Exponential Contraction in Wasserstein Distances for Diffusion Semigroups with Negative Curvature*

Feng-Yu Wang

Laboratory of Mathematical and Complex Systems, Beijing Normal University, Beijing 100875, China

Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, United Kingdom

wangfy@bnu.edu.cn, F.-Y.Wang@swansea.ac.uk

September 10, 2019

Abstract

Let P_t be the (Neumann) diffusion semigroup P_t generated by a weighted Laplacian on a complete connected Riemannian manifold M without boundary or with a convex boundary. It is well known that the Bakry-Emery curvature is bounded below by a positive constant $\lambda > 0$ if and only if

$$W_p(\mu_1 P_t, \mu_2 P_t) \le \mathrm{e}^{-\lambda t} W_p(\mu_1, \mu_2), \quad t \ge 0, p \ge 1$$

holds for all probability measures μ_1 and μ_2 on M, where W_p is the L^p Wasserstein distance induced by the Riemannian distance. In this paper, we prove the exponential contraction

$$W_p(\mu_1 P_t, \mu_2 P_t) \le c\mathrm{e}^{-\lambda t} W_p(\mu_1, \mu_2), \quad p \ge 1, t \ge 0$$

for some constants $c, \lambda > 0$ for a class of diffusion semigroups with negative curvature where the constant c is essentially larger than 1. Similar results are derived for SDEs with multiplicative noise by using explicit conditions on the coefficients, which are new even for SDEs with additive noise.

AMS subject Classification: 60J75, 47G20, 60G52.
Keywords: Wasserstein distance, diffusion semigroup, Riemannian manifold, curvature condition, SDEs with multiplicative noise.

---

*Supported in part by NNSFC (11131003, 11431014).

1 Introduction

Let M be a d-dimensional connected complete Riemannian manifold possibly with a convex boundary ∂M. Let ρ be the Riemannian distance. Consider $L = \Delta + Z$ for the Laplace-Beltrami operator Δ and some C^1-vector field Z such that the (reflecting) diffusion process generated by L is non-explosive. Then the associated Markov semigroup P_t is the (Neumann if $\partial M \neq \emptyset$) semigroup generated by L on M. In particular, it is the case when the curvature of L is bounded below; that is,

$$\text{(1.1)} \qquad \mathrm{Ric}_Z := \mathrm{Ric} - \nabla Z \geq K$$

holds for some constant $K \in \mathbb{R}$. Here and throughout the paper, we write $\mathscr{T} \geq h$ for a (not necessarily symmetric) 2-tensor $\mathscr{T}$ and a function h provided

$$\mathscr{T}(X, X) \geq h(x)|X|^2, \quad X \in T_x M, x \in M.$$

There exist many inequalities on P_t which are equivalent to the curvature condition (1.1), see [6, 20, 23, 42] for details. In particular, for any constant $K \in \mathbb{R}$, the Wasserstein distance inequality

$$\text{(1.2)} \qquad W_p(\mu_1 P_t, \mu_2 P_t) \leq \mathrm{e}^{-Kt} W_p(\mu_1, \mu_2), \quad t \geq 0, p \geq 1, \mu_1, \mu_2 \in \mathscr{P}(M)$$

is equivalent to the curvature condition (1.1). Here, $\mathscr{P}(M)$ is the class of all probability measures on M; W_p is the L^p-Warsserstein distance induced by ρ, i.e.,

$$W_p(\mu_1, \mu_2) := \inf_{\pi \in \mathscr{C}(\mu_1, \mu_2)} \|\rho\|_{L^p(\pi)}, \quad \mu_1, \mu_2 \in \mathscr{P}(M),$$

where $\mathscr{C}(\mu_1, \mu_2)$ is the class of all couplings of μ_1 and μ_2; and for a Markov operator P on $\mathscr{B}_b(M)$ (i.e. P is a positivity-preserving linear operator with $P1 = 1$),

$$(\nu P)(A) := \nu(P 1_A), \quad A \in \mathscr{B}(M), \nu \in \mathscr{P}(M),$$

where $\nu(f) := \int_M f \mathrm{d}\nu$ for $f \in L^1(\nu)$. In some references, νP is also denoted by $P^* \nu$. In the sequel we will use P_t^* to stand for the adjoint operator of P_t in $L^2(\mu)$ for the invariant probability measure μ, hence adopt the notation νP rather than $P^* \nu$ to avoid confusion. When the curvature is positive (i.e. $K > 0$), (1.2) implies the W_p-exponential contraction of P_t for $p \geq 1$.

In this paper, we aim to consider the case when (1.1) only holds for some negative constant K, and to prove the exponential contraction

$$\text{(1.3)} \qquad W_p(\mu_1 P_t, \mu_2 P_t) \leq c\mathrm{e}^{-\lambda t} W_p(\mu_1, \mu_2), \quad t \geq 0, p \geq 1, \mu_1, \mu_2 \in \mathscr{P}(M)$$

for some constants $c, \lambda > 0$. It is crucial that the exponential rate λ is independent of p. Due to the equivalence of (1.1) and (1.2), in the negative curvature case it is essential that $c > 1$.

According to [37], even when Ric_Z is unbounded below, i.e. Ric_Z goes to $-\infty$ when $\rho_o := \rho(o,\cdot) \to \infty$ for a fixed $o \in M$, there may exist the log-Sobolev inequality which implies the exponentially convergence of P_t in entropy. This suggests that (1.3) may also hold for a class of diffusion semigroups with negative curvature.

Recently, some efforts have been made in this direction for $M = \mathbb{R}^d$, see [11, 12, 18]. More precisely, let P_t be the diffusion semigroup for the solution to the following SDE on $\mathbb{R}^d$:

$$\mathrm{d}X_t = \sqrt{2}\,\mathrm{d}B_t + b(X_t)\mathrm{d}t,$$

where B_t is the d-dimensional Brownian motion and $b : \mathbb{R}^d \to \mathbb{R}^d$ is continuous. If there exist constants $K_1, K_2, r_0 > 0$ such that

$$\langle b(x) - b(y), x - y\rangle \le 1_{|x-y|\le r_0}(K_1 + K_2)|x - y|^2 - K_2|x - y|^2, \quad x, y \in \mathbb{R}^d, \tag{1.4}$$

then due to [11, 12] we have

$$W_1(\delta_x P_t, \delta_y P_t) \le c\mathrm{e}^{-\lambda t}|x - y|, \quad x, y \in \mathbb{R}^d, t \ge 0 \tag{1.5}$$

for some constants $c, \lambda > 0$, where δ_x is the Dirac measure at point x. Indeed, [11, 12] proved the W_1-exponential contraction with respect to a modified distance $f(|x - y|)$ in place of $|x - y|$ as constructed in [8, 9] for estimates of the spectral gap using the coupling by reflection. Under condition (1.4) the modified distance is comparable with the usual one so that (1.5) follows. As mentioned in [12] that there is essential difficulty to prove (1.3) for $p > 1$ even for this flat case.

In Luo and Wang [18] the estimate (1.5) was extended as

$$W_p(\delta_x P_t, \delta_y P_t) \le c\mathrm{e}^{-\lambda t/p}(|x - y| + |x - y|^{\frac{1}{p}}), \quad x, y \in \mathbb{R}^d, t \ge 0, p \ge 1 \tag{1.6}$$

for some constants $c, \lambda > 0$. Comparing with (1.3) which is equivalent to

$$W_p(\delta_x P_t, \delta_y P_t) \le c\mathrm{e}^{-\lambda t}|x - y|, \quad p \ge 1, x, y \in \mathbb{R}^d, t \ge 0$$

according to [17] (see Proposition 3.1 below), (1.6) is less sharp for small $|x-y|$ and/or large p. It is open whether (1.4), or in the Riemannian setting that Ric_Z is uniformly positive outside a compact domain, implies (1.3) for some constants $c, \lambda > 0$.

As in [16, 17], we will consider the Warsserstein distances induced by Young functions in the class

$$\mathscr{N} := \Big\{\Phi \in C^1([0,\infty);[0,\infty)) : \Phi' \text{ is nonnegative and increasing}, \\ \Phi(0) = 0, \Phi(r) > 0 \text{ for } r > 0, \lim_{r\to\infty} \frac{\Phi(r)}{r} = \infty\Big\}.$$

For any $\Phi \in \mathscr{N}$ and a measure ν on M, consider the gauge norm in $L^\Phi(\nu)$:

$$\|f\|_{L^\Phi(\nu)} := \inf\Big\{r > 0 : \nu\big(\Phi(r^{-1}|f|)\big) \le 1\Big\}, \quad \inf\emptyset := \infty.$$

In particular, we have $\|f\|_{L^{\Phi_p}(\nu)} = \|f\|_{L^p(\nu)}$ for $\Phi_p(r) := r^p,\ p \in (1,\infty)$. This is the reason why we do not take $\Phi_p(r) = \frac{1}{p}r^p$ in the characterization of Legendre conjugates. We extend the notion Φ_p to $p = 1, \infty$ by letting $\Phi_1(r) = r, \Phi_\infty = \lim_{p\to\infty} \Phi_p$ and $\|f\|_{L^{\Phi_p}(\nu)} = \|f\|_{L^p(\nu)}$ for all $p \in [1,\infty]$. Now, let

$$W_\Phi(\mu_1,\mu_2) = \inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)} \|\rho\|_{L^\Phi(\pi)}, \quad \Phi \in \bar{\mathscr{N}} := \mathscr{N} \cup \{\Phi_1, \Phi_\infty\}.$$

In particular, $W_{\Phi_p} = W_p$ for $p \in [1,\infty]$. We aim to prove the exponential decay

$$W_\Phi(\delta_x P_t, \delta_y P_t) \le \frac{c}{\Phi^{-1}(1)} \mathrm{e}^{-\lambda t}\rho(x,y), \quad x,y \in M, t \ge 0, \Phi \in \bar{\mathscr{N}} \tag{1.7}$$

when (1.1) only holds for a negative constant K, where Φ^{-1} is the inverse of $\Phi(\neq \Phi_\infty)$ and we set $\Phi_\infty^{-1}(1) = 1$ by convention.

To extend condition (1.4) to the Riemannian setting, consider the index

$$I(x,y) = \int_0^{\rho(x,y)} \sum_{i=1}^{d-1} \Big\{ |\nabla_{\dot\gamma} J_i|^2 - \langle \mathscr{R}(\dot\gamma, J_i)\dot\gamma, J_i\rangle \Big\}(\gamma_s)\mathrm{d}s, \quad x,y \in M,$$

where ρ is the Riemannian distance, $\mathscr{R}$ is the curvature tensor; $\gamma : [0,\rho(x,y)] \to M$ is the minimal geodesic from x to y with unit speed; $\{J_i\}_{i=1}^{d-1}$ are Jacobi fields along γ such that

$$J_i(y) = P_{x,y}\, J_i(x), \quad i = 1,\dots,d-1$$

holds for the parallel transform $P_{x,y} : T_xM \to T_yM$ along the geodesic γ, and $\{\dot\gamma(s), J_i(s) : 1 \le i \le d-1\}$ $(s = 0, \rho(x,y))$ is an orthonormal basis of the tangent space (at points x and y, respectively).

