundary condition if ∂M exists) listed in the increasing order counting multiplicities. See [17, 18] for further studies on the conditional empirical measure of the L-diffusion process with absorbing boundary.

In this note, we investigate the convergence rate of $\mathbb{E}[\mathbb{W}_2(\mu_t,\mu)^2]$ for non-compact Riemannian manifold M.

1.1 Upper bound estimate

We first present a result on the upper bound estimate of $\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2]$, where $\mathbb{E}^\nu$ is the expectation for the diffusion process with initial distribution ν. When $\nu=\delta_x$ is a Dirac measure, we simply denote $\mathbb{E}^x=\mathbb{E}^{\delta_x}$.

Let $p_t(x,y)$ be the heat kernel of the (Neumann) Markov semigroup P_t generated by L. We will assume

$$\gamma(t) := \int_M p_t(x,x)\mu(\mathrm{d}x)<\infty,\quad t>0. \tag{1.2}$$

By [12, Theorem 3.3] (see also [14, Theorem 3.3.19]) and the spectral representation of heat kernel, (1.2) holds if and only if L has discrete spectrum such that all eigenvalues $\{\lambda_i\}_{i\ge0}$ of $-L$ listed in the increasing order satisfy

$$\sum_{i=0}^{\infty}\mathrm{e}^{-\lambda_i t}<\infty,\quad t>0.$$

Since M is connected, the trivial eigenvalue $\lambda_0=0$ is simple, so that

$$\lambda_1 := \inf\big\{\mu(|\nabla f|^2):\ f\in C_b^1(M),\mu(f)=0,\mu(f^2)=1\big\}>0. \tag{1.3}$$

The first non-trivial eigenvalue λ_1 is called the spectral gap of L, and (1.3) is known as the Poincaré inequality.

In particular, (1.2) holds if P_t is ultracontractive, i.e.

$$\sup_{x,y\in M} p_t(x,y) = \|P_t\|_{L^1(\mu)\to L^\infty(\mu)} < \infty, \quad t>0.$$

Since $\gamma(t)$ is decreasing in t, (1.2) implies

$$\beta(\varepsilon) := 1 + \int_\varepsilon^1 \mathrm{d}s \int_s^1 \gamma(t)\mathrm{d}t < \infty, \quad \varepsilon\in(0,1]. \tag{1.4}$$

Moreover, let

$$\alpha(\varepsilon) := \mathbb{E}^\mu[\rho(X_0,X_\varepsilon)^2] = \int_M \rho(x,y)^2 p_\varepsilon(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y), \quad \varepsilon>0. \tag{1.5}$$

Finally, for any $k\ge 1$, let $\mathscr{P}_k = \{\nu\in\mathscr{P} : \nu = h_\nu\mu, \|h_\nu\|_\infty \le k\}$, where $\mathscr{P}$ is the set of all probability measures on M.

Theorem 1.1. *Assume* (1.2)*.*

(1) *For any* $k\ge 1$*,*

$$\limsup_{t\to\infty}\Big\{t\sup_{\nu\in\mathscr{P}_k}\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2]\Big\} \le \sum_{i=1}^\infty \frac{8}{\lambda_i^2}. \tag{1.6}$$

If P_t *is ultracontractive, then*

$$\limsup_{t\to\infty}\Big\{t\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2]\Big\} \le \sum_{i=1}^\infty \frac{8}{\lambda_i^2} \tag{1.7}$$

holds for $\nu\in\mathscr{P}$ *satisfying*

$$\int_0^1 \mathrm{d}s \int_M \mathbb{E}^\nu\big[\rho(x,X_s)^2\big]\mu(\mathrm{d}x) < \infty. \tag{1.8}$$

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

$$\sup_{\nu\in\mathscr{P}_k}\mathbb{E}^\nu\mathbb{W}_2(\mu_t,\mu)^2 \le ck\inf_{\varepsilon\in(0,1]}\big\{\alpha(\varepsilon)+t^{-1}\beta(\varepsilon)\big\}, \quad t,k\ge 1. \tag{1.9}$$

If P_t *is ultracontravtive, then there exists a constant* $c>0$ *such that for any* $\nu\in\mathscr{P}$ *and* $t\ge 1$*,*

$$\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2] \le c\bigg\{\frac{1}{t}\int_0^1 \mathbb{E}^\nu\big[\mu\big(\rho(X_s,\cdot)^2\big)\big]\mathrm{d}s + \inf_{\varepsilon\in(0,1]}\big\{\alpha(\varepsilon)+t^{-1}\beta(\varepsilon)\big\}\bigg\}. \tag{1.10}$$

Since the conditions (1.2), (1.5) and (1.8) are less explicit, for the convenience of applications we present the following consequence of Theorem 1.1.

Corollary 1.2. *Assume that $\partial M=\emptyset$ or ∂M is convex outside a compact set. Let $V=V_1+V_2$ for some functions $V_1,V_2\in C^1(M)$ such that*

$$\mathrm{Ric}_{V_1}:=\mathrm{Ric}-\mathrm{Hess}_{V_1}\ge -K,\quad \|\nabla V_2\|_\infty\le K \tag{1.11}$$

holds for some constant $K>0$, where Ric *is the Ricci curvature and* Hess *denotes the Hessian tensor. For any $t,\varepsilon>0$, let*

$$\tilde\gamma(t):=\int_M\frac{\mu(\mathrm{d}x)}{\mu(B(x,\sqrt{t}))},\quad \tilde\beta(\varepsilon):=1+\int_\varepsilon^1\mathrm{d}s\int_s^1\tilde\gamma(r)\mathrm{d}r.$$

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

$$\sup_{\nu\in\mathscr{P}_k}\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2]\le ck\inf_{\varepsilon\in(0,1]}\big\{\varepsilon+t^{-1}\tilde\beta(\varepsilon)\big\},\quad t,k\ge1. \tag{1.12}$$

(2) *If $\|P_t\mathrm{e}^{\lambda\rho_o^2}\|_\infty<\infty$ for $\lambda,t>0$, then for any $t\ge1$ and $\nu\in\mathscr{P}$,*

$$\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2]\le c\Big[t^{-1}\nu(|\nabla V|^2)+\inf_{\varepsilon\in(0,1]}\big\{\varepsilon+t^{-1}\tilde\beta(\varepsilon)\big\}\Big]. \tag{1.13}$$

1.2 Lower bound estimate

Consider the modified L^1-Warsserstein distance

$$\tilde W_1(\mu_1,\mu_2):=\inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)}\int_{M\times M}\{1\wedge\rho(x,y)\}\pi(\mathrm{d}x,\mathrm{d}y)\le\mathbb{W}_2(\mu_1,\mu_2).$$

We have the following result.

Theorem 1.3. (1) *In general, there exists a constant $c>0$ such that*

$$\mathbb{E}^\mu[\tilde{\mathbb{W}}_1(\mu_t,\mu)^2]\ge ct^{-1},\quad t\ge1. \tag{1.14}$$

If (1.3) *holds, then*

$$\liminf_{t\to\infty}\Big\{t\mathbb{E}^\nu[\tilde{\mathbb{W}}_1(\mu_t,\mu)^2]\Big\}>0,\quad \nu\in\mathscr{P}. \tag{1.15}$$

(2) *Let ∂M be empty or convex, and let $d\ge3$. If $\mu(|\nabla V|)<\infty$ and*

$$\mathrm{Ric}\ge -K,\quad V\le K \tag{1.16}$$

holds for some constant $K>0$, then there exists a constant $c>0$ such that

$$\inf_{\nu\in\mathscr{P}_k}\mathbb{E}^\nu[\tilde W_1(\mu_t,\mu)]\ge c(kt)^{-\frac{1}{d-2}},\quad k,t\ge1, \tag{1.17}$$

and moreover

$$\liminf_{t\to\infty}\Big\{t^{\frac{1}{d-2}}\mathbb{E}^\nu[\tilde W_1(\mu_t,\mu)]\Big\}>0,\quad d\ge4,\nu\in\mathscr{P}. \tag{1.18}$$

