Available online at www.sciencedirect.com
ScienceDirect
Journal of Functional Analysis 260 (2011) 2742–2766
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

Coupling of Brownian motions and Perelman's $\mathcal{L}$-functional

Kazumasa Kuwada [a,1], Robert Philipowski [b,*]

[a] *Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo 112-8610, Japan*
[b] *Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany*

Received 7 July 2010; accepted 28 January 2011

Communicated by C. Villani

Abstract

We show that on a manifold whose Riemannian metric evolves under backwards Ricci flow two Brownian motions can be coupled in such a way that their normalized $\mathcal{L}$-distance is a supermartingale. As a corollary, we obtain the monotonicity of the transportation cost between two solutions of the heat equation in the case that the cost function is the composition of a concave non-decreasing function and the normalized $\mathcal{L}$-distance. In particular, it provides a new proof of a recent result of Topping [P. Topping, $\mathcal{L}$-optimal transportation for Ricci flow, J. Reine Angew. Math. 636 (2009) 93–122].

Keywords: Ricci flow; $\mathcal{L}$-functional; Brownian motion; Coupling

1. Introduction

Let M be a d-dimensional differentiable manifold, $0 \leqslant \bar{\tau}_1 < \bar{\tau}_2 < T$ and $(g(\tau))_{\tau \in [\bar{\tau}_1, T]}$ a complete backwards Ricci flow on M, i.e. a smooth family of Riemannian metrics satisfying

$$\frac{\partial g}{\partial \tau} = 2 \operatorname{Ric}_{g(\tau)} \tag{1}$$

* Corresponding author.
E-mail addresses: kuwada.kazumasa@ocha.ac.jp (K. Kuwada), philipowski@iam.uni-bonn.de (R. Philipowski).
1 Partially supported by the JSPS fellowship for research abroad.

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2011.01.017

and such that $(M, g(\tau))$ is complete for all $\tau \in [\bar{\tau}_1, T]$. In this situation Perelman [23, Section 7.1] (see also [6, Definition 7.5]) defined the $\mathcal{L}$-functional of a smooth curve $\gamma : [\tau_1, \tau_2] \to M$ (where $\bar{\tau}_1 \leqslant \tau_1 < \tau_2 \leqslant T$) by

$$\mathcal{L}(\gamma) := \int_{\tau_1}^{\tau_2} \sqrt{\tau}\big[\big|\dot{\gamma}(\tau)\big|^2_{g(\tau)} + R_{g(\tau)}\big(\gamma(\tau)\big)\big]\,d\tau,$$

where $R_{g(\tau)}(x)$ is the scalar curvature at x with respect to the metric $g(\tau)$. Let $L(x, \tau_1; y, \tau_2)$ be the $\mathcal{L}$-distance between (x, τ_1) and (y, τ_2) defined by the infimum of $\mathcal{L}(\gamma)$ over smooth curves $\gamma : [\tau_1, \tau_2] \to M$ satisfying $\gamma(\tau_1) = x$ and $\gamma(\tau_2) = y$. The normalized $\mathcal{L}$-distance $\Theta_t(x, y)$ $(t \geqslant 1)$ is given by

$$\Theta_t(x, y) := 2\big(\sqrt{\bar{\tau}_2 t} - \sqrt{\bar{\tau}_1 t}\big)L(x, \bar{\tau}_1 t; y, \bar{\tau}_2 t) - 2d\big(\sqrt{\bar{\tau}_2 t} - \sqrt{\bar{\tau}_1 t}\big)^2.$$

For a measurable function $c : M \times M \to \mathbb{R}$, let us define the transportation cost $\mathcal{T}_c(\mu, \nu)$ between two probability measures μ and ν on M with respect to the cost function c by

$$\mathcal{T}_c(\mu, \nu) := \inf_{\pi} \int_{M \times M} c(x, y)\pi(dx, dy)$$

(the infimum is over all probability measures π on $M \times M$ whose marginals are μ and ν respectively). To study Perelman's work from a different aspect, Topping [30] (see also Lott [19]) showed the following result:

Theorem 1. *(See Theorem 1.1 in [30].) Assume that M is compact and that $\bar{\tau}_1 > 0$. Let $p : [\bar{\tau}_1, T] \times M \to \mathbb{R}_+$ and $q : [\bar{\tau}_2, T] \times M \to \mathbb{R}_+$ be two non-negative unit-mass solutions of the heat equation*

$$\frac{\partial p}{\partial \tau} = \Delta_{g(\tau)} p - Rp,$$

where the term Rp comes from the change in time of the volume element. Then the normalized $\mathcal{L}$-transportation cost $\mathcal{T}_{\Theta_t}(p(\bar{\tau}_1 t, \cdot)\operatorname{vol}_{g(\bar{\tau}_1 t)}, q(\bar{\tau}_2 t, \cdot)\operatorname{vol}_{g(\bar{\tau}_2 t)})$ between the two solutions evaluated at times $\bar{\tau}_1 t$ respectively $\bar{\tau}_2 t$ is a non-increasing function of $t \in [1, T/\bar{\tau}_2]$.

