Monotonicity of time-dependent transportation costs and coupling by reflection

Kazumasa Kuwada[*] and Karl-Theodor Sturm[†]

Abstract

Based on a study of the coupling by reflection of diffusion processes, a new monotonicity in time of a time-dependent transportation cost between heat distribution is shown under Bakry-Émery's curvature-dimension condition on a Riemannian manifold. The cost function comes from the total variation between heat distributions on spaceforms. As a corollary, we obtain a comparison theorem for the total variation between heat distributions. In addition, we show that our monotonicity is stable under the Gromov-Hausdorff convergence of the underlying space under a uniform curvature-dimension and diameter bound.

Keywords: transportation cost, coupling by reflection, diffusion process, curvature-dimension condition, total variation

Mathematics Subject Classification (2010): 58J65, 53C21, 60H30, 60J60, 58J35

1 INTRODUCTION

Analysis of the heat equation on manifolds or metric measure spaces is one of the central issues in the literature. Several topics such as analysis of partial differential equations, differential geometry and probability theory are interacting with each other there. As one of remarkable consequences of such an interaction, many different characterizations of the presence of lower Ricci curvature bound by means of the heat semigroup or the Brownian motion are revealed in [36]. Among those studies, recent developments in the theory of optimal transport enable us to interpret the heat distribution as a gradient curve of the relative entropy in the space of probability measures (see [3,35], for instance) along Otto's heuristic idea in [28]. This viewpoint provides a quite natural understanding of the fact that the presence of lower Ricci curvature bound implies a contraction property of heat distributions in Wasserstein distance. Significantly, this argument can bring a piece of implications between equivalent notions in [36] mentioned above. As its probabilistic counterpart, we can show the contraction by means of constructing a coupling by parallel

---

[*]Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo 112-8610, Japan, *e-mail*: `kuwada.kazumasa@ocha.ac.jp`, *tel*: +81-3-5978-5300, *fax*: +81-3-5978-5295

[†]Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany, *e-mail*: `sturm@uni-bonn.de`

transport of Brownian motions. On the other hand, there is another kind of coupling, called the coupling by reflection or the Kendall-Cranston coupling, which is also well-studied in connection with the Riemannian geometry of the underlying space. The purpose of this article is to study the coupling by reflection by formulating it in terms of the theory of optimal transport.

To state our result, we introduce the notion of transportation cost. Given a function $c : M \times M \to \mathbb{R}$ on a state space M, a transportation cost $\mathcal{T}_c(\mu, \nu)$ between two probability measures μ and ν on M is defined as follows:

$$\mathcal{T}_c(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{M \times M} c \, d\pi,$$

where $\Pi(\mu, \nu)$ is the set of all couplings of μ and ν, namely, a probability measure on $M \times M$ whose marginal distributions are μ and ν respectively. The simplest case of our result is as follows:

Theorem 1.1 *Let M be a complete Riemannian manifold with nonnegative Ricci curvature with the Riemannian distance d. Let us define $\varphi_t(a)$ for $t, a \geq 0$ by*

$$\varphi_t(a) := \begin{cases} \dfrac{1}{\sqrt{4\pi t}} \displaystyle\int_{\mathbb{R}} \left| \exp\left(-\frac{1}{4t}\left(x - \frac{a}{2} \right)^2 \right) - \exp\left(-\frac{1}{4t}\left(x + \frac{a}{2} \right)^2 \right) \right| dx, & t > 0, \\ 1_{(0, \infty)}(a), & t = 0. \end{cases}$$

Then, for $t > 0$ and two heat distributions $\mu_s^{(i)}$ ($i = 1, 2$) generated by the Laplace-Beltrami operator, we have

$$\mathcal{T}_{\varphi_{t-s_2}(d)}\big(\mu_{s_2}^{(1)}, \mu_{s_2}^{(2)}\big) \leq \mathcal{T}_{\varphi_{t-s_1}(d)}\big(\mu_{s_1}^{(1)}, \mu_{s_1}^{(2)}\big) \tag{1.1}$$

for any $0 \leq s_1 \leq s_2 \leq t$.

In the full statement in Theorem 2.3, the same result as (1.1) holds for distributions of a diffusion process with an upper bound of dimension and a lower Ricci curvature bound in the sense of Bakry and Émery with an appropriate choice of $\varphi_t(a)$, which is more complicated. Alternatively, (1.1) could be formulated as a non-expansion result of Lipschitz constants with respect to time-dependent metrics $\varphi_t(d)$; see Theorem 6.1. This allows us to interpret φ_t as the profile of the "worst case" initial data corresponding to $\varphi_0(d)$. Given $K \in \mathbb{R}$, the L^p-contraction in Wasserstein distance mentioned above means

$$e^{pKt} \mathcal{T}_{d^p}(\mu_t^{(1)}, \mu_t^{(2)}) \leq e^{pKs} \mathcal{T}_{d^p}(\mu_s^{(1)}, \mu_s^{(2)}) \tag{1.2}$$

for $t > s \geq 0$ and any heat distributions $\mu_s^{(i)}$ ($i = 1, 2$). It holds with $K = 0$ under the assumption in Theorem 1.1 and hence Theorem 1.1 can be regarded as an analogue of it. Indeed, the only difference between them is the choice of the cost function. The counterpart of Theorem 6.1 for (1.2) is the equivalence with Bakry-Émery's L^q-gradient estimates (see [21]).

Let us review the history on the study of coupling by reflection, to explain a meaning and significance of Theorem 1.1. We call $(X_1(t), X_2(t))$ a coupling of a diffusion process $X(t)$ on a state space M if (X_1, X_2) is a stochastic process on $M \times M$ and each X_i behaves

as X on M for $i = 1, 2$. The coupling by reflection on a Euclidean space, or the mirror coupling, of Brownian motions introduced in [24] is given by the global reflection with respect to the hyperplane bisecting the line segment joining initial positions. With the help of Riemannian geometry, a coupling by reflection of Brownian motions on a Riemannian manifold is constructed by Kendall [15] and Cranston [6], by making a coupling of their infinitesimal motions. In many applications, it is nice to suppose that they will coalesce after the coupling time, namely, the time when they meet. As a matter of fact, the coupling by reflection of Brownian motion on a Euclidean space, or more generally the one on a Riemannian manifold with nonnegative Ricci curvature, can meet in a finite time almost surely regardless of the dimension of the space. It is a great contrast with the case of observing two independent Brownian motions. Under a nice condition, for example, the presence of curvature bounds on the state space, this kind of coupling has provided several applications e.g. in estimating the rate of convergence to equilibrium, functional inequalities involving heat semigroups, (non-)existence of harmonic maps (see [16] and references therein, for instance). As a simple example, the coupling by reflection under nonnegative Ricci curvature easily implies the Liouville property, that is, non-existence of nonconstant bounded harmonic functions. In many of those applications, we only need to know the existence of *a coupling* π of two distributions of $X(t)$ having a good estimate of $\pi(\{(x, x) \mid x \in M\})$, which can be provided by comparing the transportation cost in Theorem 1.1 at $s = t$ with that at $s = 0$ since $\varphi_0 = 1_{(0,\infty)}$. Thus the monotonicity of transportation cost in Theorem 1.1 works sufficiently well in applications.

Recently, such topics as mentioned above is extensively studied on more singular metric measure spaces than Riemannian manifolds under a new, synthetic notion of curvature bounds (see e.g. [5, 25, 31, 32]). However, the traditional way of studying a coupling by reflection of Brownian motions is based on the theory of stochastic differential equations and hence there are many difficulties to extend the original argument directly into analysis on such singular spaces. In contrast to such an approach, the statement of Theorem 1.1 completely makes sense even on singular spaces once we introduced the notion of heat distributions on it. Thus there seems to be some possibility to extend it to such cases though our framework in Theorem 1.1 or Theorem 2.3 is still on a Riemannian manifold. Actually, we obtain a partial result in this direction by showing that the monotonicity of the transportation cost stated in Theorem 1.1(or Theorem 2.3) is stable under the Gromov-Hausdorff convergence of underlying spaces (see Theorem 7.2) with a uniform curvature-dimension and diameter bound.

One might wonder why the cost function in Theorem 1.1 appears. It is based on the fact that $\mathcal{T}_{\varphi_{t-s}(d)}\big(\mu_s^{(1)}, \mu_s^{(2)}\big)$ is a constant function in s and the infimum in the definition of the transportation cost is attained by the coupling by reflection when M is a Euclidean space [12, 19]. Our choice of the cost function is natural and sharp in this sense. Our argument in the proof of Theorem 1.1 is based on a comparison between the distance process for the coupling by reflection on M and the one on a Euclidean space. And then the sharpness on a Euclidean space plays a prominent role when we deal with the comparison process. It should be remarked that, when M is a Euclidean space, the cost function $\varphi_t(d(x, y))$ coincides with the total variation between two heat distributions at time t with initial distributions δ_x and δ_y respectively. This fact is closely related to the maximality of the coupling by reflection of Brownian motions on a Euclidean

space (see [12, 19], for instance). Indeed, by choosing the cost function as the total variation between heat distributions, exactly the same constancy holds on a spaceform since the coupling by reflection of Brownian motions is also maximal. As we will see, this characterization of our cost function leads to the following comparison theorem for the total variation between heat distributions. Let us denote the total variation norm by $\| \cdot \|_{\mathrm{TV}}$.

Corollary 1.2 *Let M be a complete Riemannian manifold whose dimension is less than or equal to $N \in \mathbb{N}$ and whose Ricci curvature is greater than or equal to $K \in \mathbb{R}$. Then, for two heat distributions $\mu_s^{(1)}$, $\mu_s^{(2)}$ with $\mu_0^{(i)} = \delta_{x_i}$ for some $x_i \in M$ $(i = 1, 2)$,*

$$\left\| \mu_t^{(1)} - \mu_t^{(2)} \right\|_{\mathrm{TV}} \leq \int_{\mathbb{M}_{K,N}} \left| \tilde{p}_t^{K,N}\left(\tilde{x}_1, y\right) - \tilde{p}_t^{K,N}\left(\tilde{x}_2, y\right) \right| \mathrm{vol}_{\mathbb{M}_{K,N}}(dy), \qquad (1.3)$$

where $\tilde{p}_t^{K,N}(x, y)$ is the heat kernel on the N-dimensional spaceform $\mathbb{M}_{K,N}$ of constant sectional curvature $K/(N-1)$ and $(\tilde{x}_1, \tilde{x}_2)$ is any pair of points in $\mathbb{M}_{K,N}$ satisfying $d(\tilde{x}_1, \tilde{x}_2) = d(x_1, x_2)$.

This is a special case of Corollary 2.4 below. It seems to be natural that we can measure the total variation as a result of a study of the coupling by reflection since the coupling by reflection has been strongly related with estimates involving the coupling time, which yields an estimate of the total variation between distributions via the coupling inequality (see [23], for instance).

Note that, to the best of the authors' knowledge, an estimate of type (1.1) with the use of a time-dependent cost function is studied first in [30, Example 4.6]. While it is discussed only on $\mathbb{R}^m$, it includes Lévy processes as an example. Also note that our cost function in Theorem 1.1 (or Theorem 2.3) is a concave function of the distance function. It corresponds to the observation in [9], which says that the optimal transport map for a concave cost function reverses the orientation. Indeed, the reflection map used in constructing the coupling by reflection does so. Finally, we remark that there is a recent related result in [7], which studies a behavior of the transportation cost of a concave cost function in connection with the coupling by reflection of a diffusion process on $\mathbb{R}^m$.

The organization of the paper is as follows. In the next section, we will give a more precise statement of our main results. For proving them, we will study the coupling by reflection in section 3. There we will follow the argument in [17, 20] in which we construct the coupling by reflection via an approximation of diffusion processes by geodesic random walks. It might be possible to follow an alternative approach in [38]. Section 4 is devoted to show several regularity properties of the function $\varphi_t^{K,N}(a)$ introduced in section 2 to describe the main theorem. Some explicit expressions of $\varphi_t^{K,N}(a)$ as well as asymptotic behavior as $t \to 0$ or $t \to \infty$ are also given there. Some results in this section might be of independent interest. The proof of our main theorem is given in section 5. Though most part will follow from the result in section 3, we need an additional argument with the aid of results in section 4 to complete the proof. We also study new monotonicity formulae for time-independent transportation costs (Corollary 5.3) as a consequences of the main theorem and results in section 4. In section 6, we give a short remark on gradient estimates for the diffusion semigroup corresponding to our main theorem. Though a similar gradient

estimate is already obtained in [6] in the same spirit, what we obtained is sharper in many respect. The stability of our main result under the Gromov-Hausdorff convergence is discussed in section 7. It ensures that all the results obtained before this section will be inherited to the measured Gromov-Hausdorff limit under a uniform curvature-dimension and diameter bound. In section 8, we will give a brief comment on the extension of results in sections 2–6 to the time-dependent metric case. Note that the assumption there is satisfied if the metric evolves according to the backward Ricci flow.

2 Framework and the main result

Let (M, g) be a complete m-dimensional Riemannian manifold with $m \geq 2$. Let d stand for the Riemannian distance on M. Let Z be a C^1-vector field and we denote the generator of the form $\Delta + Z$ by $\mathcal{L}$, where Δ is the Laplace-Beltrami operator with respect to g. Let $((X(t))_{t \in [0,\infty)}, (\mathbb{P}_x)_{x \in M})$ be a diffusion process associated with $\mathcal{L}$. Let $(\nabla Z)^\flat$ be the symmetrization of ∇Z, i.e., a $(0,2)$-tensor given by

$$(\nabla Z)^\flat(X, Y) := \frac{1}{2} \left(\langle \nabla_X Z, Y \rangle + \langle \nabla_Y Z, X \rangle \right).$$

Our basic assumption is the following condition involving the upper dimension bound and lower Ricci curvature bound formulated in terms of $\mathcal{L}$:

Assumption 1 *Given $K \in \mathbb{R}$ and $N \in [m, \infty]$, the following holds:*

$$\mathrm{Ric} - (\nabla Z)^\flat - \frac{1}{N-m} Z \otimes Z \geq Kg.$$

Here we regard the third term in the left hand side is 0 when $N = \infty$, and $N = m$ is permitted only when $Z \equiv 0$.

It is well known that Assumption 1 is equivalent to the following curvature-dimension condition of Bakry and Émery (see e.g. [4, 22]):

$$\frac{1}{2} \left(\mathcal{L} \langle \nabla f, \nabla f \rangle - 2 \langle \nabla f, \nabla \mathcal{L} f \rangle \right) \geq K \langle \nabla f, \nabla f \rangle + \frac{1}{N} (\mathcal{L} f)^2.$$

This condition is equivalent to $\dim M \leq N$ and $\mathrm{Ric} \geq K$ when $Z \equiv 0$.