(3) *Assume that P_t is ultracontractive, ∂M is either empty or convex, and $\mathrm{Ric}-\mathrm{Hess}_V\ge K$ for some constant $K\in\mathbb{R}$. Then*

$$\liminf_{t\to\infty}\inf_{\nu\in\mathscr{P}}\Big\{t^{-1}\mathbb{E}^\nu[W_2(\mu_t,\mu)^2]\Big\}\ge\sum_{i=1}^\infty\frac{2}{\lambda_i^2}. \tag{1.19}$$

Remark 1.1. According to Theorem 1.1(1) and Theorem 1.3(3), when P_t is ultracontractive, ∂M is either empty or convex, and $\text{Ric} - \text{Hess}_V \geq K$ for some constant $K \in \mathbb{R}$, we have

$$\sum_{i=1}^{\infty} \frac{2}{\lambda_i^2} \leq \liminf_{t\to\infty} \Big\{ t^{-1}\mathbb{E}^\nu[W_2(\mu_t,\mu)^2] \Big\} \leq \limsup_{t\to\infty} \Big\{ t^{-1}\mathbb{E}^\nu[W_2(\mu_t,\mu)^2] \Big\} \leq \sum_{i=1}^{\infty} \frac{8}{\lambda_i^2}, \quad \nu \in \mathscr{P}.$$

Because of (1.1) derived in [19] in the compact setting, we may hope that the same limit formula holds for the present non-compact setting. In particular, for the one-dimensional Ornstein-Uhlenbeck process where $M = \mathbb{R}, V(x) = -\frac{1}{2}|x|^2$ and $\lambda_i = i, i \geq 1$, we would guess

$$\lim_{t\to\infty} \Big\{ t\mathbb{E}^\mu[\mathbb{W}_2(\mu_t,\mu)^2] \Big\} = \sum_{i=1}^{\infty} \frac{2}{i^2}.$$

However, there is essential difficulty to prove the exact upper bound estimate as the corresponding calculations in [19] heavily depend on the estimate $\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} \leq ct^{-\frac{d}{2}}$ for some constant $c > 0$ and all $t \in (0,1]$, which is available only when M is compact.

1.3 Example

To illustrate Corollary 1.2 and Theorem 1.3, we consider a class of specific models, where the convergence rate is sharp when $d < \frac{4p-1}{p}$ as both upper and lower bounds behave as t^{-1}, and is asymptotically sharp when $d \geq 4$ and $p \to \infty$ for which both upper and lower bounds are of order $t^{-\frac{2}{d-2}}$. The assertions will be proved in Section 4.

Example 1.4. *Let $M = \mathbb{R}^d$ and $V(x) = -\kappa|x|^\alpha + W(x)$ for some constants $\kappa > 0, \alpha > 1$, and some function $W \in C^1(M)$ with $\|\nabla W\|_\infty < \infty$.*

(1) *There exists a constant $c > 0$ such that for any $t, k \geq 1$, we have*

$$\sup_{\nu\in\mathscr{P}_k} \mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2] \leq \begin{cases} ckt^{-\frac{2(\alpha-1)}{(d-2)\alpha+2}}, & \text{if } 4(\alpha-1) < d\alpha, \\ ckt^{-1}\log(1+t), & \text{if } 4(\alpha-1) = d\alpha, \\ ckt^{-1}, & \text{if } 4(\alpha-1) > d\alpha. \end{cases} \tag{1.20}$$

(2) *If $\alpha > 2$, then there exists a constant $c > 0$ such that for any $t \geq 1$,*

$$\sup_{x\in\mathbb{R}^d} \frac{\mathbb{E}^x[\mathbb{W}_2(\mu_t,\mu)^2]}{1+|x|^{2(\alpha-1)}} \leq \begin{cases} ct^{-\frac{2(\alpha-1)}{(d-2)\alpha+2}}, & \text{if } 4(\alpha-1) < d\alpha, \\ ct^{-1}\log(1+t), & \text{if } 4(\alpha-1) = d\alpha, \\ ct^{-1}, & \text{if } 4(\alpha-1) > d\alpha. \end{cases} \tag{1.21}$$

(3) *For any probability measure ν, there exists a constant $c > 0$ such that for large $t > 0$,*

$$\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2] \geq \mathbb{E}^\nu[\tilde{\mathbb{W}}_1(\mu_t,\mu)^2] \geq ct^{-\frac{2}{2\vee(d-2)}}.$$

2 Proofs of Theorem 1.1 and Corollary 1.2

By the spectral representation, the heat kernel of P_t is formulated as

$$(2.1)\qquad p_t(x,y) = 1 + \sum_{i=1}^\infty \mathrm{e}^{-\lambda_i t}\phi_i(x)\phi_i(y),\quad t>0, x,y\in M,$$

where $\{\phi_i\}_{i\ge1}$ are the associated unit eigenfunctions with respect to the non-trivial eigenvalues $\{\lambda_i\}_{i\ge1}$ of $-L$, with the Neumann boundary condition if ∂M exists.

We will use the following inequality due to [9, Theorem 2]

$$(2.2)\qquad \mathbb{W}_2(f\mu,\mu)^2 \le 4\mu(|\nabla(-L)^{-1}(f-1)|^2),\quad f\ge0, \mu(f)=1,$$

which is proved using an idea due to [2], see Theorem A.1 below for an extension to the upper bound on $\mathbb{W}_p(f_1\mu, f_2\mu)$. To apply (2.2), we consider the modified empirical measures

$$(2.3)\qquad \mu_{\varepsilon,t} := f_{\varepsilon,t}\mu,\quad \varepsilon>0, t>0,$$

where, according to (2.1),

$$(2.4)\qquad f_{\varepsilon,t} := \frac{1}{t}\int_0^t p_\varepsilon(X_s,\cdot) = 1+\sum_{i=1}^\infty \mathrm{e}^{-\lambda_i\varepsilon}\xi_i(t)\phi_i,\quad \xi_i(t) := \frac{1}{t}\int_0^t \phi_i(X_s)\mathrm{d}s.$$

Proof of Theorem 1.1. (1) It suffices to prove for $\sum_{i=1}^\infty \lambda_i^{-2}<\infty$. In this case, by [19, (2.19)] whose proof works under the condition (1.2), we find a constant $c>0$ such that

$$\sup_{\nu\in\mathscr{P}_k}\left| t\mathbb{E}^\nu[\mu(|(-L)^{-\frac12}(f_{\varepsilon,t}-1)|^2)] - \sum_{i=1}^\infty \frac{2}{\lambda_i^2\mathrm{e}^{2\varepsilon\lambda_i}}\right| \le \frac{ck}{t}\sum_{i=1}^\infty \frac{1}{\lambda_i^2\mathrm{e}^{2\varepsilon\lambda_i}}.$$

This together with (2.2) yields

$$(2.5)\qquad t\sup_{\nu\in\mathscr{P}_k}\mathbb{E}^\nu[\mathbb{W}_2(\mu_{\varepsilon,t},\mu)^2] \le \sum_{i=1}^\infty \frac{8}{\lambda_i^2} + \frac{ck}{t}\sum_{i=1}^\infty\frac{4}{\lambda_i^2},\quad \varepsilon>0.$$

To approximate μ_t using $\mu_{\varepsilon,t}$, for any $n\ge1$ let

$$\mathbb{W}_{2,n}(\mu_1,\mu_2) := \inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)}\left(\int_{M\times M}\{n\wedge\rho(x,y)^2\}\pi(\mathrm{d}x,\mathrm{d}y)\right)^{\frac12},\quad \mu_1,\mu_2\in\mathscr{P}.$$

Given $\gamma\in\mathscr{P}$, let $(X_s^\gamma)_{s\ge0}$ be the (reflecting, if $\partial M\ne\emptyset$) diffusion process generated by L with initial distribution γ, and let γP_s denote the distribution of X_s^γ. By the continuity of the diffusion process and the dominated convergence theorem, we have

$$\limsup_{\varepsilon\downarrow0}\mathbb{W}_{2,n}(\gamma P_\varepsilon,\gamma)^2 = 0,\quad n\ge1,\gamma\in\mathscr{P}.$$