By $g(\tau)$-Brownian motion, we mean the time-inhomogeneous diffusion process whose generator is $\Delta_{g(\tau)}$. As in the time-homogeneous case, the heat distribution $p(\tau, \cdot)\operatorname{vol}_{g(\tau)}$ is expressed as the law of a $g(\tau)$-Brownian motion at time τ. In view of this strong relation between heat equation and Brownian motion, it is natural to ask whether one can couple two Brownian motions on M in such a way that a pathwise analogue of this result holds. The main result of this paper answers it affirmatively as follows:

Theorem 2. *Assume that the Ricci curvature of M is bounded from below uniformly, namely there exists $K \geqslant 0$ such that*

$$\operatorname{Ric}_{g(\tau)} \geqslant -Kg(\tau) \quad \textit{for any } \tau \in [\bar{\tau}_1, T]. \tag{2}$$

Then given any points $x, y \in M$ and any $s \in [1, T/\bar{\tau}_2]$, there exist two coupled $g(\tau)$-Brownian motions $(X_\tau)_{\tau\in[\bar{\tau}_1 s,T]}$ and $(Y_\tau)_{\tau\in[\bar{\tau}_2 s,T]}$ with initial values $X_{\bar{\tau}_1 s} = x$ and $Y_{\bar{\tau}_2 s} = y$ such that the process $(\Theta_t(X_{\bar{\tau}_1 t}, Y_{\bar{\tau}_2 t}))_{t\in[s,T/\bar{\tau}_2]}$ is a supermartingale. In addition, we can take them so that the map $(x, y) \mapsto (X, Y)$ is measurable. In particular, for any $\varphi : \mathbb{R} \to \mathbb{R}$ being concave and non-decreasing, $\mathbb{E}[\varphi(\Theta_t(X_{\bar{\tau}_1 t}, Y_{\bar{\tau}_2 t}))]$ is non-increasing.

Obviously, (2) is satisfied if M is compact. Thus it includes the case of Theorem 1. As a result, Theorem 2 easily implies the following extension of Theorem 1 (see Section 5.3):

Theorem 3. *Assume that* (2) *holds. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be concave and non-decreasing. Then $\mathcal{T}_{\varphi\circ\Theta_t}(p(\bar{\tau}_1 t, \cdot)\,\mathrm{vol}_{g(\bar{\tau}_1 t)}, q(\bar{\tau}_2 t, \cdot)\,\mathrm{vol}_{g(\bar{\tau}_2 t)})$ is non-increasing in $t \in [1, T/\bar{\tau}_2]$ for non-negative unit-mass solutions p and q to the heat equation.*

We prove Theorem 2 by constructing a coupling via approximation of $g(\tau)$-Brownian motions by geodesic random walks as studied in [14]. In the next section, we demonstrate background of the problem, review related results and compare their methods with ours. Since our method superficially looks like a detour compared with other existing coupling arguments, there we explain the reason why we choose that way. To state the idea behind our proof explicitly, we prove Theorem 2 under the assumption that there is no singularity of $\mathcal{L}$-distance in Section 3. Since all technical difficulties are concentrated on the singularity of $\mathcal{L}$-distance, we can study the problem there in more direct way by using stochastic calculus. Some estimates on variations of $\mathcal{L}$-distance are gathered in Section 4. The proof of the full statement of Theorems 2 and 3 will be provided in Section 5.

Before closing this section, we give two remarks on Theorems 2 and 3.

Remark 1. As shown in [16], under backwards Ricci flow $g(\tau)$-Brownian motion cannot explode. Hence $\Theta_t(X_t, Y_t)$ is well defined for all $t \in [s, T/\bar{\tau}_2]$ in Theorem 2. This fact also ensures that $p(\tau, \cdot)\,\mathrm{vol}_{g(\tau)}$ has unit mass whenever it does at the initial time. We implicitly require this property to make $\mathcal{T}_{\varphi\circ\Theta_t}(p(\bar{\tau}_1 t, \cdot)\,\mathrm{vol}_{g(\bar{\tau}_1 t)}, q(\bar{\tau}_2 t, \cdot)\,\mathrm{vol}_{g(\bar{\tau}_2 t)})$ well defined in Theorem 3.