Note that when $(x,y) \in \mathrm{Cut}(M)$, i.e. x is in the cut-locus of y, the minimal geodesic may be not unique. As a convention in the literature, all conditions on the index I are given outside $\mathrm{Cut}(M)$. We now extend condition (1.4) to the non-flat case as follows: for some constants $K_1, K_2 > 0$,

$$\begin{aligned} I_Z(x,y) &:= I(x,y) + \langle Z, \nabla\rho(\cdot,y)\rangle(x) + \langle Z, \nabla\rho(x,\cdot,)\rangle(y) \\ &\le \big\{(K_1+K_2)1_{\{\rho(x,y)\le r_0\}} - K_2\big\}\rho(x,y), \quad x,y \in M. \end{aligned} \tag{1.8}$$

In the flat case we have $I(x,y) = 0$ and $\rho(x,y) = |x-y|$, so that this condition reduces back to (1.4). Moreover, the curvature condition (1.1) is equivalent to

$$I_Z(x,y) \le -K\rho(x,y), \quad x,y \in M,$$

so that (1.8) implies $\mathrm{Ric}_Z \ge -(K_1+K_2)$.

In the next section, we state our main results and present examples. With condition (1.8) we first extend the main results of [11, 18] to the present Riemannian setting and give the exponential convergence of P_t in W_2. Under the ultracontractivity and condition (1.1) for some $K < 0$, our the second result ensures the desired inequality (1.7). Finally, we extend these results to SDEs with multiplicative noise by using explicit conditions on the coefficients. To prove these results, we make some preparations in Section 3. Complete proofs of the main results are addressed in Sections 4-6 respectively.

2 Main Results and examples

We first consider the Riemannian setting, then extend to SDEs with multiplicative noise by using explicit conditions on the coefficients instead of the less explicit curvature condition.

2.1 The Riemannian setting

We start with condition (1.8). Besides the extension of (1.6), this condition also implies the hypercontractivity and the exponential convergence in W_2 for the semigroup P_t. For a measure μ and constants $p, q \geq 1$, let $\|\cdot\|_{L^p(\mu)\to L^q(\mu)}$ stand for the operator norm form $L^p(\mu)$ to $L^q(\mu)$. Recall that P_t is called hypercontractive if it has a unique invariant probability measure μ and $\|P_t\|_{L^2(\mu)\to L^4(\mu)} = 1$ holds for large $t > 0$. By interpolation theorem, $\|P_t\|_{L^2(\mu)\to L^4(\mu)} = 1$ can be replaced by $\|P_t\|_{L^p(\mu)\to L^q(\mu)} = 1$ for some $\infty > q > p > 1$. Let $\rho_o = (o, \cdot)$ for a fixed point $o \in M$.

Theorem 2.1. *Let* (1.8) *hold for some constants* K_1, K_2 *and* $r_0 > 0$. *Then:*

(1) *There exist two constants* $c, \lambda > 0$ *such that for any* $\Phi \in \bar{\mathscr{N}}$ *and* $x, y \in M$,

$$W_\Phi(\delta_x P_t, \delta_y P_t) \leq \inf\Big\{r > 0 : \sup_{s\in(0,1+r_0+\rho(x,y)]} \frac{\Phi(r^{-1}s)}{s} \leq \frac{e^{\lambda t}}{c\rho(x,y)}\Big\}, \quad t \geq 0. \tag{2.1}$$

In particular, there exist constants $c, \lambda > 0$ *such that*

$$W_p(\delta_x P_t, \delta_y P_t) \leq \{ce^{-\lambda t}\}^{\frac{1}{p}}(\rho(x,y) + \rho(x,y)^{\frac{1}{p}}), \quad p \geq 1, t \geq 0, x, y \in M.$$

(2) *If* P_t *has an invariant probability measure* μ *with* $\mu(e^{\varepsilon\rho_0^2}) < \infty$ *for some constant* $\varepsilon > 0$, *then the log-Sobolev inequality*

$$\mu(f^2 \log f^2) \leq C\mu(|\nabla f|^2) + \mu(f^2)\log\mu(f^2), \quad f \in C_b^1(M) \tag{2.2}$$

holds for some constant $C > 0$. *Consequently,* P_t *is hypercontractive. If moreover* $\partial M = \emptyset$, *then exist constants* $c, \lambda > 0$ *such that*

$$W_2(\nu P_t, \mu) \leq ce^{-\lambda t} W_2(\nu, \mu), \quad t \geq 0, \nu \in \mathscr{P}(M). \tag{2.3}$$

To illustrate this result, we present below a consequence with explicit conditions on Ric and ∇Z for which Ric_Z may be negative everywhere.

Corollary 2.2. *If there exist constants* $\delta_1, \delta_2 > 0$ *such that* $\mathrm{Ric} \geq \delta_1$ *and*

$$\nabla Z \leq -\delta_2 \text{ outside a compact set,} \tag{2.4}$$

then Theorem 2.1(1) *holds, and* P_t *has a unique invariant probability measure* μ *such that* (3.4) *and* (2.3) *hold for some constants* $C, c, \lambda > 0$.

Next, we introduce sufficient conditions for (1.7) to hold, these conditions also allow Ric_Z to be negative. Due to technical reason, we will need the ultracontractivity of P_t, which is essentially stronger than the hypercontractivity. We call P_t ultracontractive if $\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} < \infty$ for all $t > 0$. The ultracontractivity implies that P_t has a density $p_t(x,y)$ with respect to μ (called heat kernel) and

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

In references (see e.g. [10]), the ultracontractivity is also defined by $\|P_t\|_{L^2(\mu)\to L^\infty(\mu)} < \infty$ for $t > 0$. When P_t is symmetric in $L^2(\mu)$ we have

$$\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} \le \|P_{t/2}\|^2_{L^2(\mu)\to L^\infty(\mu)}, \quad t > 0, \tag{2.5}$$

so that these two definitions are equivalent. However, when P_t is non-symmetric, the former might be stronger than the latter. The appearance of the ultracontractivity in our study is very nature: by Theorem 2.3(1) we already have (1.7) for $\Phi = \Phi_1$ (the weakest case), and by the ultracontractivity we are able to deduce the inequality from Φ_1 to Φ_∞ (the strongest case). On the other hand, the result also indicates that (1.7) implies the hypercontractivity of P_t.

Theorem 2.3. *Assume that* Ric_Z *is bounded below.*

(1) *If P_t is ultracontractive, then there exist constants $c, \lambda > 0$ such that for any $\Phi \in \bar{\mathscr{N}}$,*

$$W_\Phi(\delta_x P_t, \delta_y P_t) \le \frac{c}{\Phi^{-1}(1)} \mathrm{e}^{-\lambda t} \min\Big\{\rho(x,y),\ G_\Phi(t)\Big\}, \quad t > 0, x, y \in M \tag{2.6}$$

holds for

$$G_\Phi(t) := \inf\Big\{r > 0 :\ (\mu\times\mu)\Big(\Phi\big(r^{-1}\rho\big)\Big) \le \|P_{t/2}\|^{-2}_{L^1(\mu)\to L^\infty(\mu)}\Big\}. \tag{2.7}$$

Consequently, for any $p \in [1,\infty], t \ge 0$ and $\mu_1, \mu_2 \in \mathscr{P}(M)$,

$$W_p(\mu_1 P_t, \mu_2 P_t) \le c\mathrm{e}^{-\lambda t} \min\Big\{W_p(\mu_1,\mu_2),\ \|\rho\|_{L^p(\mu\times\mu)} \|P_{t/2}\|^{\frac{2}{p}}_{L^1(\mu)\to L^\infty(\mu)}\Big\}. \tag{2.8}$$

(2) *On the other hand, if there exist constants $c, \lambda > 0$ such that*

$$W_\infty(\delta_x P_t, \delta_y P_t) \le c\mathrm{e}^{-\lambda t}\rho(x,y), \quad x, y \in M, t \ge 0, \tag{2.9}$$

then the log-Sobolev inequality (3.4) *holds for $c = \frac{2c^2}{\lambda}$, so that P_t is hypercontractive.*

We note that in Theorem 2.3(1) we have $\|\rho\|_{L^p(\mu\times\mu)} < \infty$ for $p \in [1,\infty)$. Indeed, since Ric_Z is bounded below, by [25, Theorem 2.1] the ultracontractivity implies the super log-Sobolev inequality (3.3) below, so that due to Herbst we have $(\mu\times\mu)(\mathrm{e}^{r\rho^2}) < \infty$ for all $r > 0$ (see e.g. [1, 2]). Therefore, $G_\Phi(t) < \infty$ for $t > 0$ and $\Phi \in \mathscr{N}$ satisfying

$$\limsup_{r\to\infty} \frac{\log\Phi(r)}{r^2} < \infty.$$

In the symmetric case (i.e. $Z = \nabla V$ for some $V \in C^2(M)$), explicit sufficient conditions for the ultracontractivity have been introduced in [37] by using the dimension-free Harnack inequality in the sense of [33]. Together with a suitable exponential estimate on the diffusion process, this inequality implies $\|P_t\|_{L^2(\mu)\to L^\infty(\mu)} < \infty$ for $t > 0$ and thus, P_t is ultracontractive due to (2.5). The conditions can be formulated as

$$-\nabla Z \geq \Psi_1 \circ \rho_o \text{ and } \mathrm{Ric} \geq -\Psi_2 \circ \rho_o \text{ hold outside a compact subset of } M, \tag{2.10}$$

where $\Psi_1, \Psi_2 : (0, \infty) \to (0, \infty)$ are increasing functions such that

$$\int_1^\infty \frac{\mathrm{d}s}{\sqrt{s}\int_0^{\sqrt{s}} \Psi_1(u)\mathrm{d}u} < \infty, \quad \lim_{r\to\infty} \min\Big\{\Psi_1(r), \frac{(\int_0^r \Psi_1(s)\mathrm{d}s)^2}{\Psi_1(r)}\Big\} = \infty, \tag{2.11}$$

and for some constants $\theta \in (0, 1/(1+\sqrt{2}))$ and $C > 0$,

$$\sqrt{\Psi_2(r+t)(d-1)} \leq \theta \int_0^r \Psi_1(s)\mathrm{d}s + \frac{1}{2}\int_0^{t/2} \Psi_1(s)\mathrm{d}s + C, \quad r, t \geq 0. \tag{2.12}$$

When Ric is bounded below, (2.12) as well as the second inequality in (2.10) hold for Ψ_2 being a large enough constant. In general, since $\int_0^r \Psi_1(s)\mathrm{d}s \geq 2\int_0^{r/2} \Psi_1(s)\mathrm{d}s$, (2.12) with $\theta = \frac{1}{4} < \frac{1}{1+\sqrt{2}}$ follows from

$$\begin{aligned} \sqrt{\Psi_2(r)(d-1)} &\leq \frac{1}{2} \inf_{t\in[0,r]} \Big\{\int_0^{t/2} \Psi_1(s)\mathrm{d}s + \int_0^{(r-t)/2} \Psi_1(s)\mathrm{d}s\Big\} + C \\ &= \int_0^{r/4} \Psi_1(s)\mathrm{d}s + C, \quad r \geq 0. \end{aligned} \tag{2.13}$$

Since (2.5) fails for non-symmetric semigroups, we apply the inequality

$$\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} \leq \|P_{t/2}\|_{L^1(\mu)\to L^2(\mu)} \|P_{t/2}\|_{L^2(\mu)\to L^\infty(\mu)}$$

due to the semigroup property. So, to ensure the ultracontractivity, we need an additional condition implying $\|P_t\|_{L^1(\mu)\to L^2(\mu)} < \infty$ (see Corollary 2.4(2) below).