Observing that $\mu_{\varepsilon,t} = \mu_t P_\varepsilon$, we have

$$\limsup_{\varepsilon\downarrow 0} \mathbb{W}_{2,n}(\mu_{\varepsilon,t},\mu_t)^2 = 0, \quad n\ge 1, t>0.$$

Since $\mathbb{W}_{2,n}(\mu_{\varepsilon,t},\mu_t)^2 \le n$ and $\nu \le k\mu$ for $\nu\in\mathscr{P}_k$, this and the dominated convergence theorem yield

$$\limsup_{\varepsilon\downarrow 0}\sup_{\nu\in\mathscr{P}_k} \mathbb{E}^\nu \mathbb{W}_{2,n}(\mu_{\varepsilon,t},\mu_t)^2 \le k \limsup_{\varepsilon\downarrow 0} \mathbb{E}^\mu \mathbb{W}_{2,n}(\mu_{\varepsilon,t},\mu_t)^2 = 0, \quad n\ge 1, t>0.$$

Combining this with (2.5) and applying the triangle inequality of $\mathbb{W}_{2,n}$, we derive

$$\begin{aligned} &t\sup_{\nu\in\mathscr{P}_k} \mathbb{E}^\nu[\mathbb{W}_{2,n}(\mu_t,\mu)^2] \le t \limsup_{\varepsilon\downarrow 0}\sup_{\nu\in\mathscr{P}_k}\big\{\mathbb{W}_{2,n}(\mu_{\varepsilon,t},\mu_t)+\mathbb{W}_{2,n}(\mu_{t,\varepsilon},\mu)\big\}^2\\ &\le \sum_{i=1}^\infty \frac{8}{\lambda_i^2} + \frac{ck}{t}\sum_{i=1}^\infty \frac{4}{\lambda_i^2}, \quad n\ge 1, t>0.\end{aligned}$$

Therefore, for any $t>0$ we have

$$t\sup_{\nu\in\mathscr{P}_k}\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\mu)^2] = t\sup_{n\ge 1,\nu\in\mathscr{P}_k}\mathbb{E}^\nu[\mathbb{W}_{2,n}(\mu_t,\mu)^2] \le \sum_{i=1}^\infty\frac{8}{\lambda_i^2}+\frac{ck}{t}\sum_{i=1}^\infty\frac{4}{\lambda_i^2}, \tag{2.6}$$

which implies (1.6).

Next, when P_t is ultracontractive, we have

$$\delta(\varepsilon) := \sup_{t\ge\varepsilon, x,y\in M} p_t(x,y) < \infty, \quad \varepsilon>0.$$

Then the distribution ν_ε of X_ε starting at ν is in the class $\mathscr{P}_{\delta(\varepsilon)}$. For any $\varepsilon\in(0,1]$, let

$$\bar\mu_{\varepsilon,t} := \frac{1}{t}\int_\varepsilon^{t+\varepsilon}\delta_{X_s}\mathrm{d}s.$$

By the Markov property and (2.6), we obtain

$$\limsup_{t\to\infty}\Big\{t\mathbb{E}^\nu[\mathbb{W}_2(\bar\mu_{\varepsilon,t},\mu)^2]\Big\} = \limsup_{t\to\infty}\Big\{t\mathbb{E}^{\nu_\varepsilon}[\mathbb{W}_2(\mu_t,\mu)^2]\Big\} \le \sum_{i=1}^\infty\frac{8}{\lambda_i^2}, \quad \varepsilon>0. \tag{2.7}$$

On the other hand, since

$$\pi := \frac{1}{t}\int_0^\varepsilon \delta_{(X_s,X_{s+t})}\mathrm{d}s + \frac{1}{t}\int_\varepsilon^t \delta_{(X_s,X_s)}\mathrm{d}s \in \mathscr{C}(\mu_t,\bar\mu_{\varepsilon,t}),$$

and since the conditional distribution of X_{s+t} given X_s is bounded above by $\delta(1)\mu$ for $t\ge 1$, we have

$$t\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\bar\mu_{\varepsilon,t})^2] \le t\mathbb{E}^\nu\int_{M\times M}\rho(x,y)^2\pi(\mathrm{d}x,\mathrm{d}y)$$

$$= \int_0^\varepsilon \mathbb{E}^\nu[\rho(X_s, X_{s+t})^2]\mathrm{d}s \le \delta(1)\int_0^\varepsilon \mathbb{E}^\nu\big[\mu\big(\rho(X_s,\cdot)^2\big)\big]\mathrm{d}s =: r_\varepsilon.$$

Combining this with (1.8), (2.7), and applying the triangle inequality of $\mathbb{W}_2$, we arrive at

$$\begin{aligned}
&\limsup_{t\to\infty}\Big\{t\mathbb{E}^\nu[\mathbb{W}_2(\bar\mu_t,\mu)^2]\Big\}\\
&\le \lim_{\varepsilon\downarrow 0}\bigg((1+r_\varepsilon^{\frac{1}{2}})\limsup_{t\to\infty}\Big\{t\mathbb{E}^\nu[\mathbb{W}_2(\bar\mu_{\varepsilon,t},\mu)^2]\Big\} + (1+r_\varepsilon^{-\frac{1}{2}})r_\varepsilon\bigg)\\
&\le \sum_{i=1}^\infty \frac{8}{\lambda_i^2}.
\end{aligned}$$

(2) By (1.3), we have

$$\int_M |P_t f-\mu(f)|^2\mathrm{d}\mu \le \mathrm{e}^{-2\lambda_1 t}\int_M |f-\mu(f)|^2\mathrm{d}\mu,\quad t\ge 0, f\in L^2(\mu). \tag{2.8}$$

By (2.1)-(2.3), and noting that $L\phi_i = -\lambda_i\phi_i$ with $\{\phi_i\}_{i\ge1}$ being orthonormal in $L^2(\mu)$, we obtain

$$\mathbb{W}_2(\mu_{\varepsilon,t},\mu)^2 \le 4\mu(|\nabla(-L)^{-1}(f_{\varepsilon,t}-1)|^2) = 4\sum_{i=1}^\infty \lambda_i^{-1}\mathrm{e}^{-2\lambda_i\varepsilon}|\xi_i(t)|^2. \tag{2.9}$$

Below we prove the desired assertions respectively.

Since for $\nu\in\mathscr{P}_k$ we have $\mathbb{E}^\nu \le k\mathbb{E}^\mu$, it suffices to prove for $\nu=\mu$. Since μ is P_t-invariant and $\mu(\phi_i^2)=1$, we have

$$\mathbb{E}^\mu[\phi_i(X_{s_1})^2] = \mu(\phi_i^2) = 1. \tag{2.10}$$

Next, the Markov property yields

$$\mathbb{E}^\mu(\phi_i(X_{s_2})|X_{s_1}) = P_{s_2-s_1}\phi_i(X_{s_1}) = \mathrm{e}^{-\lambda_i(s_2-s_1)}\phi_i(X_{s_1}),\quad s_2>s_1.$$

Combining this with (2.10) and the definition of $\xi_i(t)$, we obtain

$$\begin{aligned}
\mathbb{E}^\mu|\xi_i(t)|^2 &= \frac{2}{t^2}\int_0^t \mathrm{d}s_1\int_{s_1}^t \mathbb{E}^\mu[\phi_i(X_{s_1})\phi_i(X_{s_2})]\mathrm{d}s_2\\
&= \frac{2}{t^2}\int_0^t \mathrm{d}s_1\int_{s_1}^t \mathbb{E}^\mu[\phi_i(X_{s_1})^2]\mathrm{e}^{-\lambda_i(s_2-s_1)}\mathrm{d}s_2 \le \frac{2}{t\lambda_i}.
\end{aligned}$$