In order to state our main theorems, we introduce the notion of comparison process and associated transportation costs. Let $K \in \mathbb{R}$ and $N \in [2, \infty]$. Set $\bar{R} = \bar{R}_{K,N}$ by

$$\bar{R}_{K,N} := \begin{cases} \sqrt{\dfrac{N-1}{K}} \pi & \text{if } K > 0 \text{ and } N < \infty, \\ \infty & \text{otherwise.} \end{cases}$$

5

We define s_K and c_K as a usual comparison function as follows:

$$s_K(\theta) := \begin{cases} \dfrac{1}{\sqrt{K}}\sin(\sqrt{K}\theta) & K > 0, \\ \theta & K = 0, \\ \dfrac{1}{\sqrt{-K}}\sinh(\sqrt{-K}\theta) & K < 0, \end{cases}$$

$$c_K(\theta) := \begin{cases} \cos(\sqrt{K}\theta) & K > 0, \\ 1 & K = 0, \\ \cosh(\sqrt{-K}\theta) & K < 0 \end{cases}$$

and $t_K := s_K/c_K$. Let $\Psi = \Psi_{K,N} : (-\bar{R}, \bar{R}) \to \mathbb{R}$ be given by

$$\Psi_{K,N}(u) := \begin{cases} -2K t_{K/(N-1)}\left(\dfrac{u}{2}\right) & \text{if } N < \infty, \\ -Ku & \text{otherwise.} \end{cases}$$

Let us define a diffusion process $\rho(t) = \rho_{K,N,a}(t)$, $t \geq 0$ on $\overline{(-\bar{R}, \bar{R})} \subset \mathbb{R}$ as a solution to the following stochastic differential equation:

$$d\rho_{K,N,a}(t) = 2\sqrt{2}d\beta(t) + \Psi(\rho_{K,N,a}(t))dt, \tag{2.1}$$
$$\rho_{K,N,a}(0) = a.$$

Note that, when $\bar{R} < \infty$, both $-\bar{R}$ and $\bar{R}$ are entrance boundary for $\rho(t)$. For $t \geq 0$, let us define $\varphi_t^{K,N} : \overline{[0, \bar{R})} \to [0,1]$ by

$$\varphi_t^{K,N}(a) := \mathbb{P}\left[\inf_{0 \leq s \leq t} \rho_{K,N,a}(s) > 0\right].$$

Remark 2.1   (i) *The process $\rho_{K,N,a}$ comes from the coupling by reflection on the spaceform. Actually, when $N \in \mathbb{N}$, a simple computation implies that the distance process $d(\mathbf{X}(t))$ for the coupling by reflection $\mathbf{X}(t) = (X_1(t), X_2(t))$ of Brownian motions on the spaceform $\mathbb{M}_{N,K}$ solves the stochastic differential equation defining $\rho_{K,N,a}$ with $a = d(\mathbf{X}(0))$.*

(ii) *Since $-\rho_{K,N,a}$ has the same law as $\rho_{K,N,-a}$, the reflection map $x \mapsto -x$ on $\overline{(-\bar{R}, \bar{R})}$ provides a so-called 'reflection structure' in [19]. It is shown in [19] that the mirror coupling for $\rho_{K,N,a}$ and $\rho_{K,N,-a}$ is maximal in such a case. As a result, we have*

$$\varphi_t^{K,N}(a) = \left\| \mathbb{P} \circ (\rho_{K,N,a}(t)/2)^{-1} - \mathbb{P} \circ (\rho_{K,N,-a}(t)/2)^{-1} \right\|_{\mathrm{TV}}.$$

In particular, we can easily verify that $\varphi_t^{0,N}$ equals to φ_t in Theorem 1.1. Moreover, $\mathbb{E}[\varphi_{t-s}(|\rho(s)|)]$ is a constant function in $s \in [0,t]$ by [19, Lemma 3.4].

(iii) *When $N \in \mathbb{N}$, the coupling $\mathbf{X}(t)$ by reflection of Brownian motions $B(t)$ on $\mathbb{M}_{K,N}$ is maximal by the same reasoning (see [19, Theorem 5.1 and Example 4.6]). Thus we have*

$$\varphi_t^{K,N}(a) = \left\| \mathbb{P}_{\tilde{x}_1} \circ B(t)^{-1} - \mathbb{P}_{\tilde{x}_2} \circ B(t)^{-1} \right\|_{\mathrm{TV}}$$

for any pair of points $(\tilde{x}_1, \tilde{x}_2)$ in $\mathbb{M}_{K,N}$ satisfying $d(\tilde{x}_1, \tilde{x}_2) = a$ and $\mathbb{E}[\varphi_{t-s}(d(\mathbf{X}(s)))]$ is a constant function in $s \in [0, t]$. In particular, the right hand side of (1.3) equals to $\varphi_t^{K,N}(d(x_1, x_2))$.

Now we are in turn to state our first main theorem as follows:

Theorem 2.2 *Suppose that Assumption 1 holds. Then, for any $x_1, x_2 \in M$, there exists a coupling $\mathbf{X}(t) = (X_1(t), X_2(t))_{t \geq 0}$ of $\mathcal{L}$-diffusion processes starting from (x_1, x_2) such that, for any $t > 0$ and $s \geq 0$,*

$$\mathbb{E}\left[\varphi_t^{K,N}(d(\mathbf{X}(s)))\right] \leq \varphi_{t+s}^{K,N} d(x_1, x_2).$$

Indeed, as we will see, a coupling $\mathbf{X}(t)$ appeared in Theorem 2.2 will be given as the coupling by reflection. Theorem 2.2 yields the following corresponding property described in terms of $\mathcal{T}_{\varphi_t(d)}$. This is our second main theorem:

Theorem 2.3 *Suppose that Assumption 1 holds. For $i = 1, 2$ and $\mu^{(i)} \in \mathcal{P}(M)$, let $\mu_t^{(i)}$ be the distribution of $X(t)$ with the initial distribution $\mu^{(i)}$. Then, for any $t > 0$, $\mathcal{T}_{\varphi_{t-s}(d)}(\mu_s^{(1)}, \mu_s^{(2)})$ is a nonincreasing function of $s \in [0, t]$. That is, for $0 \leq s_1 \leq s_2 \leq t$,*

$$\mathcal{T}_{\varphi_{t-s_2}(d)}(\mu_{s_2}^{(1)}, \mu_{s_2}^{(2)}) \leq \mathcal{T}_{\varphi_{t-s_1}(d)}(\mu_{s_1}^{(1)}, \mu_{s_1}^{(2)}). \tag{2.2}$$

As a result of Theorem 2.3, we can compare $\mathcal{T}_{\varphi_{t-s}(d)}(\mu_s^{(1)}, \mu_s^{(2)})$ at $s = t$ with the one at $s = 0$ to obtain an estimate of the total variation between distributions of the diffusion process $X(t)$. In particular, when $\mu_0^{(1)}$ and $\mu_0^{(2)}$ are Dirac measures, we obtain the following comparison theorem thanks to Remark 2.1 (ii):

Corollary 2.4 *Suppose that Assumption 1 holds. Then, for $x_1, x_2 \in M$ and $t > 0$,*

$$\left\|\mathbb{P}_{x_1} \circ X(t)^{-1} - \mathbb{P}_{x_2} \circ X(t)^{-1}\right\|_{\mathrm{TV}}$$
$$\leq \left\|\mathbb{P} \circ (\rho_{K,N,d(x_1,x_2)}(t)/2)^{-1} - \mathbb{P} \circ (\rho_{K,N,-d(x_1,x_2)}(t)/2)^{-1}\right\|_{\mathrm{TV}}.$$

When $N \in \mathbb{N}$, it immediately implies Corollary 1.2 by virtue of Remark 2.1 (iii).

Note that, by taking $t \to \infty$ in (2.2) after a suitable rescaling, we can obtain a similar monotonicity formula whose cost is independent of t. See Corollary 5.3 below. Especially, when $K < 0$, it does not seem to be known in the literature.

Remark 2.5 *When $K > 0$, it is shown in [18] that under Assumption 1 the Bonnet-Myers type diameter bound*

$$\mathrm{diam}(M) \leq \pi\sqrt{\frac{N-1}{K}}$$

holds. Moreover, the equality holds only when $N = m$, $Z \equiv 0$ and M is isometric to N-dimensional sphere of constant sectional curvature $K/(N-1)$. In the case of equality, the assertion in Theorem 2.2 is obvious by Remark 2.1 (iii) and hence we may assume $\mathrm{diam}(M) < \pi\sqrt{(N-1)/K}$ in the sequel.

3 Proof of Theorem 2.2

We will show that the coupling by reflection studied in [20] (cf. [17]) satisfies the assertion of Theorem 2.2 under Assumption 1. We begin with reviewing the construction of the coupling by reflection. Let $(\xi_n)_{n\in\mathbb{N}}$ be independent random variables all of which are uniformly distributed on the unit disk on $\mathbb{R}^m$. Let $(\gamma_{xy})_{x,y\in M}$ be a measurable family of unit-speed minimal geodesics defined on $[0, d(x, y)]$ such that γ_{xy} joins x and y. Without loss of generality, we may assume that γ_{xy} is symmetric, that is, $\gamma_{xy}(d(x, y) - s) = \gamma_{yx}(s)$ holds. Let us define $\tilde{m}_{xy} : T_x M \to T_x M$ by

$$\tilde{m}_{xy}v := v - 2\langle v, \dot{\gamma}_{xy}(0)\rangle\dot{\gamma}_{xy}(0).$$

This is a reflection with respect to a hyperplane which is perpendicular to $\dot{\gamma}_{xy}$. Let $/\!/_\gamma$ be the parallel transport along a curve γ. Let us define $m_{xy} : T_x M \to T_y M$ by $m_{xy} := /\!/_{\gamma_{xy}} \circ \tilde{m}_{xy}$. Clearly m_{xy} is an isometry. Set $D(M) := \{(x, x) \mid x \in M\}$. Let $\Phi : M \to \mathcal{O}(M)$ be a measurable section of the orthonormal frame bundle $\mathcal{O}(M)$ of M. Let us define two measurable maps $\Phi_i : M \times M \to \mathcal{O}(M)$ for $i = 1, 2$ by

$$\Phi_1(x, y) := \Phi(x),$$

$$\Phi_2(x, y) := \begin{cases} m_{xy}\Phi_1(x, y), & (x, y) \in M \times M \setminus D(M), \\ \Phi(x), & (x, y) \in D(M). \end{cases}$$

Take $x_1, x_2 \in M$. Let $t_n^\alpha := \alpha^2 n$ for $n \in \mathbb{N}_0$. By using Φ_i, we define a coupled geodesic random walk $\mathbf{X}^\alpha(t) = (X_1^\alpha(t), X_2^\alpha(t))$ with a scale parameter α by $X_i^\alpha(0) = x_i$ and, for $t \in [t_n^\alpha, t_{n+1}^\alpha]$,

$$\tilde{\xi}_{n+1}^i := \sqrt{2(m+2)}\Phi_i\left(\mathbf{X}^\alpha(t_n^\alpha)\right)\xi_{n+1},$$

$$X_i^\alpha(t) := \exp_{X_i^\alpha(t_n^\alpha)}\left(\frac{t - t_n^\alpha}{\alpha^2}\left(\alpha\tilde{\xi}_{n+1}^i + \alpha^2 Z\right)\right)$$

for $i = 1, 2$, where $\exp_x$ is the exponential map at x. Let us denote $C([0, \infty) \to M \times M)$ and $C([0, \infty) \to [-\bar{R}, \bar{R}])$ equipped with the topology of compact uniform convergence by $\mathscr{C}$ and $\mathscr{C}_1$ respectively.

In what follows, we assume Assumption 1. Then, by [17, Theorem 3.1] (also see references therein), $X_i^\alpha(t)$ converges in law in $C([0, \infty) \to M)$ to an $\mathcal{L}$-diffusion process starting from x_i for $i = 1, 2$ respectively. Thus $(\mathbf{X}^\alpha)_{\alpha>0}$ is tight and hence a subsequential limit $\mathbf{X}^{\alpha_k} \to \mathbf{X} = (X_1, X_2)$ in law in $\mathscr{C}$ exists. We fix such a subsequence $(\alpha_k)_{k\in\mathbb{N}}$. In the rest of this paper, we use the same symbol $\mathbf{X}^\alpha$ for the subsequence $\mathbf{X}^{\alpha_k}$ and the term "$\alpha \to 0$" always means the subsequential limit "$\alpha_k \to 0$". Let τ^* be the first hitting time to $D(M)$ of $\mathbf{X}$. Then we define *a coupling by reflection* $\mathbf{X}^* = (X_1^*, X_2^*)$ by

$$\mathbf{X}^*(t) := \begin{cases} \mathbf{X}(t) & \text{if } t < \tau^*, \\ (X_1(t), X_1(t)) & \text{if } t \geq \tau^*. \end{cases}$$

Since τ^* is a stopping time with respect to the filtration generated by $\mathbf{X}$, and X_i ($i = 1, 2$) is a solution to the martingale problem associated with the same filtration, $\mathbf{X}^*$ is again a coupling of $\mathcal{L}$-diffusion process.

Fix a reference point $o \in M$. For $R > 0$, let $\sigma_R : \mathscr{C}_1 \to [0, \infty]$ be given by $\sigma_R(w) :=$ $\inf \{t \in [0, \infty) \mid w(t) \geq R\}$. We define $\hat{\sigma}_R^i$ $(i = 1, 2)$ and $\hat{\sigma}_R$ by $\hat{\sigma}_R^i := \sigma_R(d(o, X_i^\alpha(\cdot)))$ and $\hat{\sigma}_R := \hat{\sigma}_R^1 \wedge \hat{\sigma}_R^2$. Proposition 3.4 in [17] says that

$$\lim_{R \to \infty} \limsup_{\alpha \to 0} \mathbb{P}[\hat{\sigma}_R < \infty] = 0 \tag{3.1}$$

holds.