To estimate $G_\Phi(t)$ in (2.6) using Ψ_1, we introduce

$$\Lambda_1(r) := \frac{1}{\sqrt{r}}\int_0^{\sqrt{r}} \Psi_1(s)\mathrm{d}s, \quad \Lambda_2(r) := \int_r^\infty \frac{\mathrm{d}\, s}{\sqrt{s}\int_0^{\sqrt{s}} \Psi_1(u)\mathrm{d}u}, \quad r > 0. \tag{2.14}$$

Obviously, the inverse function Λ_2^{-1} exists on $(0, \infty)$, and since Λ_1 is increasing with $\Lambda_1(\infty) = \infty$, we have

$$\Lambda_1^{-1}(r) := \inf\{s \geq 0 : \Lambda_1(s) \geq r\} < \infty, \quad r \geq 0.$$

Corollary 2.4. *Assume that* (2.11) *and* (2.12) *hold for some constants* $\theta \in (0, 1/(1+\sqrt{2}))$ *and* $C > 0$.

(1) *If P_t is symmetric, i.e. $Z = \nabla V$ for some $V \in C^2(M)$, then there exist constants $c, \lambda > 0$ such that* (2.6) *and* (2.8) *hold for*

$$G_\Phi(t) := \inf\Big\{r > 0 :\ (\mu \times \mu)\big(\Phi(r^{-1}\rho)\big) \le \mathrm{e}^{-c-ct^{-1}\{1+\Lambda_1^{-1}(ct^{-1})-\Lambda_2^{-1}(c^{-1}t)\}}\Big\},\quad t > 0.$$

(2) *If P_t is non-symmetric but there exists continuous $h \in C([0,1];[0,\infty))$ with $h(r) > 0$ for $r > 0$ such that $\int_0^1 \frac{h(r)}{r}\mathrm{d}r < \infty$ and*

$$H(\theta) := \int_0^1 \frac{\theta}{h(r)}\Big\{1 + \Lambda_1^{-1}\big(\theta/h(r)\big) + \Lambda_2^{-1}\big(h(r)/\theta\big)\Big\}\mathrm{d}r < \infty,\quad \theta > 0,$$

then there exist constants $c, \lambda > 0$ such that (2.6) *holds for*

$$G_\Phi(t) := \inf\Big\{\lambda > 0 :\ (\mu \times \mu)\big(\Phi(\lambda^{-1}\rho)\big) \le \mathrm{e}^{-c-ct^{-1}\{1+\Lambda_1^{-1}(ct^{-1})-\Lambda_2^{-1}(c^{-1}t)\}-cH(ct^{-1})}\Big\}.$$

To conclude this part, we present a simple example to illustrate Corollary 2.4.

Example 2.1. Let M have non-positive sectional curvatures and a pole $o \in M$. Let $Z = Z_0 - \delta\nabla\rho_o^{2+\varepsilon}$ outside a compact domain, where $\delta, \varepsilon > 0$ are constants and Z_0 is a C^1 vector field with

$$\limsup_{\rho_o\to\infty} \frac{|\nabla Z_0|}{\rho_o^\varepsilon} < \delta(1+\varepsilon)(2+\varepsilon). \tag{2.15}$$

Let $\Psi_2 : (0,\infty) \to (0,\infty)$ be increasing such that

$$\mathrm{Ric} \ge -\Psi_2(\rho_o),\ \lim_{r\to\infty} \frac{\Psi_2(r)}{r^{2(1+\varepsilon)}} = 0. \tag{2.16}$$

By (2.15), (2.16) and the Hessian comparison theorem, we see that (2.10), (2.11) and (2.13) hold with $\Psi_1(r) = c_1 r^\varepsilon$ for some constant $c_1 > 0$. By (2.14), there exists a constant $C > 0$ such that

$$\Lambda_1^{-1}(\theta/h(r)) + \Lambda^{-1}(h(r)/\theta) \le C\theta^{\frac{2}{\varepsilon}} h(r)^{-\frac{2}{v}v}.$$

Taking, for instance, $h(r) = r^{\frac{\varepsilon}{4}}$ in Corollary 2.4(2), we may find out constants $c, \lambda > 0$ such that for any $p \ge 1$,

$$W_p(\mu_1 P_t, \mu_2 P_t) \le c\mathrm{e}^{-\lambda t}\min\Big\{W_p(\mu_1,\mu_2),\ \|\rho\|_{L_p(\mu\times\mu)}\exp\Big[\frac{c}{pt^{1+\frac{2}{\varepsilon}}}\Big]\Big\},\quad t > 0, \mu_1,\mu_2 \in \mathscr{P}(M).$$

2.2 SDEs with multiplicative noise

Consider the following SDE on $\mathbb{R}^d$:

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sqrt{2}\sigma(X_t)\mathrm{d}B_t, \tag{2.17}$$

where B_t is the m-dimensional Brownian motion, $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^m$ (the space of $d \times m$-matrices) are locally Lipshitz such that

$$\|\sigma\|_{HS}^2(x) + \langle b(x), x\rangle \le C(1+|x|^2), \quad x \in \mathbb{R}^d$$

holds for some constant $C > 0$, where and in the following, $\|\cdot\|_{HS}$ and $\|\cdot\|$ denote the Hilbert-Schmidt and the operator norms respectively. Then the SDE has a unique solution $\{X_t(x)\}_{t\ge 0}$ for every initial point $x \in \mathbb{R}^d$. Let P_t be the associated Markov semigroup:

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

We intend to investigate the W_p-exponential contraction for $p \in [1, \infty)$. As mentioned in Introduction that existing results only apply to $p = 1$ and $\sigma = I$, and as mentioned in [12, 18] that there is essential difficulty to prove (1.3) for $p > 1$ even for $\sigma = I$. So, the present study is non-trivial.

Corresponding to that (1.1) implies (1.2) in the Riemannian setting, we have the following assertion.

Theorem 2.5. *Let $p \in [1, \infty)$. If*

$$\begin{aligned}(2.18)\qquad &\frac{(p-2)|(\sigma(x)-\sigma(y))^*(x-y)|^2}{|x-y|^2} + \|\sigma(x)-\sigma(y)\|_{HS}^2 + \langle b(x)-b(y), x-y\rangle \\ &\le -K_p|x-y|^2, \quad x \ne y \in \mathbb{R}^d\end{aligned}$$

holds for some constant $K_p \in \mathbb{R}$, then

$$W_p(\mu_1 P_t, \mu_2 P_t) \le \mathrm{e}^{-K_p t} W_p(\mu_1, \mu_2), \quad t \ge 0, \mu_1, \mu_2 \in \mathscr{P}(\mathbb{R}^d).$$

Note that this result does apply to $p = \infty$ when σ is non-constant. Next, as in the Riemannian case, we intend to prove the exponential contraction in W_p when (2.18) only holds for some negative constant K_p. To this end, we need the SDE to be non-degenerate. The following result contains analogous assertions in Theorems 2.1 and 2.3, where the first assertion extends (1.5) to the multiplicative noise setting.

Theorem 2.6. *Assume that $\lambda_0^{-2} I \ge \sigma\sigma^* \ge \lambda_0^2 I$ for some constant $\lambda_0 \in (0, 1)$.*

(1) *If there exist constants $K_1, K_2, r_0 > 0$ such that Z and $\sigma_0 := \sqrt{\sigma\sigma^* - \lambda_0^2 I}$ satisfy*

$$\begin{aligned}(2.19)\qquad &\|\sigma_0(x)-\sigma_0(y)\|_{HS}^2 - \frac{|(\sigma(x)-\sigma(y))^*(x-y)|^2}{|x-y|^2} + \langle b(x)-b(y), x-y\rangle \\ &\le \big\{(K_1+K_2)1_{\{|x-y|\le r_0\}} - K_2\big\}|x-y|^2, \quad x, y \in \mathbb{R}^d,\end{aligned}$$

then there exist constants $c, \lambda > 0$ such that

$$W_1(\mu_1 P_t, \mu_2 P_t) \le c\mathrm{e}^{-\lambda t} W_1(\mu_1, \mu_2), \quad t \ge 0, \mu_1, \mu_2 \in \mathscr{P}(\mathbb{R}^d).$$

(2) *Let P_t have a unique invariant probability measure μ such that the log-Sobolev inequality*

$$\mu(f^2 \log f^2) \le C\mu(|\nabla f|^2), \quad f \in C_b^1(\mathbb{R}^d), \mu(f^2) = 1 \tag{2.20}$$

holds for some constant $C > 0$. If there exists a constant $K > 0$ such that

$$\|\sigma(x) - \sigma(y)\|_{HS}^2 + \langle b(x) - b(y), x - y\rangle \le K|x - y|^2, \quad x, y \in \mathbb{R}^d, \tag{2.21}$$

then (2.3) *holds for some constants $c, \lambda > 0$ and $M = \mathbb{R}^d$.*

(3) *Let P_t be ultracontractive and let* (2.21) *hold for some constant $K > 0$. Then there exist a constant $\lambda > 0$ such that for any $p \in [1, \infty)$, condition* (2.18) *implies*

$$W_p(\mu_1 P_t, \mu_2 P_t) \le c e^{-\lambda t} W_p(\mu_1, \mu_2), \quad t \ge 0, \mu_1, \mu_2 \in \mathscr{P}(\mathbb{R}^d)$$

for some constant $c = c(p) > 0$.

According to [22, Lemma 3.3], we have

$$\|\sigma_0(x) - \sigma_0(y)\|^2 \le \frac{1}{4\lambda_0}\|(\sigma\sigma^*)(x) - (\sigma\sigma^*)(y)\|_{HS}^2, \quad x, y \in \mathbb{R}^d. \tag{2.22}$$

Combining this with $\|\cdot\|_{HS}^2 \le d\|\cdot\|^2$, we see that (2.19) follows from the following more explicit condition:

$$\begin{aligned} &\frac{d-1}{4\lambda_0}\|(\sigma\sigma^*)(x) - (\sigma\sigma^*)(y)\|_{HS}^2 + \langle b(x) - b(y), x - y\rangle \\ &\le \big\{(K_1 + K_2)1_{\{|x-y| \le r_0\}} - K_2\big\}|x - y|^2, \quad x, y \in \mathbb{R}^d. \end{aligned} \tag{2.23}$$

3 Preparations

This section includes some propositions which will be used to prove the results introduced in Section 2. We first recall a link between the Wasserstein distance and gradient estimates due to [17], then deduce the hyperboundedness and the exponential convergence in entropy from the log-Sobolev inequality for non-symmetric diffusion semigroups, and finally prove the exponential contraction in gradient for ultracontractive semigroups in a general framework including both diffusion and jump Markov semigroups.