Substituting into (2.9) gives

$$\mathbb{E}^\mu[\mathbb{W}_2(\mu_{\varepsilon,t},\mu)^2] \le \frac{8}{t}\sum_{i=1}^\infty \lambda_i^{-2}\mathrm{e}^{-2\lambda_i\varepsilon} = \frac{32}{t}\sum_{i=1}^\infty\int_\varepsilon^\infty \mathrm{d}s\int_s^\infty \mathrm{e}^{-2\lambda_i r}\mathrm{d}r. \tag{2.11}$$

Noting that (2.8) and the semigroup property imply

$$p_{2t}(x,x)-1 = \int_M |p_t(x,y)-1|^2\mu(\mathrm{d}y) = \int_M |P_{\frac{t}{2}}p_{\frac{t}{2}}(x,\cdot)(y)-1|^2\mu(\mathrm{d}y)$$

$$\leq \mathrm{e}^{-\lambda_1 t}\int_M |p_{\frac{t}{2}}(x,y)-1|^2\mu(\mathrm{d}y) = \mathrm{e}^{-\lambda_1 t}\{p_t(x,x)-1\},$$

we deduce from (2.1) that

$$\sum_{i=1}^{\infty}\mathrm{e}^{-2\lambda_i t} = \int_M \big\{p_{2t}(x,x)-1\big\}\mu(\mathrm{d}x) \leq \mathrm{e}^{-\lambda_1 t}\int_M \{p_t(x,x)-1\}\mu(\mathrm{d}x) \leq \mathrm{e}^{-\lambda_1 t}\gamma(t).$$

Therefore, by (2.11) and that $\gamma(t)$ is decreasing in t, we find a constant $c_1>0$ such that

$$\begin{aligned}\mathbb{E}^\mu[\mathbb{W}_2(\mu_{\varepsilon,t},\mu)^2] &\leq \frac{32}{t}\int_\varepsilon^\infty \mathrm{d}s\int_s^\infty \mathrm{e}^{-\lambda_1 t}\gamma(t)\mathrm{d}t\\ &\leq \frac{32}{t}\int_\varepsilon^1\bigg(\int_s^1\gamma(t)\mathrm{d}t+\gamma(1)\int_1^\infty \mathrm{e}^{-\lambda_1 t}\mathrm{d}t\bigg)\mathrm{d}s + \frac{32\gamma(1)}{t}\int_1^\infty \mathrm{d}s\int_s^\infty \mathrm{e}^{-\lambda_1 r}\mathrm{d}r\\ &\leq \frac{c_1}{t}\beta(\varepsilon),\quad \varepsilon\in(0,1].\end{aligned} \tag{2.12}$$

On the other hand, (2.3) and (2.9) imply that the measure

$$\pi(\mathrm{d}x,\mathrm{d}y) := \frac{1}{t}\int_0^t \big\{\delta_{X_s}(\mathrm{d}x)p_\varepsilon(X_s,y)\mu(\mathrm{d}y)\big\}\mathrm{d}s$$

is a coupling of μ_t and $\mu_{\varepsilon,t}$. Combining this with the fact that μ is P_t-invariant, we obtain

$$\mathbb{E}^\mu[\mathbb{W}_2(\mu_t,\mu_{\varepsilon,t})^2] \leq \frac{1}{t}\mathbb{E}^\mu\int_0^t \mathrm{d}s\int_M \rho(X_s,y)^2 p_\varepsilon(X_s,y)\mu(\mathrm{d}y) = \alpha(\varepsilon).$$

By (2.12) and the triangle inequality of $\mathbb{W}_2$, this yields

$$\mathbb{E}^\mu[\mathbb{W}_2(\mu_t,\mu)^2] \leq 2\inf_{\varepsilon\in(0,1]}\big\{\alpha(\varepsilon)+c_1t^{-1}\beta(\varepsilon)\big\}.$$

Therefore, (1.9) holds for some constant $c>0$ and $\nu=\mu$.

Finally, let P_t be ultracontractive. Then there exists a constant $c_1>0$ such that

$$\sup_{t\geq 1} p_t(x,y) \leq c_1,\quad x,y\in M. \tag{2.13}$$

So, the distribution of X_1 has a distribution $\nu_1\leq c_1\mu$. Let $\bar{\mu}_t = \frac{1}{t}\int_0^t \delta_{X_{1+s}}\mathrm{d}s$. It is easy to see that

$$\pi := \frac{1}{t}\int_0^1 \delta_{(X_s,X_{s+t})}\mathrm{d}s + \frac{1}{t}\int_1^t \delta_{(X_s,X_s)}\mathrm{d}s \in \mathscr{C}(\mu_t,\bar{\mu}_t), \tag{2.14}$$

so that (2.13) yields

$$\mathbb{E}^\nu[\mathbb{W}_2(\mu_t,\bar{\mu}_t)^2] \leq \frac{1}{t}\mathbb{E}^\nu\int_0^1 |X_s-X_{s+t}|^2\mathrm{d}s \leq \frac{c_1}{t}\mathbb{E}^\nu\int_0^1 \mu\big(\rho(X_s,\cdot)^2\big)\mathrm{d}s. \tag{2.15}$$

On the other hand, by the Markov property and (1.9), we find a constant $c_2>0$ such that

$$\mathbb{E}^\nu[\mathbb{W}_2(\bar{\mu}_t,\mu)^2] = \mathbb{E}^{\nu_1}[\mathbb{W}_2(\mu_t,\mu)^2] \leq c_2\inf_{\varepsilon\in(0,1]}\big\{\alpha(\varepsilon)+t^{-1}\beta(\varepsilon)\big\}.$$

Combining this with (2.15) and using the triangle inequality of $\mathbb{W}_2$, we prove (1.10) for some constant $c>0$. □

Proof of Corollary 1.2. (1) By [16, Lemma 3.5.6] and comparing P_t with the semigroup generated by $\Delta + \nabla V_1$, see for instance [6, (2.8)], (1.11) implies that the Harnack inequality

$$(2.16)\qquad (P_t f(x))^2 \le \{P_t f^2(y)\} e^{C+Ct^{-1}\rho(x,y)^2}, \quad x, y \in M, t \in (0,1]$$

holds for some constant $C > 0$. Therefore, by [15, Theorem 1.4.1] with $\Phi(r) = r^2$ and $\Psi(x,y) = C + Ct^{-1}\rho(x,y)^2$, we obtain

$$p_{2t}(x,x) = \sup_{\mu(f^2)\le 1} (P_t f(x))^2 \le \frac{1}{\int_M e^{-C-Ct^{-1}\rho(x,y)^2}\mu(dy)} \le \frac{e^{3C}}{\mu(B(x,\sqrt{2t}))}, \quad t \in (0,1], x \in M.$$

This implies

$$(2.17)\qquad \gamma(t) \le e^{3C}\tilde{\gamma}(t), \quad t \in (0,2].$$

On the other hand, by (1.11) and Itô's formula due to [7], there exists constant $C_1 > 0$ such that

$$d\rho(x,X_t)^2 \le \Big[C_1\big(1+\rho(x,X_t)^2\big) + |\nabla V(x)|^2\Big]dt + 2\sqrt{2}\rho(x,X_t)db_t,$$

where b_t is a one-dimensional Brownian motion. So, there exists a constant $C_2 > 0$ such that

$$(2.18)\qquad \mathbb{E}^\nu[\rho(x,X_t)^2] \le (C_1 + \nu(|\nabla V|^2))te^{C_1 t} \le C_2(1+\nu(|\nabla V|^2))t, \quad t \in [0,1], x \in M.$$

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

$$\begin{aligned}\sup_{\nu\in\mathscr{P}_k} \int_M \mathbb{E}^\nu \rho(x,X_\varepsilon)^2\mu(dx) &\le k\int_M \mathbb{E}^\mu \rho(x,X_\varepsilon)^2\mu(dx)\\ &\le C_2 k(1+\mu(|\nabla V|^2))\varepsilon \le ck\varepsilon, \quad \varepsilon \in (0,1], k \ge 1.\end{aligned}$$