We next review a difference inequality of $d(\mathbf{X}^\alpha(t))$. To describe it, we will introduce some notations. For simplicity of notations, let us denote $\gamma_{X_1^\alpha(t_n^\alpha)X_2^\alpha(t_n^\alpha)}$, $m_{X_1^\alpha(t_n^\alpha)X_2^\alpha(t_n^\alpha)}$ and $d(\mathbf{X}^\alpha(t_n^\alpha))$ by γ_n, m_n and $r^\alpha(n)$ respectively. Let $\tilde{\xi}_{n+1}^\perp(0)$ be the orthogonal projection of $\tilde{\xi}_{n+1}^1$ to the hyperplane being perpendicular to $\dot{\gamma}_n(0)$, that is, $2\tilde{\xi}_{n+1}^\perp(0) := (1 + m_n)\tilde{\xi}_{n+1}^1$. We denote a vector field along γ_n given by parallel transport of $\xi_{n+1}^\perp(0)$ by $(\xi_{n+1}^\perp(s))_{s \in [0, r^\alpha(n)]}$. Let us define a weight function $h_{n+1} = h_{n+1}^{K,N}$ on $[0, r^\alpha(n)]$ and a vector field $V_{n+1} = V_{n+1}^{K,N}$ along γ_n by

$$h_{n+1}^{K,N}(s) := \begin{cases} c_{K/(N-1)}\left(\dfrac{r^\alpha(n)}{2}\right)^{-1} c_{K/(N-1)}\left(\left(\dfrac{s - r^\alpha(n)}{2}\right)\right) & \text{if } N < \infty, \\ 1 & \text{if } N = \infty, \end{cases}$$
$$V_{n+1}^{K,N}(s) := h_{n+1}(s)\tilde{\xi}_{n+1}^\perp(s).$$

Recall that we are assuming $\mathrm{diam}(M) < \pi\sqrt{(N-1)/K}$ when $K > 0$ and $N < \infty$ (see Remark 2.5). Hence h_{n+1} is well-defined. For a smooth curve γ and vector fields V and W along γ, we denote the index form by $I_\gamma(V, W)$. When $V = W$, we use the symbol $I_\gamma(V)$ for $I_\gamma(V, W)$. Take $v \in \mathbb{R}^m$. Let us define λ_{n+1} and Λ_{n+1} by

$$\lambda_{n+1} := \begin{cases} 2\sqrt{2}\langle \tilde{\xi}_{n+1}^1(0), \dot{\gamma}_n(0)\rangle & \text{if } \mathbf{X}^\alpha(t_n^\alpha) \notin D(M), \\ 2\sqrt{2}\sqrt{m+2}\langle \xi_{n+1}, v\rangle & \text{otherwise}, \end{cases}$$

$$\Lambda_{n+1} := \left(\langle Z(t_n^\alpha), \dot{\gamma}_n(s)\rangle \big|_{s=0}^{r^\alpha(n)} + I_{\gamma_n}(V_{n+1}) \right) 1_{\{\mathbf{X}^\alpha(t_n^\alpha) \notin D(M)\}}.$$

For $\delta \geq 0$, let us define $\tau_\delta : \mathscr{C}_1 \to [0, \infty]$ by $\tau_\delta(w) := \inf \{t \geq 0 \mid w(t) \leq \delta\}$. We also define $\hat{\tau}_\delta$ by $\hat{\tau}_\delta := \tau_\delta(d(\mathbf{X}^a(\cdot)))$. In the sequel, we fix $\delta \in (0, 1)$ and $R > 1$. The first goal is to prove the following difference inequality for $r^\alpha(n)$:

Proposition 3.1 *For each $\varepsilon > 0$, there exists a family of events E_ε^α with $\lim_{\alpha \to 0} \mathbb{P}[E_\varepsilon^\alpha] = 1$ such that*
$$r^\alpha(n+1) \leq r^\alpha(n) + \alpha\lambda_{n+1} + \alpha^2\Psi(r^\alpha(n)) + \varepsilon\alpha^2$$

holds for $n \in \mathbb{N}$ with $t_n^\alpha < \hat{\tau}_\delta \wedge \hat{\sigma}_R$ on $(E_\varepsilon^\alpha)^c$ for sufficiently small α.

We will prove this assertion by a similar argument as in [17, 20]. Thus we only give a brief sketch of arguments. It consists of the following three lemmata. The following is shown in the same way as [20, Lemma 3] or [17, Lemma 4.4] by using the second variation formula of arclength with a careful treatment of singularities arising from the cutlocus.

Lemma 3.2 *For $n \in \mathbb{N}_0$, we have*

$$r^\alpha(n+1) \leq r^\alpha(n) + \alpha\lambda_{n+1} + \alpha^2\Lambda_{n+1} + o(\alpha^2) \tag{3.2}$$

when $n < \hat{\tau}_\delta \wedge \hat{\sigma}_R$ and α is sufficiently small. Moreover, we can control the error term $o(\alpha^2)$ uniformly in the position of $\mathbf{X}^\alpha$.

Set $\mathscr{F}_n := \sigma(\xi_1, \ldots, \xi_n)$ and $\bar{\Lambda}_{n+1} := \mathbb{E}[\Lambda_{n+1} | \mathscr{F}_n]$. For $\varepsilon > 0$ and $R > 0$, let us define an event $\tilde{E}^\alpha_\varepsilon$ by

$$\tilde{E}^\alpha_\varepsilon := \left\{ \sup_{t^\alpha_n \leq \hat{\sigma}_R} \sum_{j=1}^n (\Lambda_j - \bar{\Lambda}_j) \leq \frac{\varepsilon}{2\alpha^2} \right\}.$$

By following arguments in [20, Lemma 6] or [17, Lemma 4.5] which are based on the Doob submartingale inequality, we obtain the following.

Lemma 3.3 *For any $\varepsilon > 0$ and $R > 0$, $\mathbb{P}[\tilde{E}^\alpha_\varepsilon]$ tends to 1 as $\alpha \to 0$.*

Lemma 3.3 ensures to replace Λ_{n+1} in Lemma 3.2 with $\bar{\Lambda}_{n+1}$ with small errors on $(\tilde{E}^\alpha_\varepsilon)^c$. Thus, the proof of Proposition 3.1 will be completed with $E^\alpha_\varepsilon = \tilde{E}^\alpha_\varepsilon$ once we show the following:

Lemma 3.4 $\bar{\Lambda}_{n+1} \leq \Psi(r^\alpha(n))$.

Proof. Note that we have

$$\langle Z, \dot{\gamma}_n(s) \rangle \big|_{s=0}^{r^\alpha(n)} = h_{n+1}(s)^2 \langle Z, \dot{\gamma}_n(s) \rangle \big|_{s=0}^{r^\alpha(n)}$$

$$= \int_0^{r^\alpha(n)} \left(h_{n+1}(s)^2 (\nabla Z)^\flat(\dot{\gamma}_n(s), \dot{\gamma}_n(s)) \right.$$

$$\left. + 2h'_{n+1}(s)h_{n+1}(s)\langle Z, \dot{\gamma}_n(s) \rangle \right) ds. \tag{3.3}$$

By an easy computation, we obtain $\mathbb{E}[\xi_1] = 0$ and $\mathrm{Cov}(\sqrt{2(m+2)}\xi_1) = 2\mathrm{Id}$. Thus we have

$$I_{\gamma_n}(V_{n+1}) = \int_0^{r^\alpha(n)} \left((m-1)h'_{n+1}(s)^2 - \mathrm{Ric}(\dot{\gamma}_n(s), \dot{\gamma}_n(s))h_{n+1}(s)^2 \right) ds. \tag{3.4}$$

Combining (3.3) and (3.4) with the definition of $\bar{\Lambda}_n$, we obtain

$$\bar{\Lambda}_{n+1} := \left(\int_0^{r^\alpha(n)} \left(h_{n+1}(s)^2 (\nabla Z)^\flat(\dot{\gamma}_n(s), \dot{\gamma}_n(s)) + 2h'_{n+1}(s)h_{n+1}(s)\langle Z, \dot{\gamma}_n(s) \rangle \right. \right.$$

$$\left. \left. + (m-1)h'_{n+1}(s)^2 - \mathrm{Ric}(\dot{\gamma}_n(s), \dot{\gamma}_n(s))h_{n+1}(s)^2 \right) ds \right) 1_{\{\mathbf{X}^\alpha(t^\alpha_n) \notin D(M)\}}. \tag{3.5}$$

Thus, when $N = \infty$, the conclusion easily follows from Assumption 1. When $N = m$, $Z \equiv 0$ holds and Assumption 1 means $\mathrm{Ric} \geq K$. Thus an easy computation in (3.5) yields the conclusion. When $m < N < \infty$, the arithmetic geometric mean inequality implies

$$2h'_{n+1}(s)h_{n+1}(s)\langle Z, \dot{\gamma}_n(s) \rangle \leq (N-m)h'_{n+1}(s)^2 + \frac{1}{N-m}h_{n+1}(s)^2\langle Z, \dot{\gamma}_n(s) \rangle^2$$

$$= (N-m)h'_{n+1}(s)^2 + \frac{1}{N-m}h_{n+1}(s)^2 Z \otimes Z(\dot{\gamma}_n(s), \dot{\gamma}_n(s)). \tag{3.6}$$

By substituting (3.6) into (3.5), we obtain

$$\bar{\Lambda}_{n+1} \le \left(\int_0^{r^\alpha(n)} \left((N-1)h'_{n+1}(s)^2 \right. \right.$$

$$\left. \left. + h_{n+1}(s)^2 \left(\frac{1}{N-m} Z \otimes Z + (\nabla Z)^\flat - \mathrm{Ric} \right) (\dot{\gamma}_n(s), \dot{\gamma}_n(s)) \right) ds \right) 1_{\{\mathbf{X}^\alpha(t_n^\alpha) \notin D(M)\}}.$$

Hence Assumption 1 reduces the assertion to the same computation as in the case $N = m$.
□

Set $a := d(x_1, x_2)$. Let $\rho^\alpha_{K,N,a}(t)$ be a discrete approximation of $\rho_{K,N,a}(t)$ defined inductively by $\rho^\alpha_{K,N,a}(0) = a$ and for $t \in [t_n^\alpha, t_{n+1}^\alpha]$

$$\rho^\alpha_{K,N,a}(t) := \rho^\alpha_{K,N,a}(t_n^\alpha) + \frac{t - t_n^\alpha}{\alpha^2} \left(\alpha \lambda_{n+1} + \alpha^2 \Psi(\rho^\alpha_{K,N,a}(t_n^\alpha)) \right).$$

For $R > 0$, let us define $R^* > 1$ by

$$R^* := \begin{cases} \bar{R} - \dfrac{1}{R} & \text{if } K > 0 \text{ and } N < \infty, \\ R & \text{otherwise.} \end{cases}$$

The following comparison theorem is crucial for the proof of Theorem 2.2.

Proposition 3.5 *For $T > 0$, $R > 0$ and $\varepsilon > 0$, there exists a constant $C(\varepsilon, T) \ge 0$ satisfying $\lim_{\varepsilon \to 0} C(\varepsilon, T) = 0$ such that*

$$d(\mathbf{X}_t^\alpha) \le \rho^\alpha_{K,N,a}(t) + C(\varepsilon, T)$$

holds for $t < \hat{\tau}_\delta \wedge \hat{\sigma}_R \wedge \sigma_{R^*}(\rho^\alpha_{K,N,a}) \wedge T$ on $(E_\varepsilon^\alpha)^c$ for sufficiently small α.

Proof. By [17, Corollary 3.6(i)], it suffices to show the assertion only when $t = t_n^\alpha$ for some $n \le n^{(\alpha)}$ (cf. [17, Lemma 3.10]). For simplicity of notations, we denote $\rho^\alpha_{K,N,a}(t_n^\alpha)$ by $\rho^\alpha(n)$. Applying Proposition 3.1, we obtain

$$r^\alpha(n+1) - \rho^\alpha(n+1) \le r^\alpha(n) - \rho^\alpha(n) + \alpha^2(\Psi(r^\alpha(n)) - \Psi(\rho^\alpha(n))) + \varepsilon\alpha^2 \qquad (3.7)$$

for $n \le n^{(\alpha)}$ with $t_n^\alpha < \hat{\tau}_\delta \wedge \hat{\sigma}_R \wedge \sigma_{R^*}(\rho^\alpha_{K,N,a})$ on E_ε^α. Under our assumption on $t = t_n^\alpha$, $r^\alpha(n) \in [\delta, R]$ and $\rho^\alpha(n) \in [0, R^*]$ hold. Note that Ψ is bounded on $[0, \mathrm{diam}(M) \wedge R^*]$. Let $f_\alpha : \mathbb{R} \to \mathbb{R}$ be a function of class C^2 satisfying the following conditions:

(i) $f_\alpha(x) = 0$ for $x \le 0$ and $f_\alpha(x) = x + \alpha/2$ for $x \ge \alpha$,

(ii) f_α is convex,

(iii) $\limsup_{\alpha \to 0} \alpha^2 \sup_{u \in \mathbb{R}} f''_\alpha(u) < C$ for some $C > 0$

11

(cf. the proof of [17, Lemma 3.10]). By (3.7), the Taylor expansion together with the condition (iii) of f_α yields

$$
\begin{aligned}
f_\alpha(r^\alpha(n+1) - \rho^\alpha(n+1)) \leq{} & f_\alpha(r^\alpha(n) - \rho^\alpha(n)) \\
& + \alpha^2 f'_\alpha(r^\alpha(n) - \rho^\alpha(n))\left(\Psi(r^\alpha(n)) - \Psi(\rho^\alpha(n))\right) + 2\varepsilon\alpha^2 \quad (3.8)
\end{aligned}
$$

for sufficiently smaller α than ε. Since Ψ is nonincreasing, properties (i) and (ii) of f_α imply

$$
f'_\alpha(r^\alpha(n) - \rho^\alpha(n))\left(\Psi(r^\alpha(n)) - \Psi(\rho^\alpha(n))\right) \leq 0.
$$

Thus, an iteration of (3.8) together with the fact $f_\alpha(x) + \alpha/2 \geq x \vee 0$ yield

$$
(r^\alpha(n) - \rho^\alpha(n))_+ \leq f_\alpha(r^\alpha(n) - \rho^\alpha(n)) + \frac{\alpha}{2} \leq 2\varepsilon\alpha^2 n + \varepsilon
$$

for $\alpha \leq 2\varepsilon$. Since $t_n^\alpha = \alpha^2 n \leq T$, the conclusion follows. $\qquad\square$

Now we are in position to give a crucial step of the proof of Theorem 2.2.

Proposition 3.6 *For any nondecreasing continuous function $\psi : \overline{[0, \bar{R})} \to [0, 1]$ with $\psi(0) = 0$, we have*

$$
\mathbb{E}\left[\psi(d(\mathbf{X}^*(s)))\right] \leq \mathbb{E}\left[\psi(\rho(s)) ; \tau_0(\rho) > s\right].
$$

Proof. Take $\delta > 0$, $R > 1$ and $t > s$. Let $\varepsilon > 0$ be so small that $C(\varepsilon, t) < \delta/2$. By virtue of Proposition 3.1, for sufficiently small α,

$$
\begin{aligned}
\mathbb{E}[\psi(d(\mathbf{X}^\alpha(s)))] \leq{} & \mathbb{E}\left[\psi(d(\mathbf{X}^\alpha(s))) ; \{\hat{\tau}_\delta > s\} \cap \{\hat{\sigma}_R > s\} \cap (E_\varepsilon^\alpha)^c\right] \\
& + \mathbb{P}[\hat{\sigma}_R \leq s] + \mathbb{E}\left[\psi(d(\mathbf{X}^\alpha(s))) ; \hat{\tau}_\delta \leq s\right] + \varepsilon. \quad (3.9)
\end{aligned}
$$

By Proposition 3.5 and the choice of ε,

$$
\begin{aligned}
\mathbb{E}\left[\psi(d(\mathbf{X}^\alpha(s))) ; \{\hat{\tau}_\delta > s\} \cap \{\hat{\sigma}_R > s\} \cap (E_\varepsilon^\alpha)^c\right] & \\
\leq \mathbb{E}\left[\psi(\rho^\alpha(s) + C(\varepsilon, t)) ; \tau_{\delta/2}(\rho^\alpha) \wedge \sigma_{R^*}(\rho^\alpha) > s\right] & + \mathbb{P}[\sigma_{R^*}(\rho^\alpha) \leq s]. \quad (3.10)
\end{aligned}
$$

Let us define $\tilde{\Psi} : [0, \infty) \to \mathbb{R}$ by

$$
\tilde{\Psi}(u) := (\Psi(u) \wedge |\Psi((2R)^*)|) \vee (-|\Psi(-(2R)^*)|).
$$