3.1 Wasserstein distance and gradient inequalities

Let (E, ρ) be a geodesic Polish space, i.e. it is a Polish space and for any two different points $x, y \in E$, there exists a continuous curve $\gamma : [0, 1] \to E$ such that $\gamma_0 = x, \gamma_1 = y$ and $\rho(\gamma_s, \gamma_t) = |s - t|\rho(x, y)$ for $s, t \in [0, 1]$. Then for any $f \in \mathrm{Lip}_b(E)$, the class of bounded Lipschitz functions on E, the length of gradient

$$|\nabla f|(x) := \limsup_{\rho(x,y)\downarrow 0} \frac{|f(x) - f(y)|}{\rho(x, y)}, \quad x \in E$$

is measurable. Moreover, let $P(x, \mathrm{d}y)$ be a Markov transition kernel and define the Markov operator

$$Pf(x) := \int_E f(y)P(x, \mathrm{d}y), \quad x \in E, f \in \mathscr{B}_b(E).$$

For any $\Phi \in \bar{\mathscr{N}} \setminus \{\Phi_\infty\}$, consider the Young norm induced by Φ with respect to P

$$\|f\|_{L_*^\Phi(P)}(x) := \sup\Big\{P(fg)(x) : g \in \mathscr{B}_b(E), P\Phi(|g|)(x) \le 1\Big\}, \quad x \in E, f \in \mathscr{B}_b(E), \tag{3.1}$$

and set $\|f\|_{L_*^{\Phi_\infty}(P)}(x) = P|f|(x)$. Then $\|\cdot\|_{L_*^{\Phi_p}} = \|\cdot\|_{L^{\Phi_q}}$ for $p \in [1, \infty], q = \frac{p}{p-1}$. The following result follows from [17, Theorem 2.2, Remark 2 and Remark 3].

Proposition 3.1 ([17])**.** *For any constant $C > 0$ and $\Phi \in \bar{\mathscr{N}}$, the following statements are equivalent to each other:*

(1) $|\nabla Pf| \le C\|\nabla f\|_{L_*^\Phi(P)}$ *for* $f \in \mathrm{Lip}_b(E)$.

(2) $W_\Phi(\delta_x P, \delta_y P) \le C\rho(x, y), \quad x, y \in E$.

When $\Phi = \Phi_p$ for $p \in [1, \infty]$, they are also equivalent to

(3) $W_p(\mu_1 P, \mu_2 P) \le CW_p(\mu_1, \mu_2), \quad \mu_1, \mu_2 \in \mathscr{P}(E)$.

3.2 Hyperboundedness and exponential convergence in entropy

When P_t is symmetric, it is well known that the hyperbounddeness, exponential convergence in entropy and the log-Sobolev inequality are equivalent each other, see [6, 36] and references within. In the non-symmetric case, the log-Sobolev inequality implies the former two properties if the generator L and the symmetric part of the Dirichlet form $\mathscr{E}$ satisfy

$$\begin{aligned} &-\mu((1 + \log f)Lf) \ge c_0\mathscr{E}\big(\sqrt{f}, \sqrt{f}\big) \text{ and} \\ &-\mu(f^{p-1}Lf) = \frac{c_0(p-1)}{p^2}\mathscr{E}(f^{\frac{p}{2}}, f^{\frac{p}{2}}), \quad p > 1, f \in \mathscr{D} \end{aligned} \tag{3.2}$$

for some constant $c_0 > 0$ and a reasonable class $\mathscr{D}$ of non-negative bounded functions, which is stable under P_t and dense in $L_+^p(\mu) := \{f \in L^p(\mu) : f \ge 0\}$ for any $p \ge 1$, see e.g. [14]. In applications, it may be not easy to figure out the class $\mathscr{D}$ such that (3.2) holds. But in general this condition can be replaced by the following approximation formula Lemma 3.2 in the spirit of [26].

Now, consider the (Neumann) semigroup P_t generated by $L := \Delta + Z$ for a locally bounded vector field Z such that P_t has a unique invariant probability measure μ. Let

$$\mathscr{D}_0 = \big\{f \in C_0^\infty(M) : \ f \text{ satisfies the Neumann condition if } \partial M \neq \emptyset\big\}.$$

Then $(L, \mathscr{D}_0)$ is dissipative (thus, closable) in $L^1(\mu)$ with closure $(L, \mathscr{D}_1(L))$ generating P_t in $L^1(\mu)$, see e.g. [29] and references therein. Let

$$\mathscr{D} = \{f \in \mathscr{D}_1(L) \cap L^\infty(\mu) : \ f \ge 0\}.$$

Lemma 3.2. *Let $f \in \mathscr{D}$ and $\psi \in C_b^\infty([\mathrm{ess}_\mu \inf f, \infty))$. There exists a sequence $\{f_n\}_{n\geq 1} \subset \mathscr{D}_0$ with $\inf f_n = \mathrm{ess}_\mu \inf f$ such that $f_n \to f$ in $L^m(\mu)$ for any $m \geq 1$, $Lf_n \to Lf$ in $L^1(\mu)$, and*

$$\mu(\psi(f)Lf) = -\lim_{n\to\infty} \mu(\psi'(f_n)|\nabla f_n|^2).$$

Proof. Since $f \in \mathscr{D} \subset \mathscr{D}_1(L) \cap L^\infty(\mu)$, there exists a uniformly bounded sequence $\{f_n\}_{n\geq 1} \subset \mathscr{D}_0$ such that $\inf f_n = \mathrm{ess}_\mu \inf f$ and $f_n \to f, Lf_n \to Lf$ in $L^1(\mu)$. By the uniform boundedness, $f_n \to f$ in $L^m(\mu)$ for any $m \geq 1$. Since $\psi \in C_b^\infty([\inf f_n, \infty))$,

$$g_n := \int_{\inf f_n}^{f_n} \psi(s)\mathrm{d}s \in \mathscr{D}_c := \{g + c : \ c \in \mathbb{R}, g \in \mathscr{D}_0\} \subset \mathscr{D}_1(L).$$

This implies $\mu(Lg_n) = 0$ since μ is P_t-invariant. So, by the dominated convergence theorem,

$$\mu(\psi(f)Lf) = \lim_{n\to\infty} \mu(\psi(f_n)Lf_n) = \lim_{n\to\infty} \mu(Lg_n - \psi'(f_n)|\nabla f_n|^2) = -\lim_{n\to\infty} \mu(\psi'(f_n)|\nabla f_n|^2).$$

□

Proposition 3.3. *Let Z be a locally bounded vector field such that the (Neumann) semigroup P_t generated by $L := \Delta + Z$ has a unique invariant probability measure μ.*

(1) *If the super log-Sobolev inequality*

$$\mu(f^2 \log f^2) \leq r\mu(|\nabla f|^2) + \beta(r), \quad r > 0, \quad f \in C_b^1(M), \mu(f^2) = 1. \tag{3.3}$$

holds for some $\beta \in C((0,\infty);(0,\infty))$, then for any constants $q > p \geq 1$ and $\gamma \in C((p,q);(0,\infty))$ such that $t := \int_p^q \frac{\gamma(r)}{r}\mathrm{d}r < \infty$, there holds

$$\|P_t\|_{L^p(\mu)\to L^q(\mu)} \leq \exp\left[\int_p^q \frac{\beta(4\gamma(r)(1-r^{-1}))}{r^2}\,\mathrm{d}r\right].$$

(2) *If the log-Sobolev inequality*

$$\mu(f^2 \log f^2) \leq C\mu(|\nabla f|^2) + \mu(f^2)\log \mu(f^2), \quad f \in C_b^1(M) \tag{3.4}$$

holds for some constant $C > 0$, then

$$\mu((P_t g)\log P_t g) \leq \mathrm{e}^{-4t/C}\mu(g \log g), \quad g \in \mathscr{B}_b(M), g \geq 0, \mu(g) = 1.$$

Proof. (1) According to Lemma 3.2, for any $f \in \mathscr{D}$ and $p > 1$, there exists $\{f_n\}_{n\geq 1} \subset \mathscr{D}_0$ such that $f_n^{\frac{2}{p}} \to f$ in $L^m(\mu)$ for all $m \geq 1$, and

$$-\mu(f^{p-1}Lf) = \frac{4(p-1)}{p^2} \limsup_{n\to\infty} \mu(|\nabla f_n|^2). \tag{3.5}$$

Applying (3.3) to f_n and using (3.5), we obtain

$$\begin{aligned} p\mu(f^p \log f) &= \lim_{n\to\infty} \mu(f_n^2 \log f_n^2) \le r \liminf_{n\to\infty} \mu(|\nabla f_n|^2) + \beta(r) \\ &\le \frac{rp^2}{4(p-1)}\Big(-\mu(f^{p-1}Lf) + \frac{4\beta(r)(p-1)}{rp^2}\Big), \quad r>0. \end{aligned}$$

Set $c(p) = \frac{rp}{4(p-1)}$, we have

$$\frac{4\beta(r)(p-1)}{rp^2} = \frac{\beta(4c(p)(1-p^{-1}))}{pc(p)}, \quad p>1,$$

so that the above inequality becomes

$$\mu(f^p \log f) \le c(p)\Big(-\mu(f^{p-1}Lf) + \gamma(p)\Big), \quad p>1, f\in \mathscr{D}$$

for $\gamma(p) := \frac{\beta(4c(p)(1-p^{-1}))}{pc(p)}$. Noting that $\mathscr{D}$ is P_t-invariant (i.e. $P_t\mathscr{D} \subset \mathscr{D}$) and dense in $L^p_+(\mu)$ for any $p \ge 1$, the desired assertion follows from the proof of [14, Corollary 3.13].

(2) It suffices to prove for $g \in \mathscr{D}$ with $\inf g > 0$. Applying Lemma 3.2 to $f = P_t g$ and $\psi(s) = 1 + \log s$, and using (3.4), we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mu((P_tg)\log P_tg) &= \mu((1+\log P_tg)LP_tg) = -4\lim_{n\to\infty}\mu\big(\big|\nabla\sqrt{f}_n\big|^2\big) \\ &\le -\frac{4}{C}\liminf_{n\to\infty}\mu(f_n\log f_n) = -\frac{4}{C}\mu((P_tg)\log P_tg), \quad t\ge 0. \end{aligned}$$

This implies the desired exponential estimate. □

3.3 Exponential contraction in gradient

In this part, we consider a general framework including both diffusion and jump processes. Let $(E, \mathscr{F}, \mu)$ be a separable complete probability space, and let P_t be a Markov semigroup on $L^2(\mu)$ with μ as invariant probability measure. Let $(L, \mathscr{D}(L))$ be the generator of P_t in $L^2(\mu)$. We assume that there exists an algebra $\mathscr{A} \subset \mathscr{D}(L)$ such that

(i) $1 \in \mathscr{A}$, $\mathscr{A}$ is dense in $L^2(\mu)$ and the algebra induced by

$$\mathscr{D} := \{P_s f : s \ge 0, f \in \mathscr{A}\}$$

is contained in $\mathscr{D}(L)$.