Combining this with (2.17), we prove the first assertion by Theorem 1.1(2). The second assertion follows from (2.18) and Theorem 1.1(2), since P_t is ultracontractive provided $\|P_t e^{\lambda\rho_o^2}\|_\infty < \infty$ for $\lambda, t > 0$, see for instance [16, Theorem 3.5.5]. □

3 Proof of Theorem 1.3

(1) We first prove that for any $0 \ne f \in L^2(\mu)$,

$$(3.1)\qquad \lim_{t\to\infty} \frac{1}{t}\mathbb{E}^\mu\bigg[\bigg|\int_0^t f(X_s)ds\bigg|^2\bigg] = 4\int_0^\infty \mu\big((P_s f)^2\big)ds > 0.$$

As shown in [3, Lemma 2.8] that the Markov property and the symmetry of P_t in $L^2(\mu)$ imply

$$(3.2)\qquad \begin{aligned}\frac{1}{t}\mathbb{E}^\mu\bigg[\bigg|\int_0^t f(X_s)ds\bigg|^2\bigg] &= \frac{2}{t}\int_0^t ds_1\int_{s_1}^t \mathbb{E}^\mu[f(X_{s_1}P_{s_2-s_1}f(X_{s_1})]ds_2\\ &= \frac{2}{t}\int_0^t ds_1\int_{s_1}^t \mu\big((P_{\frac{s_2-s_1}{2}}f)^2\big)ds_2 = \frac{4}{t}\int_0^{t/2}\mu\big((P_s f)^2\big)ds\int_s^{t-s}dr\\ &= \frac{4}{t}\int_0^{t/2}(t-2s)\mu\big((P_s f)^2\big)ds, \quad t>0,\end{aligned}$$

where we have used the variable transform $(s,r)=(\frac{s_2-s_1}{2},\frac{s_1+s_2}{2})$. This implies (3.1). On the other hand, we take $0\neq f\in L^2(\mu)$ with $\mu(f)=0$ and $\|f\|_\infty\vee\|\nabla f\|_\infty\le 1$. Then

$$t\mathbb{E}^\mu[\tilde{W}_1(\mu_t,\mu)^2]\ge\frac{1}{t}\mathbb{E}^\mu\bigg[\bigg|\int_0^t f(X_s)\mathrm{d}s\bigg|^2\bigg].$$

Combining this with (3.1), we prove (1.14) for some constant $c>0$.

If (1.3) holds, then

$$\|P_tf-\mu(f)\|_{L^2(\mu)}\le \mathrm{e}^{-\lambda_1 t}\|f-\mu(f)\|_{L^2(\mu)},\quad t\ge 0, f\in L^2(\mu). \tag{3.3}$$

Let $\nu=h_\nu\mu\in\mathscr{P}$ with $h_\nu\in L^2(\mu)$. Similarly to (3.2), for any $f\in L^2(\mu)$ with $\mu(f)=0$, we have

$$\begin{aligned}&\frac{1}{t}\bigg\{\mathbb{E}^\nu\bigg[\bigg|\int_0^t f(X_s)\mathrm{d}s\bigg|^2\bigg]-\mathbb{E}^\mu\bigg[\bigg|\int_0^t f(X_s)\mathrm{d}s\bigg|^2\bigg]\bigg\}\\
&=\frac{1}{t}\int_M\{h_\nu(x)-1\}\mathbb{E}^x\bigg[\bigg|\int_0^t f(X_s)\mathrm{d}s\bigg|^2\bigg]\mu(\mathrm{d}x)\\
&=\frac{2}{t}\int_0^t\mathrm{d}s_1\int_{s_1}^t\mu\big(\{h_\nu-1\}P_{s_1}\{fP_{s_2-s_1}f\}\big)\mathrm{d}s_2\\
&=\frac{2}{t}\int_0^t\mathrm{d}s_1\int_{s_1}^t\mu\big(\{P_{s_1}(h_\nu-1)\}\cdot\{fP_{s_2-s_1}f\}\big)\mathrm{d}s_2\\
&\ge-\frac{2\|f\|_\infty}{t}\int_0^t\mathrm{d}s_1\int_{s_1}^t\|P_{s_1}(h_\nu-1)\|_{L^2(\mu)}\|P_{s_2-s_1}f\|_{L^2(\mu)}\mathrm{d}s_2.\end{aligned}$$

Taking $0\neq f\in L^2(\mu)$ with $\mu(f)=0$ and $\|f\|_\infty\vee\|\nabla f\|_\infty\le 1$, by combining this with (3.1) and (3.3), we derive

$$\begin{aligned}&\liminf_{t\to\infty}\Big[t\mathbb{E}^\nu[\tilde{\mathbb{W}}_1(\mu_t,\mu)^2]\Big\}\ge\liminf_{t\to\infty}\bigg\{\frac{1}{t}\mathbb{E}^\nu\bigg[\bigg|\int_0^t f(X_s)\mathrm{d}s\bigg|^2\bigg]\bigg\}\\
&\ge 4\int_0^\infty\mu\big(|P_sf|^2\big)\mathrm{d}s>0,\quad \nu=h_\nu\mu\text{ with }h_\nu\in L^2(\mu).\end{aligned}\tag{3.4}$$

Next, let $\bar{\mu}_t=\frac{1}{t}\int_1^{t+1}\delta_{X_s}\mathrm{d}s,\ t>0$. By (2.14) we have

$$\tilde{\mathbb{W}}_1(\bar{\mu}_t,\mu_t)\le\int_{M\times M}1_{\{x\neq y\}}\pi(\mathrm{d}x,\mathrm{d}y)=\frac{1}{t}. \tag{3.5}$$

Noting that for any $x\in M$ we have $\nu_x:=p_1(x,\cdot)\mu$ with $p_1(x,\cdot)\in L^2(\mu)$, by the Markov property and (3.4), we obtain

$$\liminf_{t\to\infty}\Big\{t\mathbb{E}^x[\tilde{\mathbb{W}}_1(\bar{\mu}_t,\mu)^2]\Big\}=\liminf_{t\to\infty}\Big[t\mathbb{E}^{\nu_x}[\tilde{\mathbb{W}}_1(\mu_t,\mu)^2]\Big\}>0.$$

Combining this with (3.5) and the triangle inequality leads to

$$\liminf_{t\to\infty}\Big\{t\mathbb{E}^x[\tilde{\mathbb{W}}_1(\mu_t,\mu)^2]\Big\}>0,\quad x\in M.$$

Therefore, by Fatou's lemma, for any $\nu \in \mathscr{P}$ we have

$$\liminf_{t\to\infty} \Big\{t\mathbb{E}^\nu[\tilde{\mathbb{W}}_1(\mu_t,\mu)^2]\Big\} = \liminf_{t\to\infty} \int_M \Big\{t\mathbb{E}^x[\tilde{\mathbb{W}}_1(\mu_t,\mu)^2]\Big\}\nu(\mathrm{d}x)$$
$$\geq \int_M \Big(\liminf_{t\to\infty} \Big\{t\mathbb{E}^x[\tilde{\mathbb{W}}_1(\mu_t,\mu)^2]\Big\}\Big)\nu(\mathrm{d}x) > 0,$$

which implies (1.15).