Remark 2. There are plenty of examples of backwards Ricci flow satisfying (2) even when M is non-compact. Indeed, given a metric g_0 on M with bounded curvature tensor, there exists a unique solution to the Ricci flow $\partial_t g(t) = -2\,\mathrm{Ric}_{g(t)}$ with initial condition g_0 satisfying

$$\sup_{x,t} |\mathrm{Rm}_{g(t)}|_{g(t)}(x) < \infty$$

for a short time (see [28] for existence and [5] for uniqueness). Then the corresponding backwards Ricci flow satisfying (2) is obtained by time-reversal.

2. Review and remarks on background of the problem

The Ricci flow was introduced by Hamilton [9]. There he effectively used it to solve the Poincaré conjecture for 3-manifolds with positive Ricci curvature. By following his approach, Perelman [23–25] finally solved the Poincaré conjecture (see also [4,12,22]). There he used $\mathcal{L}$-functional as a crucial tool. At the same stage, he also studied the heat equation in [23] in relation with the geometry of Ricci flows. It suggests that analyzing the heat equation is still an efficient way to investigate geometry of the underlying space even in the time-dependent metric case. This guiding principle has been confirmed in recent developments in this direction. For example, we refer to [36] as one of such developments. In connection with the theory of optimal transportation, McCann and Topping [21] showed the monotonicity of $\mathcal{T}_{\rho^2_{g(\tau)}}(p(\tau,\cdot)\,\mathrm{vol}_{g(\tau)}, q(\tau,\cdot)\,\mathrm{vol}_{g(\tau)})$, where $\rho_{g(\tau)}$ is the $g(\tau)$-Riemannian distance, under backwards Ricci flow on a compact manifold. Topping's result [30] can be regarded as an extension of it to contraction in the normalized $\mathcal{L}$-transportation cost (see [19] also). By taking $\bar{\tau}_2 \to \bar{\tau}_1$, he gave a new proof of the monotonicity of Perelman's $\mathcal{W}$-entropy, which is one of fundamental ingredients in Perelman's work.

A probabilistic approach to these problems is initiated by Arnaudon, Coulibaly and Thalmaier. In [2, Section 4], they sharpened McCann and Topping's result [21] to a pathwise contraction in the following sense: There is a coupling $(X_t, Y_t)_{t\geqslant 0}$ of two Brownian motions starting from $x, y \in X$ respectively such that the $g(t)$-distance between X_t and Y_t is non-increasing in t almost surely. In their approach, probabilistic techniques based on analysis of sample paths made it possible to establish such a pathwise estimate. As an advantage of their result, the pathwise contraction easily yields that $\mathcal{T}_{\varphi\circ\rho_{g(\tau)}}(p(\tau,\cdot)\,\mathrm{vol}_{g(\tau)}, q(\tau,\cdot)\,\mathrm{vol}_{g(\tau)})$ is non-increasing for any non-decreasing φ. As an application of this sharper monotonicity, we can obtain an L^1-gradient estimate of Bakry–Émery type (see [15]) for the heat semigroup. In the time-homogeneous case, this gradient estimate has been known to be very useful in geometric analysis (see e.g. [3,17]). McCann and Topping's result only implies L^2-gradient estimate and it is weaker than the L^1-estimate. (In the time-homogeneous case, it is known that L^2-estimate also implies L^1-estimate (see [3,17,26]). However, to the best of our knowledge, an extension of such equivalence in the time-inhomogeneous case is not yet established.) As another advantage of Arnaudon, Coulibaly and Thalmaier's approach, their argument works even on non-compact M (cf. [16]). Our Theorem 2 can be regarded as an extension of their result. Indeed, our approach is the same as theirs in spirit and advantages of probabilistic approach as mentioned are also inherited to our results as we have seen in Theorem 3. We can expect that our approach makes it possible to employ several techniques in stochastic analysis to obtain more detailed behavior of $\Theta_t(X_{\bar{\tau}_1 t}, Y_{\bar{\tau}_2 t})$, especially in the limit $\bar{\tau}_2 \to \bar{\tau}_1$, in a future development.