(ii) $\Gamma(f,g) := \frac{1}{2}(L(fg) - fLg - gLf)$ gives rise to a non-degenerate positive definite bilinear form on $\mathscr{D}\times\mathscr{D}$; i.e., for any $f \in \mathscr{D}$, $\Gamma(f,f) \ge 0$ and it equals to 0 if and only if f is constant.

In particular, when P_t is the (Neumann) semigroup generated by $L := \Delta + Z$ on M with Ric_Z bounded below, the assumption holds for

$$\mathscr{A} := \{f + c : f \in C_0^\infty(M) \text{ satisfying the Neumann condition if } \partial M \neq \emptyset, c \in \mathbb{R}\}.$$

Under the above conditions,

$$\mathscr{E}(f,g) := \mu(\Gamma(f,g)), \quad f, g \in \mathscr{A}$$

is closable and the closure $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ is a conservative symmetric Dirichlet form. Although P_t is not associated to $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ when it is non-symmetric, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu((P_t f)^2) = -2\mathscr{E}(P_t f, P_t f), \quad t \geq 0, f \in \mathscr{D}. \tag{3.6}$$

If $\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} < \infty$, then P_t has a heat kernel $p_t(x,y)$ with respect to μ, i.e.

$$P_t f = \int_E p_t(\cdot, y) f(y) \mu(\mathrm{d}y), \quad f \in L^2(\mu),$$

and

$$\mathrm{ess}_{\mu\times\mu} \sup p_t = \|P_t\|_{L^1(\mu)\to L^\infty(\mu)} < \infty.$$

We consider the "gradient" length $|\nabla_\Gamma f| = \sqrt{\Gamma(f,f)}$ induced by Γ. Note that for jump processes the length is non-local and thus essentially different from the usual gradient length. As shown below that estimates of $|\nabla_\Gamma P_t|$ have a close link to functional inequalities of the associated Dirichlet form.

Proposition 3.4. *Assume that there exist $t_1 > 0$ and $\eta \in C([0,\infty); (0,\infty))$ such that*

$$\|P_{t_1}\|_{L^1(\mu)\to L^\infty(\mu)} < \infty, \quad |\nabla_\Gamma P_t f|^2 \leq \eta(t) P_t |\nabla_\Gamma f|^2, \ t \geq 0, f \in \mathscr{D}. \tag{3.7}$$

Then there exist constants $c, \lambda, t_2 > 0$ such that for any $q \geq 1$ and $\eta_q \in C([0,\infty); (0,\infty))$, the gradient estimate

$$|\nabla_\Gamma P_t f|^2 \leq \eta_q(t) (P_t |\nabla_\Gamma f|^q)^{\frac{2}{q}}, \ t \geq 0, f \in \mathscr{D} \tag{3.8}$$

implies

$$\|\nabla_\Gamma P_t f\|^2_{L^\infty(\mu)} \leq \Big(c \sup_{[0,t_2]} \eta_q\Big) \mathrm{e}^{-\lambda t} \mathrm{ess}_\mu \inf (P_t |\nabla_\Gamma f|^q)^{\frac{2}{q}}, \quad t \geq t_2, f \in \mathscr{D}. \tag{3.9}$$

Proof. (a) We first prove

$$\mathscr{E}(P_t f, P_t f) \leq C \mathrm{e}^{-\lambda t} \mathscr{E}(f,f), \quad f \in \mathscr{D}, t \geq 0 \tag{3.10}$$

for some constants $C, \lambda > 0$. By the second inequality in (3.7), for any $t > 0$ and $f \in \mathscr{D}$ we have

$$\frac{\mathrm{d}}{\mathrm{d}s} P_s (P_{t-s} f)^2 = 2 P_s |\nabla_\Gamma P_{t-s} f|^2 \leq 2\eta(t-s) P_t |\nabla_\Gamma f|^2, \quad s \in [0,t].$$

Integrating both sides over $[0,t]$ leads to

$$P_t f^2 \le (P_t f)^2 + C(t) P_t |\nabla_\Gamma f|^2, \quad C(t) := 2\int_0^t \eta(s)\mathrm{d}s,\ t>0.$$

Taking $t=t_1$ and noting that μ is the invariant probability measure of P_t, we obtain

$$\mu(f^2) \le C(t_1)\mathscr{E}(f,f) + \|P_{t_1}\|_{1\to\infty}^2 \mu(|f|)^2, \quad f\in\mathscr{D}. \tag{3.11}$$

Since $\mathscr{D}(\mathscr{E})$ is the closure of $\mathscr{D}$ under the $\mathscr{E}_1$-norm, this inequality also holds for $f\in\mathscr{D}(\mathscr{E})$. By condition (ii), the symmetric Dirichlet form is irreducible. So, according to [41, Corollary 1.2] the defective Poincaré inequality (3.11) implies the Poincaré inequality

$$\mu(f^2) \le \frac{1}{\lambda}\mathscr{E}(f,f) + \mu(f)^2, \quad f\in\mathscr{D}(\mathscr{E}) \tag{3.12}$$

for some constant $\lambda>0$. By (3.6) and that $\mathscr{D}$ is dense in $L^2(\mu)$, the Poincaré inequality is equivalent to

$$\|P_t f - \mu(f)\|_2 \le \mathrm{e}^{-\lambda t}\|f-\mu(f)\|_2, \quad t\ge 0, f\in L^2(\mu). \tag{3.13}$$

On the other hand, by the second inequality in (3.7), for any $t>0$ and $f\in\mathscr{D}$ we have

$$\frac{\mathrm{d}}{\mathrm{d}s} P_s (P_{t-s} f)^2 = 2P_s|\nabla_\Gamma P_{t-s} f|^2 \ge \frac{2}{\eta(s)}|\nabla_\Gamma P_t f|^2, \quad s\in[0,t].$$

So,

$$|\nabla_\Gamma P_t f|^2 \le \frac{P_t f^2 - (P_t f)^2}{2\int_0^t \eta(s)^{-1}\mathrm{d}s}, \quad t>0, f\in\mathscr{D}.$$

Using $P_t f - \mu(f)$ to replace f and integrating with respect to μ, we obtain

$$\mathscr{E}(P_{2t}f, P_{2t}f) \le \frac{\|P_t f - \mu(f)\|_2^2}{2\int_0^t \eta(s)^{-1}\mathrm{d}s}, \quad t>0, f\in\mathscr{D}.$$

Combining this with (3.13) and (3.12) we arrive at

$$\mathscr{E}(P_t f, P_t f) \le c_1 \mathrm{e}^{-\lambda t}\mathscr{E}(f,f), \quad t\ge 1, f\in\mathscr{D}$$

for some constant $c_1>0$; that is, (3.10) holds for $t>1$. Finally, (3.7) implies (3.10) for $t\in[0,1]$.

(b) Next, we intend to find out a constant $t_0\ge t_1$ such that

$$\frac{1}{2}\le p_t \le 2, \quad (\mu\times\mu)\text{-}a.e., t\ge t_0. \tag{3.14}$$

Indeed, by (3.13) and the first inequality in (3.7), we obtain

$$\left|\int_E (p_{t+2t_1}(\cdot,y)-1)f(y)\mu(\mathrm{d}y)\right| = |P_{t_1}(P_{t+t_1}f - \mu(f))|$$

$$\le c_0\mu(|P_{t+t_1}f-\mu(f)|)\le c_0\mathrm{e}^{-\lambda t}\|P_{t_1}f-\mu(f)\|_2\le c_0^2\mathrm{e}^{-\lambda t}\mu(|f|),\quad \mu\text{-}a.e.,\ t\ge 0,$$

where $c_0:=\|P_{t_1}\|_{L^1(\mu)\to L^\infty(\mu)}$. This implies the desired assertion for $t_0>0$ such that $c_0^2\mathrm{e}^{-\lambda t_0}\le \frac{1}{2}$.

(c) Finally, combining (3.7), (3.14), (3.10) and (3.8), we obtain

$$\begin{aligned}&\|\nabla_\Gamma P_{t+2t_0}f\|^2_{L^\infty(\mu)}\le c_1\|P_{t_0}|\nabla_\Gamma P_{t+t_0}f|^2\|_{L^\infty(\mu)}\le 2c_1\mathscr{E}(P_{t+t_0}f,P_{t+t_0}f)\\&\le c_2\mathrm{e}^{-\lambda t}\mathscr{E}(P_{t_0}f,P_{t_0}f)\le c_2\eta_q(t_0)\mathrm{e}^{-\lambda t}\mu\big((P_{t_0}|\nabla_\Gamma f|^q)^{\frac{2}{q}}\big)\\&\le c_3\eta_q(t_0)\mathrm{e}^{-\lambda t}\mathrm{ess}_\mu\inf(P_{t+2t_0}|\nabla_\Gamma f|^q)^{\frac{2}{q}}\end{aligned}$$

for some constants $c_1,c_2,c_3>0$. Then (3.9) holds for $t_2=2t_0$. □

4 Proofs of Theorem 2.1 and Corollary 2.2

The first assertion is a generalization of the main result in [18] where $M=\mathbb{R}^d$ is considered. As in [18], the key point of the proof is to construct a coupling by parallel transform for long distance but by reflection for short distance. The only difference is that we are working on a non-flat Riemannian manifold for which the curvature term appears in calculations. Since Itô's formula of the distance process has been well developed for couplings by both parallel displacement and reflection, the proof is also straightforward.

The proofs of the other two assertions are based on the log-Sobolev inequality and the log-Harnack inequality derived in [25] and [39] respectively for bounded below Ric_Z.

Proof of Theorem 2.1. (a) For two different points $x,y\in M$, let $P_{x,y}:T_xM\to T_yM$ be the parallel displacement along the minimal geodesic $\gamma:[0,\rho(x,y)]\to M$ from x to y, and let $M_{x,y}:=P_{x,y}-2\langle\cdot,\dot\gamma_0\rangle\dot\gamma_{\rho(x,y)}:T_xM\to T_yM$ be the mirror reflection. Both maps are smooth in (x,y) outside the cut-locus $\mathrm{Cut}(M)$. According to [15] and [32], the appearance of the cut-locus and/or a convex boundary helps for the success of coupling, i.e. it makes the distance between two marginal processes smaller. So, for simplicity, we may and do assume that both the cut-locus and the boundary are empty, see [3, Section 3] or [36, Chapter 2] for details.