We define $\tilde{\rho}^\alpha$ and $\tilde{\rho}$ by replacing Ψ with $\tilde{\Psi}$ in the definition of ρ^α and ρ respectively. Since $\tilde{\Psi}(u) = \Psi(u)$ for $u \in [0, R^*]$, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\psi(\rho^\alpha(s) + C(\varepsilon, t)) ; \tau_{\delta/2}(\rho^\alpha) \wedge \sigma_{R^*}(\rho^\alpha) > s\right] & \\
\leq \mathbb{E}\left[\psi(\tilde{\rho}^\alpha(s) + C(\varepsilon, t)) ; \tau_{\delta/2}(\tilde{\rho}^\alpha) > s\right], & \quad (3.11) \\
\mathbb{P}[\sigma_{R^*}(\rho^\alpha) \leq s] = \mathbb{P}[\sigma_{R^*}(\tilde{\rho}^\alpha) \leq s]. & \quad (3.12)
\end{aligned}
$$

Since $\tilde{\Psi}$ is bounded and continuous, we can easily show that $\tilde{\rho}^\alpha$ converges in law to $\tilde{\rho}$ in $C([0, \infty) \to \mathbb{R})$. Note that the following holds:

$$
\overline{\left\{w \in \mathscr{C} ; \tau_{\delta/2}(d(w)) > s\right\}} \subset \left\{w \in \mathscr{C} ; \tau_{\delta/4}(d(w)) > s\right\}.
$$

12

By combining this fact with (3.11), the Portmanteau theorem together with (3.10), (3.11) and (3.12) yields

$$\limsup_{\alpha \to 0} \mathbb{E}\left[\psi(d(\mathbf{X}^\alpha(s))) \ ; \ \{\hat{\tau}_\delta > s\} \cap \{\hat{\sigma}_R > s\} \cap (E_\varepsilon^\alpha)^c\right]$$

$$\leq \mathbb{E}\left[\psi(\tilde{\rho}(s) + C(\varepsilon, t)) \ ; \ \tau_{\delta/4}(\tilde{\rho}) > s\right] + \mathbb{P}[\sigma_{R^*}(\tilde{\rho}) \leq s]. \quad (3.13)$$

In a similar way as (3.11) and (3.12), we obtain

$$\mathbb{E}\left[\psi(\tilde{\rho}(s) + C(\varepsilon, t)) \ ; \ \tau_{\delta/4}(\tilde{\rho}) > s\right] + \mathbb{P}[\sigma_{R^*}(\tilde{\rho}) \leq s]$$

$$\leq \mathbb{E}\left[\psi(\rho(s) + C(\varepsilon, t)) \ ; \ \tau_{\delta/4}(\rho) > s\right] + 2\mathbb{P}[\sigma_{R^*}(\rho) \leq s]. \quad (3.14)$$

Here we used the fact $\psi \leq 1$. Since $\mathbf{X}^\alpha$ converges in law to $\tilde{\mathbf{X}}$ in $\mathscr{C}$, by applying the Portmanteau theorem to (3.9) together with (3.13) and (3.14), we obtain

$$\mathbb{E}\left[\psi(d(\mathbf{X}(s)))\right] = \lim_{\alpha \to 0} \mathbb{E}\left[\psi(d(\mathbf{X}^\alpha(s)))\right]$$

$$\leq \mathbb{E}\left[\psi(\rho(s) + C(\varepsilon, t)) \ ; \ \tau_{\delta/4}(\rho) > s\right] + 2\mathbb{P}[\sigma_{R^*}(\rho) \leq s]$$

$$+ \limsup_{\alpha \to 0} \mathbb{P}[\hat{\sigma}_R \leq s] + \mathbb{E}\left[\psi(d(\mathbf{X}(s))) \ ; \ \tau_\delta(d(\mathbf{X}(\cdot))) \leq s\right] + \varepsilon.$$

By letting $\varepsilon \to 0$ in this inequality, we obtain

$$\mathbb{E}\left[\psi(d(\mathbf{X}(s))) \ ; \ \tau_\delta(d(\mathbf{X}(\cdot))) > s\right]$$

$$\leq \mathbb{E}\left[\psi(\rho(s)) \ ; \ \tau_{\delta/4}(\rho) > s\right] + 2\mathbb{P}[\sigma_{R^*}(\rho) \leq s] + \limsup_{\alpha \to 0} \mathbb{P}[\hat{\sigma}_R \leq s]. \quad (3.15)$$

By the definition of $\mathbf{X}^*$ and τ^*, we have

$$\lim_{\delta \to 0} \mathbb{E}\left[\psi(d(\mathbf{X}(s))) \ ; \ \tau_\delta(d(\mathbf{X}(\cdot))) > s\right] = \lim_{\delta \to 0} \mathbb{E}\left[\psi(d(\mathbf{X}^*(s))) \ ; \ \tau_\delta(d(\mathbf{X}^*(\cdot))) > s\right]$$

$$= \mathbb{E}\left[\psi(d(\mathbf{X}^*(s))) \ ; \ \tau^* > s\right]$$

$$= \mathbb{E}\left[\psi(d(\mathbf{X}^*(s)))\right]. \quad (3.16)$$

Here the last equality follows from $\psi(0) = 0$. Similarly we obtain

$$\lim_{\delta \to 0} \mathbb{E}\left[\psi(\rho(s)) \ ; \ \tau_{\delta/4}(\rho) > s\right] = \mathbb{E}\left[\psi(\rho(s)) \ ; \ \tau_0(\rho) > s\right]. \quad (3.17)$$

Thus, by combining (3.15) with (3.16) and (3.17) and by tending $R \to \infty$ with (3.1) in mind, we obtain

$$\mathbb{E}\left[\psi(d(\mathbf{X}^*(s)))\right] \leq \mathbb{E}\left[\psi(\rho(s)) \ ; \ \tau_0(\rho) > s\right].$$

Here we used the fact that ρ cannot hit $\bar{R}$ in finite time. Hence the assertion holds. $\qquad \square$

To complete the proof of Theorem 2.2, we will use a regularity result on φ_t in the next section. As you will see, all the arguments in the next section are independent of this section. Thus there are no danger of circular arguments.

Proof of Theorem 2.2. By virtue of Proposition 4.5 (ii) below, we can apply Proposition 3.6 with $\psi = \varphi_t$. Thus we obtain

$$\mathbb{E}[\varphi_t(d(\mathbf{X}^*(s)))] \leq \mathbb{E}[\varphi_t(\rho(s)) \; ; \; \tau_0(\rho) > s]. \tag{3.18}$$

Since $-\rho_{K,N,a} \overset{d}{=} \rho_{K,N,-a}$ holds, a process $\tilde{\rho}^* = (\tilde{\rho}^{(1)}, \tilde{\rho}^{(2)})$ given by

$$\tilde{\rho}^*(t) := \begin{cases} \left(\dfrac{\rho_{K,N,a}(t)}{2}, -\dfrac{\rho_{K,N,a}(t)}{2} \right) & \text{if } t < \tau_0(\rho_{K,N,a}), \\ \left(\dfrac{\rho_{K,N,a}(t)}{2}, \dfrac{\rho_{K,N,a}(t)}{2} \right) & \text{if } t \geq \tau_0(\rho_{K,N,a}). \end{cases}$$

is a coupling of $\rho_{K,N,a}/2$ and $\rho_{K,N,-a}/2$. Since the reflection map $x \mapsto -x$ on $(-\bar{R}/2, \bar{R}/2)$ provides a reflection structure for $\rho_{K,N,a}/2$ in the sense in [19], $\tilde{\rho}^*$ is a maximal coupling of $\rho_{K,N,a}/2$ and $\rho_{K,N,-a}/2$, and $\tau_0(\rho_{K,N,a}/2) = \tau_0(\rho_{K,N,a})$ is the coupling time of $\tilde{\rho}^*$. Thus Remark 2.1 (ii) yields

$$\varphi_t(|\tilde{\rho}^*(s)|) = \varphi_{t+s}(a). \tag{3.19}$$

Since the definition of $\tilde{\rho}^*$ implies

$$\mathbb{E}[\varphi_t(\rho(s)) \; ; \; \tau_0(\rho) > s] = \mathbb{E}[\varphi_t(|\tilde{\rho}^*(s)|)],$$

the combination of it with (3.18) and (3.19) deduces the conclusion. $\qquad\square$

4 PROPERTIES OF THE COST FUNCTION

Let us define $\chi : [0, \infty] \to [0, 1]$ by

$$\chi(r) := \frac{1}{\sqrt{2\pi}} \int_{-r}^{r} e^{-u^2/2} du$$

and $\chi(\infty) = 1$. We can easily verify that χ is increasing and concave. The first goal of this section is the following expression of $\varphi_t(a)$:

Proposition 4.1 *For each $N \in [2, \infty]$, $K \in \mathbb{R}$ and $t \geq 0$, there exists a probability measure $\zeta_{t,K,N}$ on $[0, \infty)$ such that*

$$\varphi_t(a) := \int_{[0,\infty)} \chi\left(\frac{a}{2\sqrt{2u}} \right) \zeta_{t,K,N}(du) \tag{4.1}$$

holds for each $a \in [0, \infty)$. In addition, we can take $\zeta_{t,K,N}$ so that it is continuous in t with respect to the topology of weak convergence.

The expression (4.1) will be used to study some properties of $\varphi_t(a)$ in Proposition 4.5. We divide the proof of Proposition 4.1 into the following two lemmata; Lemma 4.2 when $N = \infty$ or $K = 0$ and Lemma 4.4 when $N < \infty$ and $K \neq 0$. We will give an expression of $\zeta_{t,K,N}$ there.

14

Lemma 4.2 *Suppose $N = \infty$ or $K = 0$. Then*

$$\varphi_t(a) = \chi\left(\frac{a}{2\sqrt{2\eta(t)}}\right),$$

where $\eta(t) = \eta_K(t)$ is given by

$$\eta_K(t) := \begin{cases} \dfrac{e^{2Kt} - 1}{2K} & K \neq 0, \\ t & K = 0. \end{cases}$$

In particular, Proposition 4.1 holds with $\zeta_{t,K,\infty} = \zeta_{t,0,N} = \delta_{\eta(t)}$.

Proof. In this case, $\rho_t := e^{-Kt}a + 2\sqrt{2}\int_0^t e^{K(s-t)}d\beta_s$ holds. By the martingale representation theorem, $\int_0^t e^{Ks}d\beta_s$ can be written as a deterministic time-change of a standard one-dimensional Brownian motion. By using this fact together with the expression of the hitting time distribution of the Brownian motion, the desired expression of φ_t follows. $\square$

To consider the case $N < \infty$, we begin with the following auxiliary lemma:

Lemma 4.3 *Suppose $N < \infty$. Let $\beta^i(t)$ be the standard one-dimensional Brownian motion for $i = 1, 2$. Let $\theta(t)$ be the solution to the following stochastic differential equation:*

$$d\theta(t) = \sqrt{2}d\beta^1(t) + \left(\frac{N-2}{t_{K/(N-1)}(\theta(t))} - \frac{K}{N-1}t_{K/(N-1)}(\theta(t))\right)dt,$$

$$\theta(0) = 0.$$

Let us define $\Xi(t)$ by

$$\Xi(t) := 2s_{K/(N-1)}^{-1}\left(c_{K/(N-1)}(\theta(t))\,s_{K/(N-1)}\left(\frac{a}{2} + \sqrt{2}\int_0^t \frac{d\beta^2(s)}{c_{K/(N-1)}(\theta(s))}\right)\right). \tag{4.2}$$

Then Ξ has the same law as ρ.

The alternative expression of ρ in the last lemma comes from a skew-product expression of the distance between two Brownian motions coupled by reflection on a sphere. For explaining a heuristic idea behind it, we assume $N \in \mathbb{N}$, $K = N - 1$ and $Z \equiv 0$ for a while. We identify the sphere $\mathbb{S}^N$ of constant sectional curvature 1 with an unit sphere in $\mathbb{R}^{N+1}$ as a submanifold. Let H be a (uniquely determined) 2-dimensional plane in $\mathbb{R}^{N+1}$ containing origin and given starting points of the coupling of Brownian motions by reflection. Then we can decompose the Brownian motion on $\mathbb{S}^N$ into the "circular part", that is, the projection to H and the "complementary part", that is, the projection to $H^\perp$. As a result, we can describe the distance between the two Brownian particles coupled by reflection by the scaled distance between two time-changed Brownian motions coupled by reflection on a circle whose space scaling and clock process are given by functionals of the complementary part. This description leads us to the expression in Lemma 4.3. Moreover, once we obtained this expression, we can verify it valid even when $N \notin \mathbb{N}$ or $K < 0$ as we will see in the following proof of Lemma 4.3. For the skew product expression of spherical Brownian motions, see [14, 29], for example.

Proof of Lemma 4.3. For simplicity of notations, we denote $K/(N-1)$ by $\bar{K}$ in this proof. Let $\hat{\Xi}(t) := s_{\bar{K}}(\Xi(t)/2)$ and $\hat{\rho}(t) := s_{\bar{K}}(\rho(t)/2)$. It suffices to show that both $\hat{\Xi}(t)$ and $\hat{\rho}(t)$ solves the following stochastic differential equation

$$dz(t) = \sqrt{1 - \bar{K}z(t)^2}dw(t) - \frac{N\bar{K}}{2}z(t)dt \qquad (4.3)$$

for a standard Brownian motion $w(t)$. By the Itô formula together with (2.1), we can easily verify that $\hat{\rho}$ solves (4.3) with $w(t) = \beta(t)$. The Itô formula together with (4.2) yields

$$
\begin{aligned}
d\hat{\Xi}(t) &= -\bar{K}s_{\bar{K}}(\theta(t))s_{\bar{K}}\left(\frac{a}{2} + \sqrt{2}\int_0^t \frac{d\beta^2(s)}{c_{\bar{K}}(\theta(s))}\right)d\theta(t) \\
&\quad + \sqrt{2}c_{\bar{K}}\left(\frac{a}{2} + \sqrt{2}\int_0^t \frac{d\beta^2(s)}{c_{\bar{K}}(\theta(s))}\right)d\beta^2(t) \\
&\quad - \bar{K}\left(c_{\bar{K}}(\theta(t)) + \frac{1}{c_{\bar{K}}(\theta(t))}\right)s_{\bar{K}}\left(\frac{a}{2} + \sqrt{2}\int_0^t \frac{d\beta^2(s)}{c_{\bar{K}}(\theta(s))}\right)dt \\
&= -\sqrt{2}\bar{K}s_{\bar{K}}(\theta(t))s_{\bar{K}}\left(\frac{a}{2} + \sqrt{2}\int_0^t \frac{d\beta^2(s)}{c_{\bar{K}}(\theta(s))}\right)d\beta^1(t) \\
&\quad + \sqrt{2}c_{\bar{K}}\left(\frac{a}{2} + \sqrt{2}\int_0^t \frac{d\beta^2(s)}{c_{\bar{K}}(\theta(s))}\right)d\beta^2(t) - N\bar{K}\hat{\Xi}(t)dt.
\end{aligned}
$$

Here we used the relation $c_{\bar{K}}(r)^2 + \bar{K}s_{\bar{K}}(r)^2 = 1$, which holds for any $K \in \mathbb{R}$, to obtain the last equality. By a direct computation, we have

$$\left(\bar{K}s_{\bar{K}}(\theta(t))s_{\bar{K}}\left(\frac{a}{2} + \sqrt{2}\int_0^t \frac{d\beta^2(s)}{c_{\bar{K}}(\theta(s))}\right)\right)^2 + c_{\bar{K}}\left(\frac{a}{2} + \sqrt{2}\int_0^t \frac{d\beta^2(s)}{c_{\bar{K}}(\theta(s))}\right)^2 = 1 - \bar{K}\hat{\Xi}(t)^2.$$