(2) Let $d \geq 3$, and let ∂M be empty or convex. By (1.16), we have $\mathrm{Ric} \geq -K$ for some constant $K > 0$. Then the Laplacian comparison theorem implies (see [4])

$$\Delta\rho(x,\cdot)(y) \leq \sqrt{K(d-1)}\coth\Big[\sqrt{K/(d-1)}\,\rho(x,y)\Big] \leq C\rho(x,y)^{-1}, \quad (x,y) \in \hat{M}$$

for some constant $C > 0$, where $\hat{M} := \{(x,y) : x,y \in M, x \neq y, x \notin \mathrm{cut}(y)\}$, and $\mathrm{cut}(y)$ is the cut-locus of y. So,

$$L\rho(x,\cdot)(y) \leq |\nabla V(y)| + C\big\{\rho(x,y) + \rho(x,y)^{-1}\big\}, \quad (x,y) \in \hat{M}.$$

Combining this with the Itô's formula due to [7], we obtain

$$\mathrm{d}\rho(X_0,X_t) \leq \sqrt{2}\mathrm{d}b_t + \big\{|\nabla V(X_t)| + C\rho(X_0,X_t) + C\rho(X_0,X_t)^{-1}\big\}\mathrm{d}t + \mathrm{d}l_t,$$

where b_t is a one-dimensional Brownian motion, and l_t is the local time of X_t at the initial value X_0, which is an increasing process supported on $\{t \geq 0 : X_t = X_0\}$. Thus, we find a constant $C_1 > 0$ such that

$$\mathrm{d}\Big\{\frac{\rho(X_0,X_t)^2}{1+\rho(X_0,X_t)^2}\Big\} \leq C_1(1+|\nabla V(X_t)|)\mathrm{d}t + \mathrm{d}M_t$$

for some martingale M_t. Since μ is P_t-invariant, this implies

$$\mathbb{E}^\mu\big\{\rho(X_0,X_t)\wedge 1\big\}^2 \leq C_2\big\{1+\mu(|\nabla V|)\big\}t, \quad t \geq 0, x \in M$$

for some constant $C_2 > 0$. Therefore, for any $N \in \mathbb{N}$ and $t_i := (i-1)t/N$, the probability measure

$$\tilde{\mu}_N := \frac{1}{N}\sum_{i=1}^N \delta_{X_{t_i}} = \frac{1}{t}\sum_{i=1}^N \int_{t_i}^{t_{i+1}} \delta_{X_{t_i}}\mathrm{d}s$$

satisfies

$$\mathbb{E}^\mu\tilde{W}_1(\tilde{\mu}_N,\mu_t)^2 \leq \frac{1}{t}\sum_{i=1}^N \int_{t_i}^{t_{i+1}} \mathbb{E}^\mu(\rho(X_{t_i},X_s)\wedge 1)^2\mathrm{d}s$$
$$\leq \frac{C_3}{t}\sum_{i=1}^N \int_{t_i}^{t_{i+1}} (s-t_i)\mathrm{d}s \leq \frac{C_3 t}{N}$$

for some constant $C_3 > 0$. So,

$$\sup_{\nu\in\mathscr{P}_k} \mathbb{E}^\nu[\tilde W_1(\tilde\mu_N,\mu_t)^2] \le k\mathbb{E}^\mu[\tilde W_1(\tilde\mu_N,\mu_t)^2] \le \frac{C_3kt}{N},\quad N,k\ge 1. \tag{3.6}$$

On the other hand, by Ric $\ge -K$ and $V\le K$ in (1.16) and using the volume comparison theorem, we find a constant $C_4>1$ such that

$$\mu(B(x,r)) \le C_4 r^d,\quad x\in M, r\in[0,1],$$

where $B(x,r) := \{y\in M : \rho(x,y)\wedge 1 \le r\}$. Since μ is a probability measure, this inequality holds for all $r>0$. Therefore, by [8, Proposition 4.2], there exists a constant $C_5>0$ such that

$$\tilde W_1(\tilde\mu_N,\mu) \ge C_5 N^{-\frac{1}{d}},\quad N\ge 1.$$

Combining this with (3.6) and using the triangle inequality for $\tilde W_1$, we obtain

$$\sup_{\nu\in\mathscr{P}_k} \mathbb{E}^\nu[\tilde W_1(\mu_t,\mu)] \ge C_5 N^{-\frac{1}{d}} - \sqrt{C_3kt}N^{-\frac{1}{2}},\quad N,k\ge 1.$$

maximizing in $N\ge 1$, we find a constant $c>0$ such that (1.17) holds.

Now, let $d\ge 4$. To prove (1.18) for general probability measure ν, we consider the shift empirical measure

$$\bar\mu_t := \frac{1}{t}\int_1^{t+1} \delta_{X_s}\mathrm{d}s,\quad t\ge 1,$$

and the probability measures

$$\nu_x := \delta_x P_1 = p_1(x,\cdot)\mu,\quad \nu_{x,1} := \frac{1_{B(x,1)}}{\nu_x(B(x,1))}\nu_x,\quad x\in M.$$

By the Markov property, we obtain

$$\begin{aligned}\mathbb{E}^x[\tilde W_1(\bar\mu_t,\mu]) &= \mathbb{E}^{\nu_x}[\tilde W_1(\mu_t,\mu)] = \int_M \mathbb{E}^y[\tilde W_1(\mu_t,\mu)]p_1(x,y)\mu(\mathrm{d}y)\\ &\ge \int_{B(x,1)} \mathbb{E}^y[\tilde W_1(\mu_t,\mu)]p_1(x,y)\mu(\mathrm{d}y) = \nu_x(B(x,1))\mathbb{E}^{\nu_{x,1}}[\tilde W_1(\bar\mu_t,\mu)].\end{aligned}$$

Noting that $h(x) := \sup_{y\in B(x,1)} p_1(x,y) < \infty$, this and (1.17) yield

$$\mathbb{E}^x[\tilde W_1(\bar\mu_t,\mu)] \ge g(x)t^{-\frac{1}{d-2}},\quad g(x) := c\nu_x(B(x,1))h(x)^{-\frac{1}{d-2}}, x\in M, t\ge 1.$$

Consequently, for any probability measure ν,

$$\mathbb{E}^\nu[\tilde W_1(\bar\mu_t,\mu)] = \int_M \mathbb{E}^x[\tilde W_1(\bar\mu_t,\mu)]\nu(\mathrm{d}x) \ge \nu(g)t^{-\frac{1}{d-2}},\quad t\ge 1.$$

Combining this with (3.5) and noting that $d\ge 4$ implies $t^{-\frac{1}{d-2}} \ge t^{-\frac{1}{2}}$ for $t\ge 1$, we find a constant $c_\nu>0$ such that when t is large enough,

$$\mathbb{E}^\nu[\tilde W_1(\mu_t,\mu)] \ge \mathbb{E}^\nu\big[\tilde W_1(\bar\mu_t,\mu) - \tilde{\mathbb{W}}_1(\bar\mu_t,\mu_t)\big] \ge c(\nu)t^{-\frac{1}{d-2}}.$$

(3) According to [19, Theorem 2.1], for any $\varepsilon \in (0,1]$ we have

$$\liminf_{t\to\infty} \Big\{ t \inf_{x\in M} \mathbb{E}^x[\mathbb{W}_2(\mu_{\varepsilon,t},\mu)^2]\Big\} \ge \sum_{i=1}^\infty \frac{2}{\lambda_i^2 \mathrm{e}^{2\varepsilon\lambda_i}}. \tag{3.7}$$

On the other hand, by [16, Theorem 3.3.2], the conditions that $\mathrm{Ric} - \mathrm{Hess}_V \ge K$ and ∂M is empty or convex imply

$$\mathbb{W}_2(\mu_{\varepsilon,t},\mu)^2 \le \mathrm{e}^{-2\varepsilon K}\mathbb{W}_2(\mu_t,\mu)^2, \quad \varepsilon \ge 0.$$

Combining this with (3.7), we derive

$$\liminf_{t\to\infty} \Big\{ t \inf_{x\in M} \mathbb{E}^x[\mathbb{W}_2(\mu_t,\mu)^2]\Big\} \ge \mathrm{e}^{2\varepsilon K} \sum_{i=1}^\infty \frac{2}{\lambda_i^2 \mathrm{e}^{2\varepsilon\lambda_i}}, \quad \varepsilon \in (0,1].$$

By letting $\varepsilon \downarrow 0$ we finish the proof.