Now, let X_t solve the SDE

$$\mathrm{d}_IX_t=\sqrt{2}u_t\mathrm{d}B_t+Z(X_t)\mathrm{d}t,\quad X_0=x,$$

where d_I denotes the Itô differential introduced in [13] on Riemannian manifolds, B_t is the d-dimensional Brownian motion, and u_t is the horizontal lift of X_t to the frame bundle $O(M)$. Then X_t is a diffusion process generated by L. To construct the coupling by reflection for short distance and parallel displacement for long distance, we introduce a cut-off function $h\in C^1([0,\infty))$ which is decreasing such that $h(r)=1$ for $r\le r_0$, $h(r)=0$ for $r\ge r_0+1$, and $\sqrt{1-h^2}$ is also in C^1, see e.g. [43, (3.1)] for a concrete example. To construct the coupling in the above spirit, we split the noise into two parts, i.e. to replace $\mathrm{d}B_t$ by $h(\rho(X_t,Y_t))\mathrm{d}B_t'+\sqrt{1-h(\rho(X_t,Y_t))^2}\mathrm{d}B_t''$ for two independent Brownian motions B_t' and

B_t'', then make reflection for the B_t' part and parallel displacement for the B_t'' part. More precisely, let (X_t, Y_t) solve the following SDE on $M \times M$ for $(X_0, Y_0) = (x, y)$:

$$\begin{aligned}
\mathrm{d}_I X_t &= \sqrt{2}\Big(h(\rho(X_t,Y_t))u_t\mathrm{d}B_t' + \sqrt{1-h(\rho(X_t,Y_t))^2}u_t\mathrm{d}B_t''\Big) + Z(X_t)\mathrm{d}t,\\
\mathrm{d}_I Y_t &= \sqrt{2}\Big(h(\rho(X_t,Y_t))M_{X_t,Y_t}u_t\mathrm{d}B_t' + \sqrt{1-h(\rho(X_t,Y_t))^2}P_{X_t,Y_t}u_t\mathrm{d}B_t''\Big) + Z(Y_t)\mathrm{d}t.
\end{aligned}$$

Since the coefficients of the SDE are at least C^1 outside the diagonal $\{(z,z) : z \in M\}$, it has a unique solution up to the coupling time

$$T := \inf\{t \ge 0 : X_t = Y_t\}.$$

We then let $X_t = Y_t$ for $t \ge T$ as usual. By the second variational formula and the index lemma (see e.g. the proof of [37, Lemma 2.3] and [32, (2.4)]), the process $\rho_t := \rho(X_t, Y_t)$ satisfies

$$\mathrm{d}\rho_t \le 2\sqrt{2}h(\rho_t)\mathrm{d}b_t + I_Z(X_t,Y_t)\mathrm{d}t, \quad t \le T$$

for some one-dimensional Brownian motion b_t. Thus, by condition (1.8),

$$\mathrm{d}\rho_t \le 2\sqrt{2}h(\rho_t)\mathrm{d}b_t + \big\{(K_1+K_2)1_{\{\rho_t\le r_0\}} - K_2\big\}\rho_t\mathrm{d}t, \quad t \le T. \tag{4.1}$$

Since $h(\rho_t) = 0$ for $\rho_t \ge r_0 + 1$ while $\mathrm{d}\rho_t < 0$ when $\rho_t \ge r_0 + 1$, this implies

$$\rho_t \le (r_0+1) \vee \rho_0 \le 1 + r_0 + \rho(x,y). \tag{4.2}$$

On the other hand, since $h(\rho_t) = 1$ for $\rho_t \le r_0$, as observed in [18] we have

$$\mathbb{E}\rho_t \le c\mathrm{e}^{-\lambda t}\rho(x,y), \quad t \ge 0 \tag{4.3}$$

for some constants $c, \lambda > 0$. Indeed, let

$$\bar\rho_t = \varepsilon\rho_t + 1 - \mathrm{e}^{-N\rho_t}, \quad N = \frac{r_0}{2}(K_1+K_2), \varepsilon = N\mathrm{e}^{-Nr_0}.$$

Then

$$\varepsilon\rho_t \le \bar\rho_t \le (N+\varepsilon)\rho_t, \quad \frac{4N^2}{r(\varepsilon\mathrm{e}^{Nr}+N)} \ge K_1 + K_2 \text{ for } r \in (0, r_0],$$

so that (4.1) and Itô's formula lead to

$$\begin{aligned}
\mathrm{d}\bar\rho_t &\le 2\sqrt{2}(\varepsilon + N\mathrm{e}^{-N\rho_t})h(\rho_t)\mathrm{d}b_t\\
&\quad + (\varepsilon + N\mathrm{e}^{-N\rho_t})\Big\{(K_1+K_2)1_{\{\rho_t\le r_0\}} - K_2 - \frac{4N^2}{\rho_t(\varepsilon\mathrm{e}^{N\rho_t}+N)}1_{\{\rho_t\le r_0\}}\Big\}\rho_t\mathrm{d}t\\
&\le 2\sqrt{2}(\varepsilon + N\mathrm{e}^{-N\rho_t})h(\rho_t)\mathrm{d}b_t - c_1\bar\rho_t\mathrm{d}t, \quad t \le T
\end{aligned}$$

for some constant c_1. This implies $\mathbb{E}\bar\rho_t \le \bar\rho_0\mathrm{e}^{-c_1 t}$. Then (4.3) holds for some constants $c, \lambda > 0$. Combining (4.2) with (4.3) we arrive at

$$\mathbb{E}\Phi(\rho_t/r) \le \sup_{s\in(0,1+r_0+\rho_0]}\frac{\Phi(s/r)}{s}\mathbb{E}\rho_t \le c\mathrm{e}^{-\lambda t}\rho(x,y)\sup_{s\in(0,1+r_0+\rho_0]}\frac{\Phi(s/r)}{s}.$$

So,

$$\begin{aligned} W_\Phi(\delta_x P_t, \delta_y P_t) &\le \|\rho_t\|_{L^\Phi(\mathbb{P})} = \inf\big\{r>0 : \mathbb{E}\Phi(\rho_t/r)\le 1\big\} \\ &\le \inf\Big\{r>0 : \sup_{s\in(0,1+r_0+\rho(x,y)]} \frac{\Phi(\frac{s}{r})}{s} \le \frac{\mathrm{e}^{\lambda t}}{c\rho(x,y)}\Big\}, \end{aligned}$$

which proves (2.1). Therefore, the proof of (1) is finished since the second inequality therein is a simple consequence of (2.1).

(b) To prove the log-Sobolev inequality (3.4), we follow the line of [37] to establish the dimension-free Harnack inequality using coupling by change of measures. Let (x_t, y_t) be the coupling constructed in [37] before Lemma 2.3 therein with Z replacing ∇V and

$$\xi_t = C + \frac{\rho(x,y)}{T} \tag{4.4}$$

for a constant $C > 0$. We first prove that when C is large enough our condition implies $x_T = y_T$ as in [37, Lemma 2.3]. Indeed, by (1.8) and [37, (2.4)], we have

$$\mathrm{d}\rho(x_t,y_t) = \big\{I_Z(x_t,y_t) - \xi_t\big\}\mathrm{d}t \le \Big\{(K_1+K_2)r_0 - C - \frac{\rho(x,y)}{T}\Big\}\mathrm{d}t, \quad t<\tau,$$

where $\tau := \inf\{t\ge 0 : x_t = y_t\}$ is the coupling time. Taking $C = (K_1+K_2)r_0$ we obtain

$$\mathrm{d}\rho(x_t,y_t) \le -\frac{\rho(x,y)}{T}\mathrm{d}t, \quad t<\tau,$$

so that $\tau \le T$ as desired.

Next, let R be given in the proof of [37, Proposition 3.1] for the present ξ_t in (4.4):

$$R = \exp\bigg[-\frac{1}{\sqrt{2}}\int_0^\tau \mathrm{d}M_t - \frac14\int_0^\tau \xi_t^2\mathrm{d}t\bigg],$$

where M_t is a martingale with $\mathrm{d}\langle M\rangle_t = \frac12\xi_t^2\mathrm{d}t$. Then there exists a constant $C>0$ such that for any $\alpha>1$,

$$\begin{aligned} \mathbb{E}R^{\frac{\alpha}{\alpha-1}} &= \mathbb{E}\exp\bigg[-\frac{\alpha}{(\alpha-1)\sqrt{2}}\int_0^\tau \mathrm{d}M_t - \frac{\alpha^2}{4(\alpha-1)^2}\int_0^\tau \xi_t^2\mathrm{d}t + \frac{\alpha}{4(\alpha-1)^2}\int_0^\tau \xi_t^2\mathrm{d}t\bigg] \\ &\le \exp\bigg[\frac{C\alpha}{(\alpha-1)^2}\Big(T + \frac{\rho(x,y)^2}{T}\Big)\bigg]. \end{aligned}$$

By [37, (3.2)], this implies the Harnack inequality

$$(P_t f(y))^\alpha \le P_T f^\alpha(x)\exp\bigg[\frac{C\alpha}{\alpha-1}\Big(T+\frac{\rho(x,y)^2}{T}\Big)\bigg], \quad 0\le f\in\mathscr{B}(M), x,y\in M, T>0.$$

Combining this with $\mu(\mathrm{e}^{\varepsilon\rho_o^2})<\infty$, it is easy to see that $\|P_T\|_{2\to4}<\infty$ for large $T>0$. Since (1.8) implies $\mathrm{Ric}_Z \ge -(K_1+K_2)$, by the hyperboundedness and [25, Theorem 2.1], we have the defective log-Sobolev inequality

$$\mu(f^2\log f^2) \le C_1\mu(|\nabla f|^2) + C_2, \quad f\in C_b^1(M), \mu(f^2)=1$$

for some constants $C_1, C_2 > 0$. Since the symmetric Dirichlet form $\mathscr{E}(f,g) := \mu(\langle \nabla f, \nabla g\rangle)$ with domain $H^{1,2}(\mu)$ is irreducible, according to [41] (see also [19]), the log-Sobolev inequality (3.4) holds for some constant $C > 0$.