Now we compare our method of the proof with existing arguments in coupling methods from a technical point of view. We hope that the following observation would be helpful to extend other coupling arguments than ours in this case. A common and basic idea is to couple infinitesimal motions of two Brownian particles by using a parallel transport of tangent vectors along a minimal geodesic joining them. Thus the technical difficulty arises from the singularity of ($\mathcal{L}$-)distance, or the presence of ($\mathcal{L}$-)cut locus. In our approach, we consider coupled geodesic random walks each of which approximates the Brownian motion. After we establish a difference inequality for time evolution of the $\mathcal{L}$-distance between coupled random walks, we will obtain the result by taking a limit. Note that the convergence of our random walk in law to the Brownian motion in this time-inhomogeneous case is already established in [14].

In the time-homogeneous case, there are several arguments [8,10,32–34] to construct such a coupling by approximating it with ones which move as mentioned above if they are distant from the cut locus and move independently if they are close to the cut locus. In these cases, it will be important to estimate the size of the total time when particles are close to the cut locus. To do the same in the time-inhomogeneous case, it does not seem straightforward since the ($\mathcal{L}$-)cut locus depends on time, namely it moves as time evolves. In our approach, instead of applying stochastic calculus, we only need to show a difference inequality. Thus the singularity at the $\mathcal{L}$-cut locus causes less difficulties at this stage (see Remark 7 for more details).

If we employ the theory of optimal transportation, we will work on coupling of heat distributions instead of coupling of Brownian motions. Once we move to the world of heat distributions, we can ignore the cut locus since they are of measure zero with respect to the Riemannian measure. However, in the derivation of the monotonicity results, the theory of optimal transportation at present covers only the case that the cost function is squared distance or $\mathcal{L}$-distance. It reflects the difference of results between McCann and Topping [21] and Arnaudon, Coulibaly and Thalmaier [2]. It should be remarked that such a difference still exists between these two approaches, the one by optimal transportation and the other by stochastic analysis, even in the time-homogeneous case.

Arnaudon, Coulibaly and Thalmaier [2] study the problem by developing a new method. They constructed one-parameter family of coupled particles $((X_t(u))_{t\geqslant 0})_{u\in[0,1]}$ so that $X_t(u)$ moves as a Brownian motion for any u and $(X_t(u))_{u\in[0,1]}$ is a C^1-curve whose length is non-increasing in t. Thus $(X_t(0), X_t(1))$ is the expected coupling. To construct it, they first consider a finite number of particles $((X_t(u_i))_{t\geqslant 0})_i$ which are coupled with other particles by parallel transport. Then, by increasing the number of particles, we obtain such a one-parameter family in the limit. Since they are coupled by parallel transport, the distance between two particles is of bounded variation (at least before they hit the cut locus). Thus, if adjacent particles are sufficiently close to each other at time t, we can take a deterministic $\delta > 0$ such that they cannot hit the cut locus at least until time $t + \delta$. Based on this observation, they succeeded in avoiding the problem coming from the cut locus by increasing the number of particles to make it constitute a one-parameter family of particles. In the case of this paper, we work on the $\mathcal{L}$-distance instead of the Riemannian distance and construct a coupling by space–time parallel transport instead of a coupling by parallel transport. As a result, $\mathcal{L}$-distance between coupled particle is not of bounded variation (see Remark 6 for more details). Thus, our problem differs in nature from what is studied in [1]. If we want to extend Arnaudon, Coulibaly and Thalmaier's approach in the present case, we have to be careful and need some additional arguments. Even if we succeed in constructing a one-parameter family of particles $((X_t(u))_{t\geqslant 0})_{u\in[0,1]}$ coupled by space–time parallel transport, we cannot expect that $(X_t(u))_{u\in[0,1]}$ is a C^1-curve. In our approach, such a difference causes no additional difficulty. Indeed, as studied in [13,14], we already know that it works to construct coupled particles by reflection, the distance of which is naturally regarded as a semimartingale with a non-vanishing martingale part.

3. Coupling of Brownian motions in the absence of $\mathcal{L}$-cut locus

Since the proof of Theorem 2 involves some technical arguments, first we study the problem in the case that the $\mathcal{L}$-distance L has no singularity. More precisely, we do it here under the following assumption:

Assumption 1. The $\mathcal{L}$-cut locus is empty.