4 Proof of Example 1.4

(1) Taking $V_1 \in C^\infty(\mathbb{R}^d)$ such that $V_1(x) = -\kappa|x|^\alpha$ for $|x| \ge 1$, and writing $V_2 = V + W - V_1$, we see that (1.11) holds for some constant $K \in \mathbb{R}$. By Corollary 1.2, it suffices to estimate $\tilde\gamma(t)$. For any $x \in \mathbb{R}^d$ with $|x| \ge 1$, and any $t \in (0,1]$, let $x_t = \frac{x}{|x|}\big(|x| - \frac{1}{2}\sqrt{t}\big)$. We find a constant $c_1 > 0$ and some point $z \in B(x,\sqrt{t})$ such that

$$\mu\big(B(x,\sqrt{t})\big) \ge \int_{B(x_t,\frac{1}{4}\sqrt{t})} \mathrm{e}^{-\kappa|y|^\alpha + W(y)}\mathrm{d}y \ge c_1 t^{\frac{d}{2}} \mathrm{e}^{-\kappa(|x| - \frac{1}{4}t^{\frac{1}{2}})^\alpha + W(z)}. \tag{4.1}$$

Since $|x| \ge 1$, $t \in (0,1]$ and $\alpha > 1$, we find a constant $c_2 > 0$ such that

$$\begin{aligned} |x|^\alpha - \big(|x| - t^{\frac{1}{2}}/4\big)^\alpha &= \alpha \int_{|x|-\frac{1}{4}t^{\frac{1}{2}}}^{|x|} r^{\alpha-1}\mathrm{d}r \\ &\ge \frac{\alpha t^{\frac{1}{2}}}{4}\Big(\frac{|x|}{2}\Big)^{\alpha-1} \ge c_2|x|^{\alpha-1}t^{\frac{1}{2}}. \end{aligned} \tag{4.2}$$

Moreover,

$$|W(z) - W(x)| \le \|\nabla W\|_\infty |x - z| \le \|\nabla W\|_\infty, \quad t \in (0,1], z \in B(x, t^{\frac{1}{2}}).$$

Combining this with (4.1) and (4.2), we find a $c_3 > 0$ such that

$$\mu\big(B(x,\sqrt{t})\big) \ge c_3 t^{\frac{d}{2}} \mathrm{e}^{-\kappa|x|^\alpha + c_2|x|^{\alpha-1}t^{\frac{1}{2}} + W(x)}, \quad t \in [0,1], x \in \mathbb{R}^d.$$

Noting that $-\kappa|x|^\alpha + 2|W(x)|$ is bounded from above, we find constants $c_4, c_5 > 0$ such that

$$\int_{|x|\ge 1} \frac{\mu(\mathrm{d}x)}{\mu(B(x,\sqrt{t}))} \le c_4 t^{-\frac{d}{2}} \int_1^\infty r^{d-1}\mathrm{e}^{-c_2 r^{\alpha-1}t^{\frac{1}{2}}}\mathrm{d}r \le c_5 t^{-\frac{d}{2} - \frac{d}{2(\alpha-1)}} = c_5 t^{-\frac{\alpha d}{2(\alpha-1)}}, \quad t \in (0,1].$$

On the other hand, there exists a constant $c_6 > 0$ such that $\mu(B(x,r)) \geq c_6 r^d$ for $|x| < 1$ and $r \in (0,1]$. In conclusion, there exists a constant $c_7 > 0$ such that

$$\tilde{\gamma}(t) := \int_{\mathbb{R}^d} \frac{\mu(\mathrm{d}x)}{\mu(B(x,\sqrt{t}))} \leq c_5 t^{-\frac{\alpha d}{2(\alpha-1)}} + c_6^{-1} t^{-\frac{d}{2}} \leq c_7 t^{-\frac{\alpha d}{2(\alpha-1)}}, \quad t \in (0,1].$$

Thus, there exists a constant $c_8 > 0$ such that for any $\varepsilon \in (0,1]$,

$$\tilde{\beta}(\varepsilon) \leq 1 + c_6 \int_\varepsilon^1 \mathrm{d}s \int_s^1 t^{-\frac{d\alpha}{2(\alpha-1)}} \mathrm{d}t \leq \begin{cases} c_8 \varepsilon^{2-\frac{d\alpha}{2(\alpha-1)}}, & \text{if } 2 < \frac{d\alpha}{2(\alpha-1)}, \\ c_8 \log(1+\varepsilon^{-1}), & \text{if } 2 = \frac{d\alpha}{2(\alpha-1)}, \\ c_8, & \text{if } 2 > \frac{d\alpha}{2(\alpha-1)}. \end{cases}$$

By taking $\varepsilon = t^{-\frac{2(\alpha-1)}{(d-2)\alpha+2}}$ if $4(\alpha-1) < d\alpha$, $\varepsilon = t^{-1}$ if $4(\alpha-1) = d\alpha$, and $\varepsilon \downarrow 0$ if $4(\alpha-1) > d\alpha$, we derive

$$\inf_{\varepsilon \in (0,1]} \left\{ \varepsilon + t^{-1} \tilde{\beta}(\varepsilon) \right\} \leq \begin{cases} c t^{-\frac{2(\alpha-1)}{(d-2)\alpha+2}}, & \text{if } 4(\alpha-1) < d\alpha, \\ c t^{-1} \log(1+t), & \text{if } 4(\alpha-1) = d\alpha, \\ c t^{-1}, & \text{if } 4(\alpha-1) > d\alpha \end{cases} \tag{4.3}$$

for some constant $c > 0$. Therefore, (1.20) follows from Corollary 1.2(1).

(2) Next, by [10, Corollary 3.3], when $\alpha > 2$ the Markov semigroup P_t^0 generated by $\Delta - \kappa \nabla |\cdot|^\alpha$ is ultracontractive with

$$\|P_t^0\|_{L^1(\mu_0) \to L^\infty(\mu_0)} \leq \mathrm{e}^{c_1(1+t^{-\alpha/(\alpha-2)})}, \quad t > 0 \tag{4.4}$$

for some constant $c_1 > 0$, where $\mu_0(\mathrm{d}x) := Z^{-1} \mathrm{e}^{-\kappa |x|^\alpha} \mathrm{d}x$ is probability measure with normalized constant $Z > 0$. According to the correspondence between the ultracontractivity and the log-Sobolev inequality, see [5], (4.4) holds if and only if there exists a constant $c_2 > 0$ such that

$$\mu_0(f^2 \log f^2) \leq r \mu_0(|\nabla f|^2) + c_2(1 + r^{-\frac{\alpha}{\alpha-2}}), \quad r > 0, \mu_0(f^2) = 1.$$

Replacing f by $f \mathrm{e}^{\frac{W}{2}}$ and using $\|\nabla W\|_\infty < \infty$ which implies $\mu(\mathrm{e}^{cW}) < \infty$ for any $c > 0$ due to $\alpha > 1$, we find constants c_3 such that

$$\begin{aligned} \mu(f^2 \log f^2) &\leq \mu(f^2 W) + 2r\mu(|\nabla f|^2) + 2\|\nabla W\|_\infty^2 + c_2(1 + r^{-\frac{\alpha}{\alpha-2}}) \\ &\leq \frac{1}{2}\mu(f^2 \log f^2) + \frac{1}{2} \log \mu(\mathrm{e}^{2W}) + 2r\mu(|\nabla f|^2) + 2\|\nabla W\|_\infty^2 + c_2(1 + r^{-\frac{\alpha}{\alpha-2}}) \\ &\leq \frac{1}{2}\mu(f^2 \log f^2) + 2r\mu(|\nabla f|^2) + c_3(1 + r^{-\frac{\alpha}{\alpha-2}}), \quad r > 0, \mu(f^2) = 1, \end{aligned}$$

where in the second line we have used the Young inequality [1, Lemma 2.4]

$$\mu(f^2 g) \leq \mu(f^2 \log f^2) + \log \mu(\mathrm{e}^g), \quad \mu(f^2) = 1, g \in L^1(f^2\mu).$$

Hence, for some constant $c_4 > 0$ we have

$$\mu(f^2 \log f^2) \leq r\mu(|\nabla f|^2) + c_4(1 + r^{-\frac{\alpha}{\alpha-2}}), \quad r > 0, \mu(f^2) = 1.$$

By the above mentioned correspondence of the log-Sobolev inequality and semigroup estimate, this implies

$$\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} \le \mathrm{e}^{c_5(1+t^{-\alpha/(\alpha-2)})},\quad t>0$$

for some constant $c_5>0$. In particular, this and $\mu(\mathrm{e}^{\lambda|\cdot|^2})<\infty$ imply $\|P_t\mathrm{e}^{\lambda|\cdot|^2}\|_\infty<\infty$ for $t,\lambda>0$, so that by Corollary 1.2(2), (1.21) follows from (4.3) and the fact that $|\nabla V(x)|^2\le c'(1+|x|^{2(\alpha-1)})$ holds for some constant $c'>0$.