(c) When $\partial M = \emptyset$, the log-Sobolev inequality implies the Talagrand inequality

$$W_2(f\mu,\mu)^2 \le \frac{C}{2}\mu(f\log f), \quad f\ge 0, \mu(f)=1, \tag{4.5}$$

see [5, 35, 21]. Next, let P_t^* be the adjoint of P_t in $L^2(\mu)$. Since μ is P_t-invariant, P_t^* is generated by $L^* := \Delta + Z^*$ for the C^1-vector field $Z^* = 2\nabla V - Z$, where $V = \log \frac{\mathrm{d}\mu}{\mathrm{d}x}$. So, by Proposition 3.3 for P_t^* in place of P_t, the log-Sobolev inequality implies

$$\mu((P_t^*f)\log P_t^* f) \le \mathrm{e}^{-4t/C}\mu(f\log f), \quad t\ge 0, f\ge 0, \mu(f)=1. \tag{4.6}$$

Moreover, according to [39, Theorem 1.1], the curvature condition $\mathrm{Ric}_Z \ge -(K_1+K_2) =: -K$ is equivalent to the log-Harnack inequality

$$P_t(\log f)(x) \le \log P_t f(y) + \frac{K\rho(x,y)^2}{2(1-\mathrm{e}^{-2Kt})}, \quad t\ge 0, x,y\in M, 0\le f\in \mathscr{B}_b(M).$$

By [42, Proposition 1.4.4(3)], this implies

$$\mu((P_t^*f)\log P_t^* f) \le \frac{K}{2(1-\mathrm{e}^{-2Kt})}W_2(f\mu,\mu)^2, \quad f\ge 0, \mu(f)=1, t>0. \tag{4.7}$$

Combining (4.5), (4.6) and (4.7), we obtain

$$\begin{aligned} &W_2((f\mu)P_{1+t},\mu)^2 = W_2((P_{1+t}^*f)\mu,\mu)^2 \le \frac{C}{2}\mu((P_{1+t}^*f)\log P_{1+t}^*f)\\ &\le \frac{C}{2}\mathrm{e}^{-4t/C}\mu((P_1^*f)\log P_1^*f) \le c_1\mathrm{e}^{-4t/C}W_2(f\mu,\mu)^2, \quad t\ge 0, f\ge 0, \mu(f)=1 \end{aligned} \tag{4.8}$$

for some constant $c_1 > 0$. Noting that $\mathrm{Ric}_Z \ge -K$ implies $|\nabla P_t f| \le \mathrm{e}^{Kt}P_t|\nabla f|$ (see e.g. [39]), by Proposition 3.1 we have

$$W_2((f\mu)P_t,\mu) = W_2((f\mu)P_t,\mu P_t) \le c_2 W_2(f\mu,\mu), \quad t\in[0,1], f\ge 0, \mu(f)=1.$$

Combining with (4.8) yields

$$W_2((f\mu)P_t,\mu) \le c\mathrm{e}^{-\lambda t}W_2(f\mu,\mu), \quad t\ge 0, f\ge 0, \mu(f)=1$$

for some constants $c,\lambda > 0$. Therefore, the proof of (3) is finished. □

Proof of Corollary 2.2. By Theorem 2.1, it suffices to verify condition (1.8) and to prove that P_t has a unique invariant probability measure μ with $\mu(\mathrm{e}^{\varepsilon\rho_o^2}) < \infty$ for some constant $\varepsilon > 0$.

Since $\mathrm{Ric} \ge -\delta_1$, [37, (2.5)] applies to $K(x,y) = \delta_1$. So,

$$I(x,y) \le 2\sqrt{\delta_1(d-1)}. \tag{4.9}$$

On the other hand, by $\nabla Z \le -\delta_2$ outside a compact set, there exist constants $c, r > 0$ such that

$$\nabla Z \le -\delta_2 + c1_{\{\rho_o \le r\}}.$$

Then, letting $\gamma : [0, \rho(x,y)] \to M$ be the minimal geodesic from x to y with $|\dot\gamma| = 1$, we obtain

$$\begin{aligned}\langle Z, \nabla\rho(\cdot,y)\rangle(x) + \langle Z, \nabla\rho(x,\cdot)\rangle(y) &= \int_0^{\rho(x,y)} \frac{\mathrm{d}}{\mathrm{d}s}\langle Z(\gamma_s), \dot\gamma_s\rangle \mathrm{d}s \\ = \int_0^{\rho(x,y)} \langle \nabla_{\dot\gamma_s} Z(\gamma_s), \dot\gamma_s\rangle \mathrm{d}s &\le -\delta_2\rho(x,y) + 2cr.\end{aligned}$$

Combining this with (4.9) we prove (1.8) for some constants $K_1, K_2 > 0$.

Next, by $\mathrm{Ric} \ge -\delta_1$ and $\nabla Z \le -\delta_2$ outside a compact set, there exist constants $c_1, c_2 > 0$ such that

$$L\rho_o^2 \le c_1 - c_2\rho_0^2$$

for some constants $c_1, c_2 > 0$. So, when $\varepsilon > 0$ is small enough,

$$L\mathrm{e}^{\varepsilon\rho_o^2} \le c(\varepsilon) - \gamma(\varepsilon)\mathrm{e}^{\varepsilon\rho_o^2}$$

holds for some constants $c(\varepsilon), \delta(\varepsilon) > 0$. By a standard argument this implies that P_t has an invariant probability measure μ with $\mu(\mathrm{e}^{\varepsilon\rho_o^2}) < \infty$. The uniqueness of μ follows from the irreducibility and strong Feller property which is well known for the present framework. □

5 Proof of Theorem 2.3 and Corollary 2.4

Proof of Theorem 2.3. (1) Since $\mathrm{Ric}_Z \ge -K$ for some constant $K \ge 0$, we have (see e.g. [39])

$$|\nabla P_t f| \le \mathrm{e}^{Kt} P_t|\nabla f|, \quad f \in C_b^1(M).$$

Combining this with Proposition 3.4 for $q = 1$ and noting that $P_t|\nabla f|$ is continuous, we obtain

$$|\nabla P_t f| \le c_0\mathrm{e}^{-\lambda t} P_t|\nabla f|, \quad t \ge t_0, f \in C_b^1(M)$$

for some constants $c_0, \lambda, t_0 > 0$. Obviously, (3.1) implies

$$\|\cdot\|_{L^1(P_t)} \le \frac{\|\cdot\|_{L_*^\Phi(P_t)}}{\Phi^{-1}(1)}, \quad \Phi \in \mathscr{N}.$$

Then

$$|\nabla P_t f| \le \frac{c_0}{\Phi^{-1}(1)} \mathrm{e}^{-\lambda t} \|\nabla f\|_{L_*^\Phi(P_t)}, \quad t \ge 0, \Phi \in \mathscr{N}, f \in C_b^1(M).$$

According to Proposition 3.1, this is equivalent to

$$W_\Phi(\delta_x P_t, \delta_y P_t) \le \frac{c_0}{\Phi^{-1}(1)} \mathrm{e}^{-\lambda t} \rho(x,y), \quad t \ge 0, x, y \in M. \tag{5.1}$$

On the other hand, noting that

$$\mathscr{C}(\delta_x P_t, \delta_y P_t) \ni \pi_t := (\delta_x P_t) \times (\delta_y P_t) \le \|P_t\|^2_{L^1(\mu)\to L^\infty(\mu)}(\mu\times\mu),$$

we obtain

$$W_\Phi(\delta_x P_t, \delta_y P_t) \le \|\rho\|_{L^\Phi(\pi_t)} \le G_\Phi(2t), \quad t>0.$$

Combining this with (5.1) and the semigroup property, we arrive at

$$W_\Phi(\delta_x P_t, \delta_y P_t) \le \frac{c_0}{\Phi^{-1}(1)}\mathrm{e}^{-\lambda t/2} W_\Phi(\delta_x P_{t/2}, \delta_y P_{t/2}) \le \frac{c_0}{\Phi^{-1}(1)}\mathrm{e}^{-\lambda t/2} G_\Phi(t).$$

This together with (5.1) implies (2.6) for some constants $c, \lambda > 0$. Moreover, (2.8) follows from (2.6) according to Proposition 3.1.

(2) By Proposition 3.1, (2.9) implies

$$|\nabla P_t f| \le c\mathrm{e}^{-\lambda t} P_t|\nabla f|, \quad t\ge 0, f\in C_b^1(M).$$

Then using the standard semigroup calculation of Bakry-Emery, this implies

$$\begin{aligned} &P_t(f^2\log f^2) - (P_t f^2)\log P_t f^2 = \int_0^t \frac{\mathrm{d}}{\mathrm{d}s} P_s\big\{(P_{t-s}f^2)\log P_{t-s}f^2\big\}\mathrm{d}s \\ &= \int_0^t P_s\Big(\frac{|\nabla P_{t-s}f^2|^2}{P_{t-s}f^2}\Big)\mathrm{d}s \le 4c^2\int_0^t \mathrm{e}^{-2\lambda(t-s)} P_s\Big(\frac{(P_{t-s}\{f|\nabla f|\})^2}{P_{t-s}f^2}\Big)\mathrm{d}s \\ &\le 4c^2\int_0^t \mathrm{e}^{-2\lambda(t-s)}(P_t|\nabla f|^2)\mathrm{d}s = \frac{2c^2(1-\mathrm{e}^{-2\lambda t})}{\lambda} P_t|\nabla f|^2, \quad t\ge 0. \end{aligned}$$

Since $\lim_{t\to\infty} P_t g = \mu(g)$ for $g\in\mathscr{B}_b(M)$ due to the ergodicity, by letting $t\to\infty$ we prove the log-Sobolev inequality for (3.4) for $C = \frac{2c^2}{\lambda}$. Indeed, by the local Poincaré inequality and [24, Theorem 3.1], the weak Poincaré inequality

$$\mu(f^2) \le \alpha(r)\mu(|\nabla f|^2) + r\|f\|_\infty^2, \quad r>0, \mu(f)=0$$

holds for some $\alpha: (0,\infty)\to(0,\infty)$. By [24, Theorem 2.1], this implies

$$\lim_{t\to 0}\sup_{\|g\|_\infty\le 1}\mu(|P_t g - \mu(g)|^2) = 0.$$

□

Proof of Corollary 2.4. We first observe that the proof of [37, Theorem 4.2] works also for the non-symmetric case with ∇Z in place of Hess_V, so that for some constant $c_1 > 0$,

$$\|P_t\|_{L^2(\mu)\to L^\infty(\mu)} \le \exp\Big[c_1 + \frac{c_1}{t}\Big(1 + \Lambda_1^{-1}(c_1 t^{-1}) + \Lambda_2^{-1}(c_1^{-1}t)\Big)\Big], \quad t>0. \tag{5.2}$$

Since in the symmetric case we have $\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} \le \|P_{t/2}\|^2_{L^2(\mu)\to L^\infty(\mu)}$, the first assertion follows immediately from Theorem 2.3.

As for the non-symmetric case, since

$$\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} \le \|P_{t/2}\|_{L^1(\mu)\to L^2(\mu)} \|P_{t/2}\|_{L^2(\mu)\to L^\infty(\mu)},$$

by Theorem 2.3 and (5.2) it suffices to prove

$$\|P_t\|_{L^1(\mu)\to L^2(\mu)} \le \exp\big[c' + c' H(c' t^{-1})\big], \quad t>0 \tag{5.3}$$

for some constant $c'>0$. According to [25, Theorem 2.1], (5.2) implies the super log-Sobolev inequality (3.3) for

$$\beta(r) := c + \frac{c}{r}\Big\{1 + \Lambda_1^{-1}(cr^{-1}) + \Lambda_2^{-1}(c^{-1}r)\Big\}, \quad r>0$$

for some (possibly different) constant $c>0$. Then Proposition 3.3 with $p=1, q=2$ and $\gamma(r) := \frac{trh(r-1)}{(r-1)\int_0^1 s^{-1}h(s)\mathrm{d}s}$ implies (5.3).