Note that $1 - \bar{K}\hat{\Xi}(t)^2 > 0$ holds for any $t \geq 0$ almost surely since $\theta(t)$ never hits $\bar{R}/2$. Thus, $\hat{\Xi}$ solves (4.3) with $w(t) = \beta^*(t)$ given by

$$
\begin{aligned}
\beta^*(t) := \int_0^t \frac{1}{\sqrt{1 - \bar{K}\hat{\Xi}(t)^2}}\Bigg(&-\bar{K}s_{\bar{K}}(\theta(s))s_{\bar{K}}\left(\frac{a}{2} + \sqrt{2}\int_0^s \frac{d\beta^2(u)}{c_{\bar{K}}(\theta(u))}\right)d\beta^1(s) \\
&+ c_{\bar{K}}\left(\frac{a}{2} + \sqrt{2}\int_0^s \frac{d\beta^2(u)}{c_{\bar{K}}(\theta(u))}\right)d\beta^2(s)\Bigg)
\end{aligned}
$$

and hence the conclusion follows. $\qquad\square$

Lemma 4.4 *Suppose $N < \infty$. We denote the law of $\int_0^t c_{K/(N-1)}(\theta(s))^{-2}ds$ by $\zeta_{t,K,N}$ for each $t \geq 0$, where $\theta(\cdot)$ is as in Lemma 4.3. Then the conclusion of Proposition 4.1 holds true.*

Proof. The continuity in t of $\zeta_{t,K,N}$ directly follows from the definition. Let Ξ be as in Lemma 4.3. By the martingale representation theorem, there exists a standard one-dimensional Brownian motion $B(t)$ such that

$$B\left(2\int_0^{\cdot} \frac{ds}{c_{K/(N-1)}(\theta(s))^2}\right) \stackrel{d}{=} \sqrt{2}\int_0^{\cdot} \frac{d\beta^2(s)}{c_{K/(N-1)}(\theta(s))}$$

16

holds. Since β^1 and β^2 are independent, $B(\cdot)$ behaves as a standard Brownian motion even under the conditional probability measure $\mathbb{P}[\,\cdot\,|\sigma(\beta^1)]$. Thus the definition of $\varphi_t(a)$ and Lemma 4.3 yield

$$
\begin{aligned}
\varphi_t(a) &= \mathbb{P}\left[\inf_{0 \le s \le t} \rho(s) > 0\right] = \mathbb{P}\left[\inf_{0 \le s \le t}\left(\frac{a}{2} + \sqrt{2}\int_0^s \frac{d\beta^2(u)}{c_{K/(N-1)}(\theta(u))}\right) > 0\right] \\
&= \mathbb{P}\left[\inf\left\{\frac{a}{2} + B(s) \,\Big|\, 0 \le s \le 2\int_0^t \frac{du}{c_{K/(N-1)}(\theta(u))^2}\right\} > 0\right] \\
&= \mathbb{E}\left[\mathbb{P}\left[\inf\left\{\frac{a}{2} + B(s) \,\Big|\, 0 \le s \le 2\int_0^t \frac{du}{c_{K/(N-1)}(\theta(u))^2}\right\} > 0 \Big| \sigma(\beta^1)\right]\right] \\
&= \mathbb{E}\left[\chi\left(\frac{a}{2\sqrt{2}}\left(\int_0^t \frac{du}{c_{K/(N-1)}(\theta(u))^2}\right)^{-1/2}\right)\right].
\end{aligned}
$$

Hence the desired result holds. $\qquad\square$

Now we state some consequences of the expressions of $\varphi_t(a)$ in Proposition 4.1:

Proposition 4.5   (i) *For $a \in [0, \bar{R}]$, $[0, \infty) \ni t \mapsto \varphi_t(a)$ is continuous.*

  (ii) *φ_t is continuous on $\overline{[0, \bar{R})}$ and smooth on $(0, \bar{R})$ for $t > 0$.*

  (iii) *φ_t is concave on $\overline{[0, \bar{R})}$ for $t \ge 0$.*

  (iv) *For $t \ge 0$, $K, K' \in \mathbb{R}$ with $K \ge K'$, $N, N' \in [2, \infty]$ with $N \le N'$ and $a \in [0, \bar{R}_{K,N}]$,*

$$
\varphi_t^{K,N}(a) \le \varphi_t^{K',N'}(a).
$$

  (v) *For $t > 0$, $K \in \mathbb{R}$ and $N \in [2, \infty]$, $\varphi_t^{K,N}$ is differentiable at 0. Moreover, for $K' \in \mathbb{R}$ and $N' \in [2, \infty]$ with $K' \le K$, $N' \ge N$,*

$$
(\varphi_t^{K,N})'(0) = \int_{[0,\infty)} \frac{\zeta_{t,K,N}(du)}{4\sqrt{\pi u}} \le (\varphi_t^{K',N'})'(0). \tag{4.4}
$$

  In particular, $(\varphi_t^{K,N})'(0) \le (\varphi^{K,\infty})'(0) = (\pi \eta(t))^{-1/2}/4$. Here $\eta(t) = \eta_K(t)$ is as in Lemma 4.2.

  (vi) *For $K \in \mathbb{R}$ and $N \in [2, \infty]$, $\displaystyle\lim_{t \downarrow 0} \sqrt{t}(\varphi_t^{K,N})'(0) = \frac{1}{4\sqrt{\pi}}$.*

Proof.   (i) It is obvious by the continuity of $\zeta_{t,K,N}$ in (4.1).

  (ii) Note that the derivative of $\chi(a/(2\sqrt{2u}))$ of any order with respect to a-variable is a bounded function of u for $a \in (0, \bar{R})$ in (4.1). Thus the dominated convergence theorem yields that φ_t is smooth on $(0, \bar{R})$. We can show the continuity of φ_t on $\overline{[0, \bar{R})}$ similarly.

  (iii) Since $\varphi_0 = 1_{(0,\infty)}$ by definition, it is obviously concave. Thus it suffices to consider the case $t > 0$. As we did in the proof of (ii), we can compute $\varphi_t''(a)$ at $a \in (0, \bar{R})$ by

using the dominated convergence theorem. Since χ is concave, $\varphi_t''(a) \le 0$ and hence the conclusion holds because φ_t is continuous on $\overline{[0, \bar{R})}$ by (ii).

(iv) By a direct computation, we can verify $\Psi_{K,N}(u) \le \Psi_{K',N'}(u)$ for any $u \in [0, \bar{R}_{K,N})$. Thus the comparison theorem for stochastic differential equations (see [13] for instance) yields that $\rho_{K,N,a}(t) \le \rho_{K',N',a'}(t)$ for $a' > a$ and $t \ge 0$. It implies $\varphi_t^{K,N}(a) \le \varphi_t^{K',N'}(a')$ by the definition of $\varphi_t^{K,N}(a)$. Since φ_t is continuous, the asserted inequality follows by tending $a' \downarrow a$.

(v) Since χ is concave and $\chi(0) = 0$, $\chi(r)/r$ is nonincreasing. Thus the monotone convergence theorem yields

$$\lim_{a \downarrow 0} \frac{\varphi_t^{K,N}(a) - \varphi_t^{K,N}(0)}{a} = \lim_{a \downarrow 0} \frac{\varphi_t^{K,N}(a)}{a} = \int_{[0,\infty)} \frac{\zeta_{t,K,N}(du)}{4\sqrt{\pi u}}.$$

By combining this identity with (iv), we obtain

$$\int_{[0,\infty)} \frac{\zeta_{t,K,N}(du)}{4\sqrt{\pi u}} \le \int_{[0,\infty)} \frac{\zeta_{t,K',N'}(du)}{4\sqrt{\pi u}} \le \int_{[0,\infty)} \frac{\zeta_{t,K',\infty}(du)}{4\sqrt{\pi u}} = \frac{1}{4\sqrt{\pi \eta_{K'}(t)}} < \infty$$

and hence the conclusion follows.

(vi) We use the expression of $(\varphi_t^{K,N})'(0)$ in (v). When $N = \infty$, it easily follows from Lemma 4.2. Next we consider the case $K > 0$ with the expression of $\zeta_{t,K,N}$ given in Lemma 4.4. By the definition of $\theta(t)$ in Lemma 4.3, $c_{K/(N-1)}(\theta(t)) \in (0, 1]$ holds for each $t > 0$. Thus we have $\int_0^t (c_{K/(N-1)}(\theta(s)))^{-2} ds \ge t$ and therefore the dominated convergence theorem yields

$$\lim_{t \downarrow 0} \sqrt{t}(\varphi_t^{K,N})'(0) = \lim_{t \downarrow 0} \frac{1}{4\sqrt{\pi}} \mathbb{E}\left[\left(\frac{1}{t}\int_0^t \frac{ds}{c_{K/(N-1)}(\theta(s))^2}\right)^{-1/2}\right] = \frac{1}{4\sqrt{\pi}}.$$

Finally, for the general $K \in \mathbb{R}$ and $N \in [2, \infty)$, the conclusion follows from (4.4) together with the above-mentioned two cases. $\qquad \square$

Since $\varphi_r(0) = 0$, Proposition 4.5 (iii) yields the following corollary:

Corollary 4.6 *We have $\varphi_r(a+a') \le \varphi_r(a) + \varphi_r(a')$ for $r > 0$ and $a, a' \ge 0$. In particular, for $t > 0$, $\varphi_t(d(\cdot, \cdot))$ is a bounded distance function being compatible with the topology on M.*

Though the preparation of the proof of Theorem 2.3 is already finished in Proposition 4.5, we will discuss further properties of φ_t in the rest of this section. First we will study more explicit expression of $\varphi_t(a)$ than the one in Lemma 4.4 in the case $N < \infty$ and $K \ne 0$. Lemma 4.7 and Corollary 4.8 below study the case $K < 0$. Based on the expression of the Brownian motion on the hyperbolic space by a stochastic differential equation (see [27], for instance), we can show the following in a similar way as Lemma 4.3:

Lemma 4.7 *Suppose $N < \infty$ and $K < 0$. Let $\beta^1(t)$ and $\beta^2(t)$ be independent, one-dimensional standard Brownian motions. Let $\Xi'(t)$ and $\theta'(t)$ be given by*

$$\theta'(t) := \exp\left(\sqrt{\frac{-2K}{N-1}}\beta^1(t) + Kt\right),$$

$$\Xi'(t) := s_{K/(N-1)}\left(\frac{a}{2}\right) + \sqrt{2}\int_0^t \theta'(s)d\beta^2(s).$$

Then $2s_{K/(N-1)}^{-1}(\Xi'(\cdot)/\theta'(\cdot))$ has the same law as ρ.

Proof. As in the proof of Lemma 4.3, we denote $K/(N-1)$ by $\bar{K}$. We already know in the proof of Lemma 4.3 that $s_{\bar{K}}(\rho(t)/2)$ solves the stochastic differential equation (4.3) with $w(t) = \beta(t)$. Thus it suffices to show that $\Xi'(t)/\theta'(t)$ also solves (4.3) for a standard Brownian motion $w(t)$. The Itô formula yields

$$d\left(\frac{\Xi'(t)}{\theta'(t)}\right) = -\frac{\Xi'(t)}{\theta'(t)^2}d\theta'(t) + \frac{\Xi'(t)}{\theta'(t)^3}d\langle\theta'\rangle(t) + \frac{1}{\theta'(t)}d\Xi'(t)$$

$$= -\sqrt{-\bar{K}}\frac{\Xi'(t)}{\theta'(t)}\left(\sqrt{2}d\beta^1(t) - (N-2)\sqrt{-\bar{K}}dt\right) - 2\bar{K}\frac{\Xi'(t)}{\theta'(t)}dt + \sqrt{2}d\beta^2(t)$$

$$= \sqrt{2}d\beta^2(t) - \sqrt{-2\bar{K}}\frac{\Xi'(t)}{\theta'(t)}d\beta^1(t) - N\bar{K}\frac{\Xi'(t)}{\theta'(t)}dt.$$

Thus $\Xi'(t)/\theta'(t)$ solves (4.3) with $w(t) = \beta^{**}(t)$ given by

$$\beta^{**}(t) := \int_0^t \left(1 - \bar{K}\left(\frac{\Xi'(s)}{\theta'(s)}\right)^2\right)^{-1/2}\left(d\beta^2(s) - \sqrt{-\bar{K}}\left(\frac{\Xi'(s)}{\theta'(s)}\right)d\beta^1(s)\right).$$

$\square$

Corollary 4.8 *Suppose $N < \infty$ and $K < 0$. Then*

$$\varphi_t(a) = \mathbb{E}\left[\chi\left(\frac{1}{2\sqrt{2}}s_{K/(N-1)}\left(\frac{a}{2}\right)\left(\int_0^t \theta'(s)^2 ds\right)^{-1/2}\right)\right], \tag{4.5}$$

where $\theta'(t)$ is as in Lemma 4.7. Moreover,

$$\varphi_t(a) = \int_{-\infty}^\infty \int_0^\infty \chi\left(\frac{1}{2}\sqrt{\frac{-K}{(N-1)u}}s_{K/(N-1)}\left(\frac{a}{2}\right)\right)$$

$$\times \exp\left(\frac{(N-1)}{2}(Kt - x) - \frac{1+e^{2x}}{2u}\right)\vartheta\left(\frac{e^x}{u}, \frac{-2Kt}{N-1}\right)\frac{du}{u}dx,$$

where

$$\vartheta(r, t) := \frac{r}{2\pi^3 t}e^{\pi^2/(2t)}\int_0^\infty e^{-\xi^2/(2t)}e^{-r\cosh(\xi)}\sinh(\xi)\sin\left(\frac{\pi\xi}{t}\right)d\xi.$$

Proof. Let $\Xi'(t)$ and $\theta'(t)$ be as in Lemma 4.7. By the martingale representation theorem, there exists a Brownian motion $B(t)$ such that

$$\Xi'(t) \stackrel{d}{=} s_{K/(N-1)}\left(\frac{a}{2}\right) + B\left(2\int_0^t \theta'(s)^2 ds\right).$$

Hence, as in the proof of Lemma 4.4, the definition of $\varphi_t(a)$ and Lemma 4.7 yield

$$\varphi_t(a) = \mathbb{P}\left[\inf_{0 \le s \le t} \rho(s) > 0\right] = \mathbb{P}\left[\inf_{0 \le s \le t} \Xi'(s) > 0\right]$$

$$= \mathbb{P}\left[\inf\left\{s_{K/(N-1)}\left(\frac{a}{2}\right) + B(s) \;\Big|\; 0 \le s \le 2\int_0^t \theta'(u)^2 du\right\} > 0\right]$$

$$= \mathbb{E}\left[\chi\left(\frac{1}{2\sqrt{2}} s_{K/(N-1)}\left(\frac{a}{2}\right)\left(\int_0^t \theta'(u)^2 du\right)^{-1/2}\right)\right].$$

This is nothing but (4.5). Now the conclusion follows by using an explicit expression of the distribution of $\int_0^t \theta'(u)^2 du$ in [26, Theorem 4.1] (also see references therein). $\square$

In the case $K > 0$, we use several properties on the Gegenbauer, or ultraspherical, polynomials to obtain alternative expression of φ_t in Lemma 4.9 below. We refer to [33] for basics on Gegenbauer polynomials.