(3) By [11, Corollary 1.4], the Poincaré inequality (1.3) holds for some constant $\lambda_1>0$. Moreover, it is trivial that the condition (1.16) holds for some constant $K\ge 0$. So, the desired lower bound estimate is implied by Theorem 1.3.

Acknowledgement. The author would like to thank the referees for useful comments and careful corrections.

References

[1] M. Arnaudon, A. Thalmaier, F.-Y. Wang, *Gradient estimates and Harnack inequalities on non-compact Riemannian manifolds,* Stoch. Proc. Appl. 119(2009), 3653–3670.

[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] P. Cattiaux, D. Chafai, A. Guillin, *Central limit theorems for additive functionals of ergodic Markov diffusions processes,* ALEA Lat. Am. J. Probab. Math. Stat. 9(2012), 337–382.

[4] J. Cheeger and D. G. Ebin, *Comparison Theorems in Riemannian Geometry,* Amsterdam: North-Holland, 1975.

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

[6] F. Z. Gong, F.-Y. Wang, *Heat kernel estimates with applications to compactness of manifolds,* Quart J. Math. 52(2001), 1–10.

[7] W. S. Kendall, *The radial part of Brownian motion on a manifold: a semimartingale property,* Ann. Probab. 15(1987), 1491–1500.

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

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

[10] M. Röckner, F.-Y. Wang, *Supercontractivity and ultracontractivity for (non-symmetric) diffusion semigroups on manifolds,* Forum Math. 15(2003), 893–921.

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

[12] F.-Y. Wang, *Functional inequalities, semigroup properties and spectrum estimates,* Infinite Dimensional Analysis, Quantum Probability and Related Topics 3:2(2000), 263–295.

[13] F.-Y. Wang, *Harnack inequalities on manifolds with boundary and applications,* J. Math. Pures Appl. 94(2010), 304–321.

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

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

[16] F.-Y. Wang, *Analysis for Diffusion Processes on Riemnnian Manifolds,* World Sicentific, 2014.

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

[18] F.-Y. Wang, *Convergence in Wasserstein distance for empirical measures of Dirichlet diffusion processes on manifolds,* arXiv:2005.09290, to appear in J. Eur. Math. Soc.

[19] F.-Y. Wang, J.-X. Zhu, *Limit theorems in Wasserstein distance for empirical measures of diffusion processes on Riemannian manifolds,* aXiv:1906.03422.

A Upper bound estimate on $\mathbb{W}_p(f_1\mu, f_2\mu)$

For $p \geq 1$, let $\mathbb{W}_p$ be the L^p-Wasserstein distance induced by ρ, i.e.

$$\mathbb{W}_p(\mu_1, \mu_2) = \inf_{\pi \in \mathscr{C}(\mu_1, \mu_2)} \|\rho\|_{L^p(\pi)}.$$

According to [9, Theorem 2], for any probability density f of μ, we have

$$\mathbb{W}_p(f\mu, \mu)^p \leq p^p \mu\big(|\nabla(-L)^{-1}(f-1)|^p\big). \tag{A.1}$$

The idea of the proof goes back to [2], in which the following estimate is presented for probability density functions f_1, f_2:

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

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

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

In this Appendix, we extend estimates (A.1) and (A.2) as follows, which might be useful for further studies.

Theorem A.1. *For any probability density functions f_1 and f_2 with respect to μ such that $f_1\vee f_2>0$,*

$$\mathbb{W}_p(f_1\mu,f_2\mu)^p \leq \min\left\{p^p2^{p-1}\int_M\frac{|\nabla(-L)^{-1}(f_2-f_1)|^p}{(f_1+f_2)^{p-1}}\mathrm{d}\mu,\ p^p\int_M\frac{|\nabla(-L)^{-1}(f_2-f_1)|^p}{f_1^{p-1}}\mathrm{d}\mu,\right.$$
$$\left.\int_M\frac{|\nabla(-L)^{-1}(f_2-f_1)|^2}{\mathscr{M}_p(f_1,f_2)}\mathrm{d}\mu\right\}.$$

Proof. It suffices to prove for $p>1$. Let $\mathrm{Lip}_b(M)$ be the set of bounded Lipschitz continuous functions on M. Consider the Hamilton-Jacobi semigroup $(Q_t)_{t>0}$ on $\mathrm{Lip}_b(M)$:

$$Q_t\phi := \inf_{x\in M}\left\{\phi(x)+\frac{1}{pt^{p-1}}\rho(x,\cdot)^p\right\},\quad t>0,\phi\in\mathrm{Lip}_b(M).$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_t\phi = -\frac{p-1}{p}|\nabla Q_t\phi|^{\frac{p}{p-1}},\quad t>0. \tag{A.3}$$

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 (A.3) and the integration by parts formula, we obtain

$$\begin{aligned}
\mu_1(Q_1f)-\mu_2(f) &= \int_0^1\left\{\frac{\mathrm{d}}{\mathrm{d}s}\mu\big([f_1+\theta_s(f_2-f_1)]Q_sf\big)\right\}\mathrm{d}s\\
&= \int_0^1\mathrm{d}s\int_M\left\{\theta_s'(f_2-f_1)Q_sf-\frac{f_1+\theta_s(f_2-f_1)}{q}|\nabla Q_sf|^q\right\}\mathrm{d}\mu\\
&= \int_0^1\mathrm{d}s\int_M\left\{\theta_s'\langle\nabla(-L)^{-1}(f_2-f_1),\nabla Q_sf\rangle-\frac{f_1+\theta_s(f_2-f_1)}{q}|\nabla Q_sf|^q\right\}\mathrm{d}\mu\\
&\leq \frac{1}{p}\int_M|\nabla(-L)^{-1}(f_2-f_1)|^p\mathrm{d}\mu\int_0^1\frac{|\theta_s'|^p}{[f_1+\theta_s(f_2-f_1)]^{p-1}}\mathrm{d}s,
\end{aligned}$$

where the last step is due to Young's inequality $ab\leq a^p/p+b^q/q$ for $a,b\geq 0$. By Kantorovich duality formula

$$\frac{1}{p}\mathbb{W}_p(\mu_1,\mu_2)^p = \sup_{f\in C_b^1(M)}\big\{\mu_1(Q_1f)-\mu_2(f)\big\},$$

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 derive

$$\mathbb{W}_p(\mu_1,\mu_2)^p \leq \int_0^1 \frac{|\theta_s'|^p}{\min\{\theta_s,1-\theta_s\}^{p-1}}\mathrm{d}s\int_M \frac{|\nabla(-L)^{-1}(f_1-f_2)|^p}{(f_1+f_2)^{p-1}}\mathrm{d}\mu. \tag{A.4}$$

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 (A.4) that

$$\mathbb{W}_p(f_1\mu,f_2\mu)^p \leq p^p2^{p-1}\int_M \frac{|(-L)^{-\frac{1}{2}}(f_2-f_1)|^p}{(f_1+f_2)^{p-1}}\mathrm{d}\mu.$$

Next, (A.4) with $\theta_s = 1-(1-s)^p$ implies

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

Finally, with $\theta_s = s$ we deduce from (A.4) that

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

Then the proof is finished. □