See Section 5.1 or [6,30,35] for the definition of $\mathcal{L}$-cut locus. Under Assumption 1, the following hold:

1. For all $x, y \in M$ and all $\bar{\tau}_1 \leqslant \tau_1 < \tau_2 \leqslant T$ there is a unique minimizer $\gamma_{xy}^{\tau_1\tau_2}$ of $L(x, \tau_1; y, \tau_2)$ (existence of $\gamma_{xy}^{\tau_1\tau_2}$ is proved in [6, Lemma 7.27], while uniqueness follows immediately from the characterization of $\mathcal{L}$-cut locus, see Section 5.1).
2. The function L is globally smooth.

Thus, in this case, we can freely use stochastic analysis on the frame bundle without taking any care on regularity of L.

3.1. Construction of the coupling

A $g(\tau)$-Brownian motion $\tilde{X}$ on M (scaled in time by the factor $\bar{\tau}_1$) starting at a point $x \in M$ at time $s \in [1, T/\bar{\tau}_2]$ can be constructed in the following way [1,7,16]: Let $\pi : \mathcal{F}(M) \to M$ be the frame bundle and $(e_i)_{i=1}^d$ the standard basis of $\mathbb{R}^d$. For each $\tau \in [\bar{\tau}_1, T]$ let $(H_i(\tau))_{i=1}^d$ be the associated $g(\tau)$-horizontal vector fields on $\mathcal{F}(M)$ (i.e. $H_i(\tau, u)$ is the $g(\tau)$-horizontal lift of ue_i). Moreover let $(\mathcal{V}^{\alpha,\beta})_{\alpha,\beta=1}^d$ be the canonical vertical vector fields, i.e. $(\mathcal{V}^{\alpha,\beta} f)(u) := \frac{\partial}{\partial m_{\alpha\beta}}|_{\mathbf{m}=\mathrm{Id}}(f(u(\mathbf{m})))$ ($\mathbf{m} = (m_{\alpha\beta})_{\alpha,\beta=1}^d \in GL_d(\mathbb{R})$), and let $(W_t)_{t\geqslant 0}$ be a standard $\mathbb{R}^d$-valued Brownian motion. By $\mathcal{O}^{g(\tau)}(M)$, we denote the $g(\tau)$-orthonormal frame bundle.

We first define a scaled horizontal Brownian motion as the solution $\tilde{U} = (\tilde{U}_t)_{t\in[s,T/\bar{\tau}_1]}$ of the Stratonovich SDE

$$d\tilde{U}_t = \sqrt{2\bar{\tau}_1}\sum_{i=1}^d H_i(\bar{\tau}_1 t, \tilde{U}_t) \circ dW_t^i - \bar{\tau}_1 \sum_{\alpha,\beta=1}^d \frac{\partial g}{\partial \tau}(\bar{\tau}_1 t)(\tilde{U}_t e_\alpha, \tilde{U}_t e_\beta)\mathcal{V}^{\alpha\beta}(\tilde{U}_t)\, dt \tag{3}$$

on $\mathcal{F}(M)$ with initial value $\tilde{U}_s = u \in \mathcal{O}_x^{g(\bar{\tau}_1 s)}(M)$, and then define a scaled Brownian motion $\tilde{X}$ on M as

$$\tilde{X}_t := \pi \tilde{U}_t.$$

Note that $\tilde{X}_t$ does not move when $\bar{\tau}_1 = 0$. The last term in (3) ensures that $\tilde{U}_t \in \mathcal{O}^{g(\bar{\tau}_1 t)}(M)$ for all $t \in [s, T/\bar{\tau}_1]$ (see [1, Proposition 1.1], [7, Proposition 1.2]), so that by Itô's formula for all smooth $f : [s, T/\bar{\tau}_1] \times M \to \mathbb{R}$

$$df(t, \tilde{X}_t) = \frac{\partial f}{\partial t}(t, \tilde{X}_t)\, dt + \sqrt{2\bar{\tau}_1}\sum_{i=1}^d (\tilde{U}_t e_i) f(t, \tilde{X}_t)\, dW_t^i + \bar{\tau}_1 \Delta_{g(\tau_1 t)} f(t, \tilde{X}_t)\, dt.$$

Let us define $(X_\tau)_{\tau\in[\bar{\tau}_1 s, T]}$ by $X_{\bar{\tau}_1 t} := \tilde{X}_t$. Then X_τ becomes a $g(\tau)$-Brownian motion when $\bar{\tau}_1 > 0$.