Lemma 4.9 *Suppose $N < \infty$ and $K > 0$. Then, for all $a \in [0, \bar{R}]$,*

$$\varphi_t(a) = \sum_{n=0}^{\infty} e^{-(2n+1)(2n+N)Kt/(N-1)} \frac{(-1)^n(4n+N+1)}{\pi(2n+N)} B\left(\frac{N-1}{2}, n+\frac{1}{2}\right) P_{2n+1}(\tilde{a}),$$

where $B(\cdot, \cdot)$ is the Beta function, $\tilde{a} := \sin(\sqrt{K/(N-1)}a/2)$ and $P_n(x)$ is the n-th Gegenbauer polynomial of parameter $(N-1)/2$.

Proof. Let us define $\hat{\rho}(t)$ by

$$\hat{\rho}(t) := \sin\left(\frac{1}{2}\sqrt{\frac{K}{N-1}}\rho\left(\frac{(N-1)t}{2K}\right)\right).$$

Then $\hat{\rho}(t)$ solves the following stochastic differential equation:

$$d\hat{\rho}(t) = \sqrt{1-\hat{\rho}(t)^2}d\beta(t) - \frac{N}{2}\hat{\rho}(t)dt,$$

$$\hat{\rho}(0) = \tilde{a},$$

where $\beta(t)$ is a one-dimensional standard Brownian motion. Thus $\hat{\rho}(t)$ is the Legendre process, or the diffusion process on $(-1, 1)$ generated by L_N given as follows:

$$L_N := \frac{1}{2}(1-x^2)\frac{\partial^2}{\partial x^2} - \frac{N}{2}x\frac{\partial}{\partial x}.$$

20

It is well-known that $\mu_N(dx) = (1-x^2)^{N/2-1}dx$ is the symmetrizing measure of L_N. Moreover, L_N is essentially selfadjoint on $L^2(\mu_N)$, the spectra of L_N on $L^2(\mu_N)$ is $\{-n(n+N-1)/2\}_{n\in\mathbb{N}_0}$ all of which are eigenvalues of multiplicity one, and the normalized eigenfunction corresponding to the n-th eigenvalue is n-th normalized Gegenbauer polynomials $\bar{P}_n$ defined by $\bar{P}_n(x) = Z_n^{-1}P_n(x)$ and

$$Z_n := \left\{ \int_0^1 P_n(x)^2 \mu_N(dx) \right\}^{1/2} = \frac{2^{1-N/2}\sqrt{\pi\Gamma(n+N-1)}}{\sqrt{n!(n+(N-1)/2)}\Gamma((N-1)/2)}.$$

As a result, the transition density $p_1(t,x,y)$ of $\hat{\rho}(t)$ with respect to μ_N is given by

$$p_1(t,x,y) = \sum_{n=0}^{\infty} e^{-n(n+N-1)t/2}\bar{P}_n(x)\bar{P}_n(y), \tag{4.6}$$

where the sum converges in $L^2(\mu_N \otimes \mu_N)$. We claim that the infinite sum in the right hand side of (4.6) converges uniformly in x and y. Recall that the Gegenbauer polynomial P_n satisfies the following recursion relation:

$$P_n(x) = \frac{2n+N-3}{n}xP_{n-1}(x) - \frac{(n+N-3)}{n}P_{n-2}(x), \tag{4.7}$$
$$P_0(x) = 1, \quad P_1(x) = (N-1)x.$$

By induction, we can easily show that there exist $c_0 > 0$ and $q > 0$ such that

$$\sup_{x\in(-1,1)} |P_n(x)| \le c_0 q^n. \tag{4.8}$$

(for instance, we can dominate the left hand side by $(N-1)4^n \prod_{k=1}^n (1 + |N-3|/k)$). It is not difficult to see that there exists $c_1 > 0$ such that $Z_n^{-1} \le c_1\sqrt{n}$ for all $n \in \mathbb{N}$ since $N \ge m \ge 2$. Then these estimates imply the claim.

Now the reflection principle yields

$$\varphi_t(a) = \mathbb{P}\left[\inf_{0\le s\le t} \rho(s) > 0 \right]$$
$$= \mathbb{P}\left[\inf_{0\le s\le 2Kt/(N-1)} \hat{\rho}(s) > 0 \right]$$
$$= \int_0^1 \left(p_1(\frac{2Kt}{N-1}, \tilde{a}, x) - p_1(\frac{2Kt}{N-1}, -\tilde{a}, x) \right) \mu_N(dx)$$
$$= \sum_{n=0}^{\infty} \int_0^1 e^{-n(n+N-1)Kt/(N-1)} \left(\bar{P}_n(\tilde{a}) - \bar{P}_n(-\tilde{a}) \right) \bar{P}_n(x)\mu_N(dx)$$
$$= 2\sum_{n=0}^{\infty} e^{-(2n+1)(2n+N)Kt/(N-1)} P_{2n+1}^*(\tilde{a}) \int_0^1 P_{2n+1}^*(x)\mu_N(dx).$$

The Rodrigues formula for the Gegenbauer polynomial asserts

$$P_n(x) = \frac{(-2)^n}{n!}\frac{\Gamma(n+(N-1)/2)\Gamma(n+N-1)}{\Gamma((N-1)/2)\Gamma(2n+N-1)}(1-x^2)^{1-N/2}\frac{d^n}{dx^n}(1-x^2)^{n+N/2-1}.$$

By using this formula twice, we obtain

$$\int_0^1 P_{2n+1}(x)\mu_N(dx) = \frac{(-2)^{2n+1}}{(2n+1)!}\frac{\Gamma(2n+(N+1)/2)\Gamma(2n+N)}{\Gamma((N-1)/2)\Gamma(4n+N+1)}\left[\frac{d^{2n}}{dx^{2n}}(1-x^2)^{2n+N/2}\right]_{x=0}^1$$

$$= \frac{(N-1)}{(2n+1)(2n+N)}P_{2n}^{(N+2)}(0),$$

where $P_{2n}^{(N+2)}$ is the $(2n)$-th Gegenbauer polynomial of parameter $(N+1)/2$ (associated with L_{N+2}). By the recursion formula (4.7), we obtain

$$P_{2n}^{(N+2)}(0) = (-1)^n\frac{\Gamma(n+(N+1)/2)}{\Gamma((N+1)/2)n!}.$$

Thus the duplication formula of the Gamma function yields

$$\varphi_t(a) = \sum_{n=0}^\infty e^{-(2n+1)(2n+N)Kt/(N-1)}\frac{(-1)^n 2(N-1)\Gamma(n+(N+1)/2)}{Z_{2n+1}^2(2n+1)(2n+N)\Gamma((N+1)/2)n!}P_{2n+1}(\tilde{a})$$

$$= \sum_{n=0}^\infty e^{-(2n+1)(2n+N)Kt/(N-1)}\frac{(-1)^n(4n+N+1)}{\pi(2n+N)}B\left(\frac{N-1}{2}, n+\frac{1}{2}\right)P_{2n+1}(\tilde{a}).$$

This is nothing but the desired identity. $\qquad\square$

Based on expressions of $\varphi_t(a)$ in Lemma 4.2, Corollary 4.8 and Lemma 4.9, we will obtain the asymptotic behavior of $\varphi_t(a)$ as $t \to \infty$ in the following corollary:

Corollary 4.10 *The following convergence holds compact uniformly in $a \in [0, \bar{R})$:*

(i) *When $N = \infty$ and $K \geq 0$, or $N < \infty$ and $K = 0$,*

$$\lim_{t\to\infty}\sqrt{\eta_K(t)}\varphi_t(a) = \frac{a}{4\sqrt{\pi}}.$$

In addition, this is an increasing limit.

(ii) *When $N = \infty$ and $K < 0$,*

$$\lim_{t\to\infty}\varphi_t(a) = \chi\left(\frac{a\sqrt{-K}}{2}\right).$$

(iii) *When $N < \infty$ and $K > 0$,*

$$\lim_{t\to\infty}e^{NKt/(N-1)}\varphi_t(a) = \frac{(N^2-1)}{\pi N}B\left(\frac{N-1}{2}, \frac{1}{2}\right)\sin\left(\frac{1}{2}\sqrt{\frac{K}{N-1}}a\right).$$

In addition, $\sup\{e^{NKt/(N-1)}\varphi_t(a) \mid t \geq 1, a \in [0, \bar{R}]\} < \infty$.

(iv) *When $N < \infty$ and $K < 0$,*

$$\lim_{t\to\infty} \varphi_t(a) = \frac{1}{\Gamma((N-1)/2)} \int_0^\infty \chi\left(\sqrt{\frac{-Ku}{2(N-1)}} s_{K/(N-1)}\left(\frac{a}{2}\right)\right) u^{(N-3)/2} e^{-u} du.$$
(4.9)

Proof. (i) (ii) The convergence easily follows by elementary calculus. The monotonicity in t in (i) follows from the concavity of χ. In both cases, the Dini theorem ensures the uniformity of the convergence on each compact set.

(iii) The computation of the limit as well as the uniformity on $[0, \bar{R}]$ and the finiteness of the supremum directly follows from Proposition 4.9 and (4.8).

(iv) By (4.5) and the monotone convergence theorem,

$$\lim_{t\to\infty} \varphi_t(a) = \mathbb{E}\left[\chi\left(\frac{1}{2\sqrt{2}} s_{K/(N-1)}\left(\frac{a}{2}\right)\left(\int_0^\infty \theta'(s)^2 ds\right)^{-1/2}\right)\right].$$
(4.10)

Then the distribution $\int_0^\infty \theta'(s)^2 ds$ can be described with the aid of [26, Theorem 6.2] (also see references therein) to obtain (4.9). Since the convergence in (4.10) is monotone, the compact uniformity of the convergence follows from the Dini theorem. $\qquad\square$

5   Monotonicity of transportation costs

Based on Proposition 4.5 and Corollary 4.6, we will show some continuity properties for $\varphi_t(a)$ and $\mathcal{T}_{\varphi_t(d)}$ with respect to t in the following two lemmata.

Lemma 5.1 *Let $c_n : M \times M \to [0, \infty)$ be a family of continuous functions converging to $c : M \times M \to [0, \infty)$ pointwisely. Let $\mu, \nu \in \mathcal{P}(M)$.*

(i) *If $\sup_{n,x,y} c_n(x,y) < \infty$ or c_n is nondecreasing in n, then*

$$\limsup_{n\to\infty} \mathcal{T}_{c_n}(\mu, \nu) \leq \mathcal{T}_c(\mu, \nu).$$

(ii) *If the convergence $c_n \to c$ is uniform on each compact set or c_n is nondecreasing in n, then*

$$\liminf_{n\to\infty} \mathcal{T}_{c_n}(\mu, \nu) \geq \mathcal{T}_c(\mu, \nu).$$

Proof. (i) Under the assumption on c_n, for each $\pi \in \Pi(\mu, \nu)$,

$$\limsup_{n\to\infty} \mathcal{T}_{c_n}(\mu, \nu) \leq \limsup_{n\to\infty} \int_{M\times M} c_n \, d\pi = \int_{M\times M} c \, d\pi.$$

Thus the assertion holds by taking infimum over $\pi \in \Pi(\mu, \nu)$.

(ii) Take a subsequence $(c_{n_k})_k$ of $(c_n)_n$ so that

$$\lim_{k\to\infty} \mathcal{T}_{c_{n_k}}(\mu, \nu) = \liminf_{n\to\infty} \mathcal{T}_{c_n}(\mu, \nu).$$

23

Since $\Pi(\mu, \nu)$ is compact and c_n is continuous and nonnegative, a usual variational argument implies that there is a minimizer of $\mathcal{T}_{c_{n_k}}(\mu, \nu)$, i.e. there exists $\pi_k \in \Pi(\mu, \nu)$ such that $\mathcal{T}_{c_{n_k}}(\mu, \nu) = \int_{M \times M} c_{n_k} d\pi_k$. We may assume that π_k converges as $k \to \infty$ by taking a subsequence if necessary. We denote the limit by π_∞.

First we consider the case that c_n converges to c compact uniformly. Take $\varepsilon > 0$ and choose a compact set $K \subset M \times M$ such that $\pi_k(K^c) < \varepsilon$. Then, for any $R > 0$, the assumption on c_n implies

$$
\begin{aligned}
\lim_{k \to \infty} \mathcal{T}_{c_{n_k}}(\mu, \nu) &\geq \liminf_{k \to \infty} \int_K c_{n_k} \wedge R \, d\pi_k \\
&\geq \liminf_{k \to \infty} \int_K c \wedge R \, d\pi_k - \varepsilon \\
&\geq \liminf_{k \to \infty} \int_M c \wedge R \, d\pi_k - (R+1)\varepsilon. \\
&= \int_M c \wedge R \, d\pi_\infty - (R+1)\varepsilon.
\end{aligned}
$$

Thus, by taking $\varepsilon \downarrow 0$ and $R \uparrow \infty$, we obtain

$$
\lim_{k \to \infty} \mathcal{T}_{c_{n_k}}(\mu, \nu) \geq \int_M c \, d\pi_\infty \geq \mathcal{T}_c(\mu, \nu)
$$

and hence the assertion holds.

Next we consider the case that c_n is nondecreasing in n. Then, for $k \in \mathbb{N}$,

$$
\int_{M \times M} c_{n_k} d\pi_\infty \leq \liminf_{l \to \infty} \int_{M \times M} c_{n_k} d\pi_l \leq \liminf_{l \to \infty} \int_{M \times M} c_{n_l} d\pi_l = \liminf_{n \to \infty} \mathcal{T}_{c_n}(\mu, \nu).
$$

By taking $k \to \infty$, the monotone convergence theorem implies that

$$
\mathcal{T}_c(\mu, \nu) \leq \limsup_{k \to \infty} \int_{M \times M} c_{n_k} d\pi_\infty.
$$

Thus, the conclusion follows by combining these two estimates. $\qquad\square$

For later use, we will state the following lemma in a slightly more general form than what we will use in the proof of Theorem 2.3.