□

6 Proofs of Theorems 2.5-2.6

Proof of Theorems 2.5. Let $X_t(x)$ solve (2.17) with initial point x. By Itô's formula and condition (2.18) we obtain

$$\begin{aligned}
&\mathrm{d}|X_t(x)-X_t(y)|^p\\
&\le \mathrm{d}M_t + p|X_t(x)-X_t(y)|^{p-2}\bigg\{\frac{(p-2)|(\sigma(X_t(x))-\sigma(X_t(y))^*(X_t(x)-X_t(y))|^2}{|X_t(x)-X_t(y)|^2}\\
&\quad + \|\sigma(X_t(x))-\sigma(X_t(y))\|_{HS}^2 + \langle b(X_t(x)-b(X_t(y)), X_t(x)-X_t(y)\rangle\bigg\}\mathrm{d}t\\
&\le \mathrm{d}M_t - pK_p|X_t(x)-X_t(y)|^p\mathrm{d}t
\end{aligned}$$

for some martingale M_t. This implies

$$\mathbb{E}|X_t(x)-X_t(y)|^p \le \mathrm{e}^{-pK_pt}|x-y|^p, \quad t\ge 0, x,y\in\mathbb{R}^d,$$

and thus,

$$\begin{aligned}
|\nabla P_t f(x)| &\le \limsup_{y\to x}\mathbb{E}\Big(\frac{|f(X_t(x))-f(X_t(y))|}{|X_t(x)-X_t(y)|}\cdot\frac{|X_t(x)-X_t(y)|}{|x-y|}\Big)\\
&\le \mathrm{e}^{-K_pt}(P_t|\nabla f|^{\frac{p}{p-1}})^{\frac{p-1}{p}}.
\end{aligned} \tag{6.1}$$

Then the desired assertion follows from Proposition 3.1. □

Proof of Theorem 2.6. (1) We reformulate (2.17) as

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sqrt{2}\big(\sigma_0(X_t)\mathrm{d}B_t' + \lambda_0\mathrm{d}B_t''\big), \tag{6.2}$$

where B'_t and B''_t are independent d-dimensional Brownian motions. For any $x \neq y$, let X_t solve this SDE with $X_0 = x$, and let Y_t solve the following coupled SDE with $Y_0 = y$:

$$\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + \sqrt{2}\,\sigma_0(Y_t)\mathrm{d}B'_t + \lambda_0\sqrt{2}\left(\mathrm{d}B''_t - 2\frac{\langle X_t - Y_t, \mathrm{d}B''_t\rangle(X_t - Y_t)}{|X_t - Y_t|^2}\right).$$

That is, under the flat metric we have made coupling by reflection for B''_t and coupling by parallel displacement for B'_t. Obviously, the coupled SDE has a unique solution up to the coupling time

$$T_{x,y} := \inf\{t \geq 0 : X_t = Y_t\}.$$

We set $Y_t = X_t$ for $t \geq T_{x,y}$ as usual. Then by (2.19) and Itô's formula, we obtain

$$\mathrm{d}|X_t - Y_t| \leq \mathrm{d}M_t + \big\{(K_1 + K_2)1_{\{|X_t - Y_t| \leq r_0\}} - K_2\big\}|X_t - Y_t|\mathrm{d}t, \quad t \leq T_{x,y} \tag{6.3}$$

for

$$\mathrm{d}M_t := \frac{\sqrt{2}\langle 2\lambda_0\mathrm{d}B''_t + (\sigma_0(X_t) - \sigma_0(Y_t))\mathrm{d}B'_t, X_t - Y_t\rangle}{|X_t - Y_t|}$$

being a martingale with

$$\mathrm{d}\langle M\rangle_t \geq 8\lambda_0^2\mathrm{d}t. \tag{6.4}$$

By repeating the argument leading to (4.3), it is easy see that (6.3) and (6.4) imply

$$\mathbb{E}|X_t - Y_t| \leq c\mathrm{e}^{-\lambda t}|x - y|, \quad t \geq 0$$

for some constants $c, \lambda > 0$ independent of x, y. Therefore,

$$|\nabla P_t f| \leq c\mathrm{e}^{-\lambda t}\|\nabla f\|_\infty, \quad t \geq 0, f \in C_b^1(\mathbb{R}^d),$$

so that the first assertion follows from Proposition 3.1.

(2) According to [40, Theorem 1.1], $\sigma\sigma^* \geq \lambda_0^2 I$ and (2.21) imply the log-Harnack inequality

$$P_t(\log f)(x) \leq \log P_t f(y) + \frac{c_1|x - y|^2}{1 - \mathrm{e}^{-c_2 t}}, \quad t \geq 0, x, y \in \mathbb{R}^d, 0 \leq f \in \mathscr{B}_b(\mathbb{R}^d)$$

for some constants $c_1, c_2 > 0$. Next, condition (2.21) implies

$$|\nabla P_t f|^2 \leq \mathrm{e}^{2Kt}P_t|\nabla f|^2,$$

see e.g. [**?**]RW10. Combining these with the log-Sobolev inequality, we prove the second assertion as in (c) in the proof of Theorem 2.1.

(3) According to the proof of Theorem 2.5, the condition (2.18) implies the gradient estimate (6.1). Next, by Proposition 3.4, the ultracontractivity and (6.1) imply

$$|\nabla P_t f| \leq c(p)\mathrm{e}^{-\lambda t}(P_t|\nabla f|^{\frac{p}{p-1}})^{\frac{p-1}{p}}, \quad t \geq 0, f \in C_b^1(\mathbb{R}^d)$$

for some $c(p) > 0$ and $\lambda > 0$ independent of p. Then the proof if finished by Proposition 3.1. □

Acknowledgement. The author would like to thank Jian Wang and the referee for helpful comments and corrections.

References

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

[2] S. Aida, D. Stroock, *Moment estimates derived from Poincaré and logarithmic Sobolev inequalities,* Math. Res. Lett. 1(1994), 75–86.

[3] M. Arnaudon, A. Thalmaier, F.-Y. Wang, *Harnack inequality and heat kernel estimate on manifolds with curvature unbounded below,* Bull. Sci. Math. 130(2006), 223–233.

[4] D. Bakry, M. Emery, *Hypercontractivité de semi-groupes de diffusion*, C. R. Acad. Sci. Paris. Sér. I Math. 299(1984), 775–778.

[5] D. Bakry, I. Gentil, M. Ledoux, *Hypercontractivity of Hamilton-Jacobi equations,* J. Math. Pures Appl. 80(2001), 669–696.

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

[7] D. Bakry, I. Gentil, M. Ledoux, *On Harnack inqualities and optimal transportation,* Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) XIV(2015), 705–727.

[8] M.-F. Chen, F.-Y. Wang, *Estimates of logarithmic Sobolev constant: an improvement of Bakry-Emery criterion,* J. Funct. Anal. 144(1997), 287–300.

[9] M.-F. Chen, F.-Y. Wang, *Estimation of spectral gap for elliptic operators,* Trans. Amer. Math. Soc. 349(1997), 1239–1267.

[10] E. B. Davies, B. Simons, *Ultracontractivity and heat kernel for Schrördinger operators and Dirichlet Laplacians,* J. Funct. Anal. 59(1984), 335–395.

[11] A. Eberle, *Reflection coupling and Warsserstein contractivity without convecity,* C. R. Acad. Sci. Paris. Sér. I Math. 349(2011), 1101–1104.

[12] A. Eberle, *Reflection couplings and contraction rates for diffusions,* to appear in Probab. Thero. Relat. Fields.

[13] M. Emery, *Stochastic Calculus in Manifolds,* Springer-Verlag, Berlin, 1989, with an appendix by P.-A. Meyer.

[14] L. Gross, *Logarithmic Sobolev inequalities and contractivity properties of semigroups*, Lecture Notes in Math. 1563, Springer-Verlag, 1993.

[15] W. S. Kendall, *Nonnegative Ricci curvature and the Brownian coupling property,* Stochastics 19(1986), 111–129.

[16] K. Kuwada, *Duality on gradient estimates and Wasserstein controls,* J. Funct. Anal. 258 (2010), 3758–3774.

[17] K. Kuwada, *Gradient estimate for Markov kernels, Wasserstein control and Hopf-Lax formula,* In "Potential Theory and Its Related Fields", 61–80, RIMS Kôkyûroku Bessatsu, B43, Res. Inst. Math. Sci. (RIMS), Kyoto, 2013.

[18] D. Luo, J. Wang, *Exponential convergence in L^p-Warsserstein distance for diffusion processes without uniform dissipative drift,* to appear in Math. Nachr.

[19] L. Miclo, *On hyperboundedness and spectrum of Markov operators,* Inven. Math. 200(2015), 311–343.

[20] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation,* Comm. Part. Diff. Equat. 26(2001), 101–174.

[21] F. Otto, C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality,* J. Funct. Anal. 173(2000), 361–400.

[22] E. Priola, F.-Y. Wang, *Gradient estimates for diffusion semigroups with singular coefficients,* J. Funct. Anal. 236(2006), 244–264.

[23] M.-K. von Renesse, K.-T. Sturm, *Transport inequalities, gradient estimates, entropy, and Ricci curvature,* Comm. Pure Appl. Math. 58(2005), 923–940.

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

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

[26] M. Röckner, F.-Y. Wang, *Spectrum for a class of (nonsymmetric) diffusion operators,* Bull. London Math. Soc. 36(2004), 95–104.

[27] M. Rockner, F.-Y. Wang, *Log-Harnack Inequality for Stochastic differential equations in Hilbert spaces and its consequences,* Infinite Dimensional Analysis, Quantum Probability and Related Topics 13(2010), 27–37.

[28] J. Shao, *Hamilton-Jacobi semi-groups in infinite dimensional spaces,* Bull. Sci. math. 130(2006), 720–738.

[29] W. Stannat, *(Nonsymmetric) Dirichlet operators on L^1: Existence, uniqueness and associated Markov processes,* Ann. Sc. Norm. Super. Pisa Cl. Sci. 28(1999) 99–140.

[30] C. Villani, *Topics in Optimal Transportation,* Amer. Math. Soc. 2003.

[31] C. Villani, *Optimal Transpot, Old and New,* Springer, 2009.

[32] F.-Y. Wang, *Application of coupling method to the Neumann eigenvalue problem,* Probab. Theory Relat. Fields 98(1994), 299–306.

[33] F.-Y. Wang, *Logarithmic Sobolev inequalities on noncompact Riemannian manifolds,* Probab. Theory Relat. Fields 109(1997), 417–424.

[34] F.-Y. Wang, *Functional inequalities for empty essential spectrum,* J. Funct. Anal. 170(2000), 219–245.

[35] F.-Y. Wang, *Probability distance inequalities on Riemannian manifolds and path spaces,* J. Funct. Anal. 206(2004), 167–190.

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

[37] F.-Y.Wang, *Log-Sobolev inequalities: different roles of Ric and Hess,* Ann. Probab. 37(2009), 1587–1604.

[38] F.-Y.Wang, *Log-Sobolev inequality on non-convex manifolds,* Adv. Math. 222(2009), 1503–1520.

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

[40] F.-Y. Wang, *Harnack inequality for SDE with multiplicative noise and extension to Neumann semigroup on nonconvex manifolds,* Ann. Probab. 39(2011), 1449–1467.

[41] F.-Y.Wang, *Criteria on spectral gap of Markov operators,* J. Funct. Anal. 266(2014), 2137–2152.

[42] F.-Y. Wang, *Analysis for Diffusion Processes on Riamnnian Manifolds,* Springer, 2014.

[43] F.-Y. Wang, *Asymptotic couplings by reflection and applications for nonlinear monotone SPDEs,* Nonlinear Anal. 117(2015), 169–188.