Remark 3. Intuitively, it might be helpful to think that X_τ lives in $(M, g(\tau))$, or $\tilde{X}_t$ lives in $(M, g(\bar{\tau}_1 t))$. The same is true for Y and $\tilde{Y}$ which will be defined below. Similarly, for all curves $\gamma : [\tau_1, \tau_2] \to M$ appearing in connection with $\mathcal{L}$-distance, we can naturally regard $\gamma(\tau)$ as in $(M, g(\tau))$.

We now want to construct a second scaled Brownian motion $\tilde{Y}$ on M in such a way that its infinitesimal increments $d\tilde{Y}_t$ are "space–time parallel" to those of $\tilde{X}$ (up to scaling effect) along the minimal $\mathcal{L}$-geodesic (namely, the minimizer of L) from $(\tilde{X}_t, \bar{\tau}_1 t)$ to $(\tilde{Y}_t, \bar{\tau}_2 t)$. To make this idea precise, we first define the notion of space–time parallel vector field:

Definition 1 *(Space–time parallel vector field).* Let $\bar{\tau}_1 \leqslant \tau_1 < \tau_2 \leqslant T$ and $\gamma : [\tau_1, \tau_2] \to M$ be a smooth curve. We say that a vector field Z along γ is *space–time parallel* if

$$\nabla^{g(\tau)}_{\dot{\gamma}(\tau)} Z(\tau) = -\operatorname{Ric}^{\#}_{g(\tau)}\bigl(Z(\tau)\bigr) \tag{4}$$

holds for all $\tau \in [\tau_1, \tau_2]$. Here $\nabla^{g(\tau)}$ stands for the covariant derivative associated with the $g(\tau)$-Levi-Civita connection and $\operatorname{Ric}^{\#}_{g(\tau)}$ is defined by regarding the $g(\tau)$-Ricci curvature as a $(1, 1)$-tensor via $g(\tau)$. Since (4) is a linear first-order ODE, for any $\xi \in T_{\gamma(\tau_1)}M$ there exists a unique space–time parallel vector field Z along γ with $Z(\tau_1) = \xi$.

Remark 4. Whenever Z and Z' are space–time parallel vector fields along a curve γ, their $g(\tau)$-inner product is constant in τ:

$$\begin{aligned}
\frac{d}{d\tau}\bigl\langle Z(\tau), Z'(\tau)\bigr\rangle_{g(\tau)} &= \frac{\partial g}{\partial \tau}(\tau)\bigl(Z(\tau), Z'(\tau)\bigr) + \bigl\langle \nabla^{g(\tau)}_{\dot{\gamma}(\tau)} Z(\tau), Z'(\tau)\bigr\rangle_{g(\tau)} + \bigl\langle Z(\tau), \nabla^{g(\tau)}_{\dot{\gamma}(\tau)} Z'(\tau)\bigr\rangle_{g(\tau)} \\
&= 2\operatorname{Ric}_{g(\tau)}\bigl(Z(\tau), Z'(\tau)\bigr) - \operatorname{Ric}_{g(\tau)}\bigl(Z(\tau), Z'(\tau)\bigr) - \operatorname{Ric}_{g(\tau)}\bigl(Z(\tau), Z'(\tau)\bigr) \\
&= 0.
\end{aligned}$$

Definition 2 *(Space–time parallel transport).* For $x, y \in M$ and $\bar{\tau}_1 \leqslant \tau_1 < \tau_2 \leqslant T$, we define a map $m^{\tau_1\tau_2}_{xy} : T_xM \to T_yM$ as follows: $m^{\tau_1\tau_2}_{xy}(\xi) := Z(\tau_2)$, where Z is the unique space–time parallel vector field along $\gamma^{\tau_1\tau_2}_{xy}$ with $Z(\tau_1) = \xi$. By Remark 4, $m^{\tau_1\tau_2}_{xy}$ is an isometry from $(T_xM, g(\tau_1))$ to $(T_yM, g(\tau_2))$. In addition, it smoothly depends on x, τ_1, y, τ_2 under Assumption 1.

Remark 5. The emergence of the Ricci curvature in (4) is based on the Ricci flow equation (1). Indeed, we can introduce the notion of space–time parallel transport even in the absence of (1) with keeping the property in Remark 4 by using $2^{-1}\partial_\tau g(\tau)^{\#}$ instead of $\operatorname{Ric}^{\#}_{g(\tau)}$ in (4). This would be a natural extension in the sense that it coincides with the usual parallel transport when $g(\tau)$ is constant in τ.