Lemma 5.2 *Let $(\mu_s)_{s \in [0,\infty)}$ and $(\nu_s)_{s \in [0,\infty)}$ be families of probability measures on M which is continuous in s with respect to the topology of weak convergence. For $t > 0$, let $\hat{d} : [0, t] \times M \times M \to [0, \infty)$ be a continuous function such that $\hat{d}(s, \cdot, \cdot)$ is a distance function on M for each $s \in [0, t]$. Then $s \mapsto \mathcal{T}_{\varphi_{t-s}(\hat{d}(s, \cdot, \cdot))}(\mu_s, \nu_s)$ is continuous on $[0, t)$ and lower semi-continuous at t.*

Proof. For simplicity of notations, we denote $\hat{d}(s, x, y)$ by $d_s(x, y)$. Let $s_0 \in [0, t)$ and take a decreasing sequence $(s_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} s_n = s_0$. Let $\varepsilon > 0$. Since $(\mu_{s_n})_{n \in \mathbb{N}}$ and $(\nu_{s_n})_{n \in \mathbb{N}}$ are tight in $\mathcal{P}(M)$, there exist a compact set $K \subset M$ such that $\pi((K \times K)^c) < \varepsilon$

for any $\pi \in \bigcup_{n\in\mathbb{N}} \Pi(\mu_{s_n}, \nu_{s_n})$. Since $\varphi_{t-s_n}(a)$ is nonincreasing in n, the Dini theorem yields that $\varphi_{t-s_n}(d_{s_0})$ converges to $\varphi_{t-s_0}(d_{s_0})$ as $n \to \infty$ uniformly on $K \times K$. By Corollary 4.6,

$$|\varphi_{t-s_n}(d_{s_n}(x,y)) - \varphi_{t-s_n}(d_{s_0}(x,y))| \leq \varphi_{t-s_n}(|d_{s_n}(x,y) - d_{s_0}(x,y)|).$$

By the assumption on d_s, we have

$$\lim_{n\to\infty} \sup_{x,y\in K} |d_{s_n}(x,y) - d_{s_0}(x,y)| = 0.$$

By combining these estimates, we obtain

$$\limsup_{n\to\infty} \left| \mathcal{T}_{\varphi_{t-s_n}(d_{s_n})}(\mu_{s_n}, \nu_{s_n}) - \mathcal{T}_{\varphi_{t-s_0}(d_{s_0})}(\mu_{s_n}, \nu_{s_n}) \right|$$

$$\leq \limsup_{n\to\infty} \left(\sup_{x,y\in K} |\varphi_{t-s_n}(d_{s_0}(x,y)) - \varphi_{t-s_0}(d_{s_0}(x,y))| \right.$$

$$\left. + \sup_{x,y\in K} \varphi_{t-s_1}\left(|d_{s_n}(x,y) - d_{s_0}(x,y)|\right) \right) + \varepsilon$$

$$= \varepsilon. \tag{5.1}$$

By virtue of Corollary 4.6 and [34, Theorem 7.12], the weak convergences $\mu_{s_n} \to \mu_{s_0}$ and $\nu_{s_n} \to \nu_{s_0}$ imply that $\mathcal{T}_{\varphi_{s_0}(d_{s_0})}(\mu_{s_n}, \nu_{s_n})$ converges to $\mathcal{T}_{\varphi_{s_0}(d_{s_0})}(\mu_{s_0}, \nu_{s_0})$. By combining this fact with (5.1), we obtain

$$\lim_{n\to\infty} \mathcal{T}_{\varphi_{t-s_n}(d_{s_n})}(\mu_{s_n}, \nu_{s_n}) = \mathcal{T}_{\varphi_{t-s_0}(d_{s_0})}(\mu_{s_0}, \nu_{s_0}).$$

It proves that $\mathcal{T}_{\varphi_{t-s}(d_s)}(\mu_s, \nu_s)$ is right-continuous at s_0. In a similar way, we can show the left-continuity of $\mathcal{T}_{\varphi_{t-s}(d_s)}(\mu_s, \nu_s)$ at s_0. Finally we will show the lower semi-continuity at t. Since $\varphi_r(a)$ is nonincreasing in r, for $t' > t$, we have

$$\liminf_{s\uparrow t} \mathcal{T}_{\varphi_{t-s}(d_s)}(\mu_s, \nu_s) \geq \lim_{s\uparrow t} \mathcal{T}_{\varphi_{t'-s}(d_s)}(\mu_s, \nu_s) = \mathcal{T}_{\varphi_{t'-t}(d_t)}(\mu_t, \nu_t).$$

Hence the conclusion follows from Lemma 5.1 (ii) by letting $t' \downarrow t$. $\qquad\square$

Proof of Theorem 2.3. Recall that, for $t' > 0$, $s' \geq 0$ and $x_1, x_2 \in M$, Theorem 2.2 yields

$$\mathcal{T}_{\varphi_{t'}(d)}(\mathbb{P}_{x_1} \circ X(s')^{-1}, \mathbb{P}_{x_2} \circ X(s')^{-1}) \leq \varphi_{t'+s'}(d(x_1,x_2)). \tag{5.2}$$

Let $0 \leq s_1 \leq s_2 < t$. For each $y_1, y_2 \in M$, take $\pi^{y_1 y_2}_{s_2-s_1} \in \Pi(\mathbb{P}_{y_1} \circ X(s_2 - s_1)^{-1}, \mathbb{P}_{y_2} \circ X(s_2 - s_1)^{-1})$ so that

$$\mathcal{T}_{\varphi_{t-s_2}(d)}(\mathbb{P}_{y_1} \circ X(s_2 - s_1)^{-1}, \mathbb{P}_{y_2} \circ X(s_2 - s_1)^{-1}) = \int_{M\times M} \varphi_{t-s_2}(d)\, d\pi^{y_1 y_2}_{s_2-s_1}.$$

We can choose $\pi^{y_1 y_2}_{s_2-s_1}$ so that $(y_1, y_2) \mapsto \pi^{y_1 y_2}_{s_2-s_1}$ is measurable (see [35, Corollary 5.22], for instance). Let us take a minimizer $\pi \in \Pi(\mu^{(1)}_{s_1}, \mu^{(2)}_{s_1})$ of $\mathcal{T}_{\varphi_{t-s_1}(d)}(\mu^{(1)}_{s_1}, \mu^{(2)}_{s_1})$ and define $\pi^* \in \Pi(\mu^{(1)}_{s_2}, \mu^{(2)}_{s_2})$ by

$$\pi^*(A) := \int_{M\times M} \pi^{y_1 y_2}_{s_2-s_1}(A)\pi(dy_1 dy_2).$$

Then, by applying (5.2) with $t' = t - s_2$ and $s' = s_2 - s_1$, we obtain

$$
\begin{aligned}
\mathcal{T}_{\varphi_{t-s_2}(d)}(\mu_{s_2}^{(1)}, \mu_{s_2}^{(2)}) &\leq \int_{M \times M} \varphi_{t-s_2}(d)\, d\pi^* \\
&= \int_{M \times M} \mathcal{T}_{\varphi_{t-s_2}(d)}(\mathbb{P}_{y_1} \circ X(s_2 - s_1)^{-1}, \mathbb{P}_{y_2} \circ X(s_2 - s_1)^{-1}) \pi(dy_1 dy_2) \\
&\leq \int_{M \times M} \varphi_{t-s_1}(d(y_1, y_2)) \pi(dy_1 dy_2) \\
&= \mathcal{T}_{\varphi_{t-s_1}(d)}(\mu_{s_1}^{(1)}, \mu_{s_1}^{(2)}).
\end{aligned}
$$

Thus the assertion holds when $t > s_2$. When $t = s_2$, the assertion follows by taking $s_2 \uparrow t$ together with Lemma 5.2 with $\hat{d}(t, x, y) := d(x, y)$. $\qquad\square$

In Theorem 2.3, the cost function $\varphi_{t-s}(d)$ depends on time parameter t. Thus it seems to be natural to consider the limit $t \to \infty$, under a suitable scaling if necessary. We can realize it by combining Corollary 4.10 with Theorem 2.3 with the aid of Lemma 5.1. Then we obtain the following monotonicity of transportation costs:

Corollary 5.3 *Let us define* $\Theta = \Theta_{K,N} : [0, \bar{R}) \to [0, \infty)$ *and* $\kappa = \kappa(K, N) \in \mathbb{R}$ *by*

$$
\Theta_{K,N}(a) := \begin{cases}
a & (K = 0), \\
a & (N = \infty \text{ and } K > 0), \\
\chi\left(\dfrac{a\sqrt{-K}}{2}\right) & (N = \infty \text{ and } K < 0), \\
\sin\left(\dfrac{1}{2}\sqrt{\dfrac{K}{N-1}}\,a\right) & (N < \infty \text{ and } K > 0), \\
\displaystyle\int_0^\infty \chi\left(\sqrt{\dfrac{-Ku}{2(N-1)}}\, s_{K/(N-1)}\left(\dfrac{a}{2}\right)\right) u^{(N-3)/2} e^{-u} du & (N < \infty \text{ and } K < 0),
\end{cases}
$$

$$
\kappa(K, N) := \begin{cases}
K \vee 0 & (N = \infty), \\
\dfrac{NK}{N-1} \vee 0 & (N < \infty).
\end{cases}
$$

For $i = 1, 2$ *and* $\mu^{(i)} \in \mathcal{P}(M)$, *let* $\mu_t^{(i)}$ *be the distribution of* $X(t)$ *with the initial distribution* $\mu^{(i)}$. *Then* $e^{\kappa s} \mathcal{T}_{\Theta(d)}(\mu_s^{(1)}, \mu_s^{(2)})$ *is nonincreasing in* s.

When $K = 0$ or $N = \infty$ and $K > 0$, what the last corollary states is nothing but the L^1-Wasserstein contraction. When $N < \infty$ and $K > 0$, what we obtained is essentially well-known (see [37] for the statement formulated in terms of optimal transport theory). Thus the most interesting assertion is in the case $K < 0$. In the usual L^p-Wasserstein contraction in (1.2), The upper bound grows exponentially fast as time increases when $K < 0$. The last corollary says that a nonincreasing property still holds even when $K < 0$ by choosing a cost function appropriately.

For a bounded and measurable function $f : M \to \mathbb{R}$, we define the action of the diffusion semigroup $P_t f$ by $P_t f(x) := \mathbb{E}_x[f(X(t))]$. We denote the dual action of P_t to $\mathcal{P}(M)$ by P_t^*. That is,

$$P_t^* \mu(A) = \int_M \mathbb{P}_x[X(t) \in A] \, \mu(dx).$$

Since $\varphi_t(d(x, y))$ is a distance function by Corollary 4.6, the Kantorovich-Rubinstein duality easily implies the following (cf. [21, 30]):

Theorem 6.1 *Given $t, s \geq 0$, the following are equivalent:*

(i) *For $\mu_1, \mu_2 \in \mathcal{P}(M)$,*

$$\mathcal{T}_{\varphi_t^{K,N}(d)}(P_s^* \mu_1, P_s^* \mu_2) \leq \mathcal{T}_{\varphi_{t+s}^{K,N}(d)}(\mu_1, \mu_2).$$

(ii) *For any $\varphi_t^{K,N}(d)$-Lipschitz function f on M,*

$$\sup_{x \neq y} \frac{|P_s f(x) - P_s f(y)|}{\varphi_{t+s}^{K,N}(d(x, y))} \leq \sup_{x \neq y} \frac{|f(x) - f(y)|}{\varphi_t^{K,N}(d(x, y))}.$$

The condition (i) in the last theorem comes from the consequence of Theorem 2.3. Note that, in the condition (ii), those supremums may be attained at $(x, y) \in M \times M$ with $d(x, y) > 0$ since $\varphi_t(d)$ is not a geodesic distance. As an easy consequence of Theorem 6.1, we obtain the following gradient estimate.

Corollary 6.2 *Under Assumption 1, we have*

$$\|\nabla P_t f\|_\infty \leq \varphi_t'(0) \operatorname{osc}(f)$$

for any bounded measurable function f on M.

Recall that an expression of $\varphi_t'(0)$ is given in Proposition 4.5. Note that a gradient estimate like in Corollary 6.2 also follows from the reverse Poincaré inequality (see [4, 22] for instance; also see [2]). When $K \geq 0$, Corollary 6.2 and Proposition 4.5 (v) easily imply the Liouville property, that is, there are no nonconstant bounded $\mathcal{L}$-harmonic functions, by taking f as a bounded harmonic function (so that $P_t f = f$) and $t \to \infty$.

Proof. Theorem 2.3 tells us that the condition (i) holds with $t = 0$ under Assumption 1. Then the definition of φ_0 yields

$$\sup_{x \neq y} \frac{|f(x) - f(y)|}{\varphi_0^{K,N}(d(x, y))} = \operatorname{osc}(f).$$

Recall that the differentiability of $\varphi_s^{K,N}$ at 0 is given in Proposition 4.5 (v). Thus Theorem 6.1 implies that

$$\frac{1}{(\varphi_s^{K,N})'(0)} \|\nabla P_s f\|_\infty \leq \sup_{x \neq y} \frac{|P_s f(x) - P_s f(y)|}{\varphi_s(d(x, y))} \leq \operatorname{osc}(f)$$

and hence the conclusion holds. $\square$

In this section we consider a sequence of Riemannian manifolds (M_n, g_n) $(n \in \mathbb{N})$. By technical reasons, we restrict ourselves into the case that each M_n is compact. Let f_n be a positive C^1-function on M_n and $Z_n := \nabla f_n$ for $n \in \mathbb{N}$. Suppose that, given $N < \infty$ and K, (M_n, g_n) and Z_n satisfies Assumption 1 for all $n \in \mathbb{N}$ where the parameters N and K are independent of n. Let vol_n be the Riemannian volume measure on (M_n, g_n) and set $\nu_n = \mathrm{e}^{f_n} \mathrm{vol}_n$. Under Assumption 1, the metric measure space (M_n, d_n, ν_n) satisfies the curvature-dimension condition $\mathsf{CD}(K, \infty)$ (see [25, 31, 32]). Thus the gradient flow of the relative entropy functional Ent_{ν_n} on L^2-Wasserstein space over (M_n, d_n, ν_n) is identified with the gradient flow of the Dirichlet energy functional on $L^2(M_n, \nu_n)$ (see [1, 8, 11]).

Definition 7.1 (i) *Let (M_1, d_1) and (M_2, d_2) be metric spaces. For $\varepsilon > 0$, we call $f : M_1 \to M_2$ an ε-isometry if the following hold:*

$$\sup_{x,y \in M_1} |d_1(x, y) - d_2(f(x), f(y))| \leq \varepsilon, \qquad \sup_{y \in M_2} d_2(y, f(M_1)) \leq \varepsilon.$$

(ii) *Let $((M_n, d_n, \nu_n))_{n \in \mathbb{N}}$ and (M, d, ν) be metric measure spaces. We say (M_n, d_n, ν_n) converges to (M, d, ν) as $n \to \infty$ in the measured Gromov-Hausdorff sense if there exist $\varepsilon_n > 0$ $(n \in \mathbb{N})$ with $\lim_{n \to \infty} \varepsilon_n = 0$ and ε_n-isometry $f_n : M_n \to M$ so that $f_n^\# \nu_n$ converges to $f^\# \nu$ in the vague topology.*

In the sequel, we assume that (M_n, d_n, ν_n) converges to a compact metric measure space (M, d, ν) in the measured Gromov-Hausdorff sense via ε_n-isometries $f_n : M_n \to M$. Note that, in this framework, the convergence with respect to the measured Gromov-Hausdorff distance is equivalent to the convergence with respect to the distance $\mathbf{D}$ introduced in [31] (see [31, Subsection 3.4]). Under the assumption, (M, d, ν) satisfies $\mathsf{CD}(K, \infty)$ again (see [2, 25, 31]). Thus, for $\mu_0 \in \mathcal{P}(M)$ with $\mathrm{Ent}_\nu(\mu_0) < \infty$, there exists a unique gradient curve μ_t of Ent_ν on $\mathcal{P}(M)$ starting from μ_0 (see [1, 2, 10]).

The following theorem asserts that these gradient curves enjoy the same monotonicity as shown in Theorem 2.3:

Theorem 7.2 *For $i = 1, 2$, let $\mu_0^{(i)} \in \mathcal{P}(M)$ with $\mathrm{Ent}_\nu(\mu_0^{(i)}) < \infty$ and $\mu_t^{(i)}$ a gradient curve of Ent_ν with initial distribution $\mu_0^{(i)}$. Then, for any $t \in [0, \infty)$, $\mathcal{T}_{\varphi_{t-s}(d)}(\mu_s^{(1)}, \mu_s^{(2)})$ is a nonincreasing function of $s \in [0, t]$.*

Proof. By virtue of Lemma 5.2, it suffices to show the assertion for $s \in (0, t)$. For $i = 1, 2$, there exists $\mu_0^{(i,n)} \in \mathcal{P}(M_n)$ for $n \in \mathbb{N}$ such that $\mathrm{Ent}_{\nu_n}(\mu_0^{(i,n)}) < \infty$ and that $f_n^\# \mu_0^{(i,n)}$ converges to $\mu_0^{(i)}$ by following an argument in the proof of [25, Theorem 4.15]. Let $\mu_t^{(i,n)}$ be the gradient curve of Ent_{ν_n} on $\mathcal{P}(M_n)$ with the initial distribution $\mu_0^{(i,n)}$. Then, by virtue of [10, Theorem 21], $f_n^\# \mu_t^{(i,n)}$ converges to $\mu_t^{(i)}$ for $i = 1, 2$ and $t > 0$.

We claim that for each $s \in (0, t)$,

$$\lim_{n \to \infty} \left(\mathcal{T}_{\varphi_{t-s}(d_n)}(\mu_s^{(1,n)}, \mu_s^{(2,n)}) - \mathcal{T}_{\varphi_{t-s}(d)}(f_n^\# \mu_s^{(1,n)}, f_n^\# \mu_s^{(2,n)}) \right) = 0. \tag{7.1}$$

Take $\pi^{(n)} \in \Pi(\mu_s^{(1,n)}, \mu_s^{(2,n)})$ and set $\tilde{\pi}^{(n)} := (f_n \times f_n)^{\#} \pi^{(n)}$. Then we can easily see that $\tilde{\pi}^{(n)} \in \Pi(f_n^{\#} \mu_s^{(1,n)}, f_n^{\#} \mu_s^{(2,n)})$ holds. Since f_n is ε_n-isometry and $\varphi_{t-s}(\cdot)$ is nondecreasing,

$$
\begin{aligned}
\mathcal{T}_{\varphi_{t-s}(d)}(f_n^{\#} \mu_s^{(1,n)}, f_n^{\#} \mu_s^{(2,n)}) &\leq \int_{M \times M} \varphi_{t-s}(d(x,y)) \tilde{\pi}^{(n)}(dxdy) \\
&= \int_{M \times M} \varphi_{t-s}(d(f_n(x), f_n(y))) \pi^{(n)}(dxdy) \\
&= \int_{M \times M} \varphi_{t-s}(d_n(x,y) + \varepsilon_n) \pi^{(n)}(dxdy). \quad (7.2)
\end{aligned}
$$

Since our choice of $\pi^{(n)} \in \Pi(\mu_s^{(1,n)}, \mu_s^{(2,n)})$ can be arbitrary, (7.2) and Corollary 4.6 yield

$$
\mathcal{T}_{\varphi_{t-s}(d)}(f_n^{\#} \mu_s^{(1,n)}, f_n^{\#} \mu_s^{(2,n)}) \leq \mathcal{T}_{\varphi_{t-s}(d_n)}(\mu_s^{(1,n)}, \mu_s^{(2,n)}) + \varphi_{t-s}(\varepsilon_n). \quad (7.3)
$$

To complete the proof of the claim, let us take an approximate inverse g_n for each $n \in \mathbb{N}$, that is, $g_n : M \to M_n$ satisfies

$$
\lim_{n \to \infty} \sup_{x \in M_n} d(x, g_n(f_n(x))) = 0, \qquad \lim_{n \to \infty} \sup_{x \in M} d_n(x, f_n(g_n(x))) = 0.
$$

We may assume g_n is ε_n'-isometry for some ε_n' with $\lim_{n \to \infty} \varepsilon_n' = 0$ without loss of generality. By a similar argument as what we used to obtain (7.3),

$$
\begin{aligned}
\mathcal{T}_{\varphi_{t-s}(d_n)} \left((g_n \circ f_n)^{\#} \mu_s^{(1,n)}, (g_n \circ f_n)^{\#} \mu_s^{(2,n)} \right) & \\
&\leq \mathcal{T}_{\varphi_{t-s}(d)} \left(f_n^{\#} \mu_s^{(1,n)}, f_n^{\#} \mu_s^{(2,n)} \right) + \varphi_{t-s}(\varepsilon_n'). \quad (7.4)
\end{aligned}
$$

Since $(\mathrm{id} \times (g_n \circ f_n))^{\#} \mu_s^{(i,n)} \in \Pi(\mu_s^{(i,n)}, (g_n \circ f_n)^{\#} \mu_s^{(i,n)})$,

$$
\begin{aligned}
\mathcal{T}_{\varphi_{t-s}(d_n)} \left(\mu_s^{(i,n)}, (g_n \circ f_n)^{\#} \mu_s^{(i,n)} \right) &\leq \int_{M_n \times M_n} d_n(x, g_n(f_n(x))) \mu^{(i,n)}(dx) \\
&\leq \sup_{x \in M_n} d_n(x, g_n(f_n(x)))
\end{aligned}
$$

for $i = 1, 2$. By combining this estimate with (7.4), we obtain

$$
\begin{aligned}
\mathcal{T}_{\varphi_{t-s}(d_n)} \left(\mu_s^{(1,n)}, \mu_s^{(2,n)} \right) & \\
&\leq \mathcal{T}_{\varphi_{t-s}(d)} \left(f_n^{\#} \mu_s^{(1,n)}, f_n^{\#} \mu_s^{(2,n)} \right) + 2 \sup_{x \in M_n} d_n(x, g_n(f_n(x))) + \varphi_{t-s}(\varepsilon_n'). \quad (7.5)
\end{aligned}
$$

Hence (7.3) and (7.5) imply the claim since $\varphi_{t-s}(\cdot)$ is continuous.

By Corollary 4.6 and [34, Theorem 7.12], $\mathcal{T}_{\varphi_{t-s}(d)}(f_n^{\#} \mu_s^{(i,n)}, \mu_s^{(i)})$ converges to 0 as $n \to \infty$ for $i = 1, 2$. Hence (7.1) yields $\lim_{n \to \infty} \mathcal{T}_{\varphi_{t-s}(d_n)}(\mu_s^{(1,n)}, \mu_s^{(2,n)}) = \mathcal{T}_{\varphi_{t-s}(d)}(\mu_s^{(1)}, \mu_s^{(2)})$. Since $\mathcal{T}_{\varphi_{t-s}(d_n)}(\mu_s^{(1,n)}, \mu_s^{(2,n)})$ is nonincreasing in s by Theorem 2.3, the conclusion holds. $\square$

8 Time-dependent metrics

Let $(g(t))_{t\in[T_1,T_2]}$ be a family of smooth complete Riemannian metrics on M depending smoothly in t. Let $Z(t)$ be a time-dependent vector field on M depending continuously in t and consider the time-inhomogeneous diffusion process $((X(t))_{t\in[T_1,T_2]},(\mathbb{P}_x)_{x\in M})$ generated by $\mathcal{L}_t := \Delta_{g(t)} + Z(t)$. The following assumption corresponds to Assumption 1 with $N = \infty$:

Assumption 2 *Given $K \in \mathbb{R}$, the following holds for each t:*

$$(\nabla Z(t))^\flat + \frac{1}{2}\partial_t g(t) \leq \mathrm{Ric}_{g(t)} - Kg(t).$$

An important example of the time-dependent metrics $g(t)$ satisfying Assumption 2 is the backward Ricci flow, that is,

$$\frac{1}{2}\partial_t g(t) = \mathrm{Ric}_{g(t)}.$$

Under Assumption 2, the coupling by reflection of $X(t)$ is already studied in [17] via the approximation by geodesic random walks (The notation in [17] looks slightly different since we considered the diffusion process generated by $\Delta_{g(t)}/2 + Z(t)$ there). By modifying arguments in previous sections, we can obtain the results corresponding to Theorem 2.2, Theorem 2.3, Corollary 2.4 and Corollary 5.3 with $N = \infty$ by replacing d which measures the distribution at s with $d_{g(T_1+s)}$. For example, the conclusion of the statement corresponding to Theorem 2.3 is as follows: Let $\mu_s^{(i)}$ be the distribution of $X(t)$ at $t = s + T_1$ with initial distribution $\mu_{T_1}^{(i)}$ for $i = 1, 2$. Then, for $t \geq s_2 > s_1 \geq 0$,

$$\mathcal{T}_{\varphi_{t-s_2}(d_{g(s_2+T_1)})}(\mu_{s_2+T_1}^{(1)}, \mu_{s_2+T_1}^{(2)}) \leq \mathcal{T}_{\varphi_{t-s_1}(d_{g(s_1+T_1)})}(\mu_{s_1+T_1}^{(1)}, \mu_{s_1+T_1}^{(2)}). \tag{8.1}$$

For reader's convenience, let us explain briefly why the time derivative with respect to the metric appears in Assumption 2. When we follow the argument in the time-independent metric case in Proposition 3.1, we consider $d_{g(t_n^\alpha)}(\mathbf{X}^\alpha(t_n^\alpha))$ instead of $r^\alpha(n) = d(\mathbf{X}^\alpha(t_n^\alpha))$. Then, in the Taylor expansion in the proof of Lemma 3.2, there appears the time derivative of $d_{g(t)}$ as an additional term. It can be described in terms of the time derivative of $g(t)$. Then our condition in Assumption 2 will be used to implement this additional term into the lower bound of Bakry-Émery Ricci tensor. For more details, see [17].

Note that, by [17, Lemma 2.5], $\hat{d}(t, x, y) := d_{g(t+T_1)}(x, y)$ satisfies the assumption of Lemma 5.2. This fact will be used to complete the proof of (8.1) when $t = s_2$.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, preprint. arXiv:1106.2090.

[2] _____, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, preprint. arXiv:1109.0222.

[3] _____, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Birkhäuser Verlag, Basel, 2008.

[4] D. Bakry, *On Sobolev and logarithmic Sobolev inequalities for Markov semigroups*, New trends in stochastic analysis (Charingworth, 1994), World Sci. Publ. River Edge, NJ, 1997, pp. 43–75.

[5] D. Burago, Yu. Burago, and S. Ivanov, *A course in metric geometry*, Graduate studies in mathematics, 33, American mathematical society, Providence, RI, 2001.

[6] M. Cranston, *Gradient estimates on manifolds using coupling*, J. Funct. Anal. **99** (1991), no. 1, 110–124.

[7] A. Eberle, *Reflection coupling and Wasserstein contractivity without convexity*, C. R. Math. Acad. Sci. Paris **349** (2011), no. 19–20, 1101–1104.

[8] M. Erbar, *The heat equation on manifolds as a gradient flow in the Wasserstein space*, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), no. 1, 1–23.

[9] W. Gangbo and R.J. McCann, *The geometry of optimal transportation*, Acta Math. **177** (1996), no. 2, 113–161.

[10] N. Gigli, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations **39** (2010), no. 1–2, 101–120.

[11] N. Gigli, K. Kuwada, and S. Ohta, *Heat flow on Alexandrov spaces*, to appear in Comm. Pure. Appl. Math.

[12] E. P. Hsu and K.-Th. Sturm, *Maximal coupling of Euclidean Brownian motions*, Preprint.

[13] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes*, second ed., North-Holland Mathematical Library, 24, North-Holland Publishing Co., Amsterdam-New York; Kodansha, Ltd., Tokyo, 1989.

[14] K. Itô and H.P. McKean, *Diffusion processes and their sample paths*, second printing ed., Die Grundlehren der mathematischen Wissenschaften, Band 125, Springer-Verlag, Berlin-New York, 1974.

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

[16] W.S. Kendall, *From stochastic parallel transport to harmonic maps*, New directions in Dirichlet forms, AMS/IP Studies in Advanced Mathematics, 8, Amer. Math. Soc., Providence, RI; International Press, Cambridge, MA, 1998, pp. 49–115.

[17] K. Kuwada, *Convergence of time-inhomogeneous geodesic random walks and its application to coupling methods*, to appear in Ann. Probab.

[18] ———, *A probabilistic approach to the maximal diameter theorem*, Preprint. available at: http://www.math.ocha.ac.jp/kuwada/papers.html.

[19] ———, *On uniqueness of maximal coupling for diffusion processes with a reflection*, J. Theoret. Probab. **20** (2007), no. 4, 935–957.

[20] ———, *Couplings of the Brownian motion via discrete apporoximation under lower Ricci curvature bounds*, Probabilistic Approach to Geometry (Tokyo), Adv. Stud. Pure Math. 57, Math. Soc. Japan, 2010, pp. 273–292.

[21] ———, *Duality on gradient estimates and Wasserstein controls*, J. Funct. Anal. **258** (2010), no. 11, 3758–3774.

[22] M. Ledoux, *The geometry of Markov diffusion generators*, Ann. Fac. Sci. Toulouse Math. (6) **9** (2000), no. 2, 305–366.

[23] T. Lindvall, *Lectures on the coupling method*, John Wiley & Sons, Chichester and New York, 1992.

[24] T. Lindvall and L.C.G. Rogers, *Coupling of multidimensional diffusions by reflection*, Ann. Probab. **14** (1986), no. 3, 860–872.

[25] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. Math. **169** (2009), no. 3, 903–991.

[26] H. Matsumoto and M. Yor, *Exponential functional of Brownian motion, I.: Probability laws at fixed time*, Probability surveys **2** (2005), 312–347.

[27] ———, *Exponential functional of Brownian motion, II.: Some related diffusion processes*, Probability surveys **2** (2005), 348–384.

[28] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1–2, 101–174.

[29] E. J. Pauwels and L. C. G. Rogers, *Skew-product decompositions of Brownian motions*, Geometry of random motion (Ithaca, N.Y., 1987), Contemp. Math., 73, Amer. Math. Soc., Providence, RI, 1988, pp. 237–262.

[30] K.-Th. Sturm, *A semigroup approach to harmonic maps*, Potential Anal. **23** (2005), no. 3, 225–277.

[31] ———, *On the geometry of metric measure spaces. I*, Acta. Math. **196** (2006), no. 1, 65–131.

[32] ———, *On the geometry of metric measure spaces. II*, Acta. Math. **196** (2006), no. 1, 133–177.

[33] G. Szegő, *Orthogonal polynomials*, fourth ed., Colloquium Publications, Vol. XXIII, American Mathematical Society, Providence, R.I., 1975.

[34] C. Villani, *Topics in optimal transportations*, Graduate studies in mathematics, 58, American mathematical society, Providence, RI, 2003.

[35] ⸻, *Optimal transport, old and new*, Grundlehren der Mathematischen Wissenschaften, 338, Springer-Verlag, 2008.

[36] M.-K. von Renesse and K.-Th. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Comm. Pure. Appl. Math. **58** (2005), no. 7, 923–940.

[37] F.-Y. Wang, *Equivalent semigroup properties for curvature-dimension condition*, preprint. arXiv:1012.5686.

[38] ⸻, *Functional inequalities, Markov semigroups, and spectral theory*, Mathematics Monograph Series 4, Science Press, Beijing, China, 2005.