Similarly as in [10, Formula (6.5.1)], we now define a second scaled horizontal Brownian motion $\tilde{V} = (\tilde{V}_t)_{t\in[s,T/\bar{\tau}_2]}$ on $\mathcal{F}(M)$ as the solution of

$$\begin{aligned}
d\tilde{V}_t &= \sqrt{2\bar{\tau}_2}\sum_{i=1}^{d} H_i(\bar{\tau}_2 t, \tilde{V}_t) \circ dB^i_t - \bar{\tau}_2 \sum_{\alpha,\beta=1}^{d} \frac{\partial g}{\partial \tau}(\bar{\tau}_2 t)(\tilde{V}_t e_\alpha, \tilde{V}_t e_\beta)\mathcal{V}^{\alpha\beta}(\tilde{V}_t)\,dt, \\
dB_t &= \tilde{V}_t^{-1} m^{\tau_1,\tau_2}_{\pi\tilde{U}_t,\pi\tilde{V}_t}\tilde{U}_t\,dW_t
\end{aligned}$$

with initial value $\tilde{V}_s = v \in \mathcal{O}^{g(\bar{\tau}_2 s)}_y(M)$, and we set $\tilde{Y}_t := \pi\tilde{V}_t$. As we did for $\tilde{X}$, let us define $(Y_\tau)_{\tau\in[\bar{\tau}_2 s,T]}$ by $Y_{\bar{\tau}_2 t} := \tilde{Y}_t$ to make Y a $g(\tau)$-Brownian motion. From a theoretical point of view, it seems to be natural to work with (X_τ, Y_τ) (see Remark 3). However, for technical simplicity, we will prefer to work with $(\tilde{X}_t, \tilde{Y}_t)$ instead in the sequel.

3.2. Proof of Theorem 2 in the absence of $\mathcal{L}$-cut locus

Our argument in this section is based on the following Itô formula for $(\tilde{X}_t, \tilde{Y}_t)$:

Lemma 1. *Let f be a smooth function on $[s, T/\bar{\tau}_2] \times M \times M$. Then*

$$df(t, \tilde{X}_t, \tilde{Y}_t) = \frac{\partial f}{\partial t}(t, \tilde{X}_t, \tilde{Y}_t)\,dt + \sum_{i=1}^{d} \bigl[\sqrt{2\bar{\tau}_1}\tilde{U}_t e_i \oplus \sqrt{2\bar{\tau}_2}\tilde{V}_t e_i^*\bigr] f(t, \tilde{X}_t, \tilde{Y}_t)\,dW_t^i$$
$$+ \sum_{i=1}^{d} \operatorname{Hess}_{g(\bar{\tau}_1 t)\oplus g(\bar{\tau}_2 t)} f|_{(t,\tilde{X}_t,\tilde{Y}_t)}\bigl(\sqrt{\bar{\tau}_1}\tilde{U}_t e_i \oplus \sqrt{\bar{\tau}_2}\tilde{V}_t e_i^*, \sqrt{\bar{\tau}_1}\tilde{U}_t e_i \oplus \sqrt{\bar{\tau}_2}\tilde{V}_t e_i^*\bigr)\,dt.$$

Here the Hessian of f is taken with respect to the product metric $g(\bar{\tau}_1 t) \oplus g(\bar{\tau}_2 t)$, e_i^* stands for $e_i^*(\tilde{U}_t, \bar{\tau}_1 t; \tilde{V}_t, \bar{\tau}_2 t)$, where

$$e_i^*(u, \tau_1; v, \tau_2) := v^{-1} m_{\pi u, \pi v}^{\tau_1, \tau_2} u e_i,$$

and for tangent vectors $\xi_1 \in T_x M$, $\xi_2 \in T_y M$ we write $\xi_1 \oplus \xi_2 := (\xi_1, \xi_2) \in T_{(x,y)}(M \times M)$.

Proof. As in [11, Formula (2.11)], Itô's formula applied to a smooth function $\tilde{f}$ on $[s, T/\bar{\tau}_2] \times \mathcal{F}(M) \times \mathcal{F}(M)$ gives

$$\begin{aligned}
d\tilde{f}(t, \tilde{U}_t, \tilde{V}_t) = {} & \frac{\partial \tilde{f}}{\partial t}(t, \tilde{U}_t, \tilde{V}_t)\,dt \\
& + \sum_{i=1}^{d} \bigl[\sqrt{2\bar{\tau}_1}\bigl(H_{i,1}(\bar{\tau}_1 t, \cdot)\tilde{f}\bigr)(t, \tilde{U}_t, \tilde{V}_t)\,dW_t^i \\
& + \sqrt{2\bar{\tau}_2}\bigl(H_{i,2}(\bar{\tau}_2 t, \cdot)\tilde{f}\bigr)(t, \tilde{U}_t, \tilde{V}_t)\,dB_t^i\bigr] \\
& + \sum_{i=1}^{d} \bigl[\bar{\tau}_1\bigl(H_{i,1}^2(\bar{\tau}_1 t, \cdot)\tilde{f}\bigr)(t, \tilde{U}_t, \tilde{V}_t) + \bar{\tau}_2\bigl(H_{i,2}^2(\bar{\tau}_2 t, \cdot)\tilde{f}\bigr)(t, \tilde{U}_t, \tilde{V}_t)\bigr]\,dt \\
& + 2\sqrt{\bar{\tau}_1\bar{\tau}_2} \sum_{i,j=1}^{d} \bigl(H_{i,1}(\bar{\tau}_1 t, \cdot) H_{j,2}(\bar{\tau}_2 t, \cdot)\tilde{f}\bigr)(t, \tilde{U}_t, \tilde{V}_t)\,d\langle W^i, B^j\rangle_t \\
& - \sum_{\alpha,\beta=1}^{d} \Bigl[\bar{\tau}_1 \frac{\partial g}{\partial \tau}(\bar{\tau}_1 t)(\tilde{U}_t e_\alpha, \tilde{U}_t e_\beta)\mathcal{V}^{\alpha\beta}(\tilde{U}_t) \\
& \oplus \bar{\tau}_2 \frac{\partial g}{\partial \tau}(\bar{\tau}_2 t)(\tilde{V}_t e_\alpha, \tilde{V}_t e_\beta)\mathcal{V}^{\alpha\beta}(\tilde{V}_t)\Bigr] \tilde{f}(t, \tilde{U}_t, \tilde{V}_t)\,dt,
\end{aligned}$$

where $H_{i,1}$ respectively $H_{i,2}$ means H_i applied with respect to the first respectively second space variable. By the definition of B, this equals

$$
\begin{aligned}
&\frac{\partial \tilde f}{\partial t}(t,\tilde U_t,\tilde V_t)\,dt\\
&\quad+\sum_{i=1}^d\big[\sqrt{2\bar\tau_1}H_i(\bar\tau_1 t,\tilde U_t)\oplus\sqrt{2\bar\tau_2}H_i^*(\tilde U_t,\bar\tau_1 t;\tilde V_t,\bar\tau_2 t)\big]\tilde f(t,\tilde U_t,\tilde V_t)\,dW_t^i\\
&\quad+\sum_{i=1}^d\big[\sqrt{\bar\tau_1}H_i(\bar\tau_1 t,\tilde U_t)\oplus\sqrt{\bar\tau_2}H_i^*(\tilde U_t,\bar\tau_1 t;\tilde V_t,\bar\tau_2 t)\big]^2\tilde f(t,\tilde U_t,\tilde V_t)\,dt\\
&\quad-\sum_{\alpha,\beta=1}^d\bigg[\bar\tau_1\frac{\partial g}{\partial\tau}(\bar\tau_1 t)(\tilde U_t e_\alpha,\tilde U_t e_\beta)\mathcal V^{\alpha\beta}(\tilde U_t)\oplus\bar\tau_2\frac{\partial g}{\partial\tau}(\bar\tau_2 t)(\tilde V_t e_\alpha,\tilde V_t e_\beta)\mathcal V^{\alpha\beta}(\tilde V_t)\bigg]\tilde f(t,\tilde U_t,\tilde V_t)\,dt,
\end{aligned}
$$

where $H_i^*(u,\tau_1;v,\tau_2)$ is the $g(\tau_2)$-horizontal lift of $ve_i^*(u,\tau_1;v,\tau_2)$.

The claim follows by choosing $\tilde f(t,u,v):=f(t,\pi u,\pi v)$ because this $\tilde f$ is constant in the vertical direction so that the term involving $\mathcal V^{\alpha\beta}\tilde f$ vanishes. □

Let $\Lambda(t,x,y):=L(x,\bar\tau_1 t;y,\bar\tau_2 t)$. In order to apply Lemma 1 to the function Θ we need the following proposition, whose proof is given in the next section. Since we will use it again in Section 5, we state it without assuming Assumption 1.

Proposition 1. *Take $x,y\in M$, $u\in\mathcal O_x^{g(\bar\tau_1 t)}(M)$ and $v\in\mathcal O_y^{g(\bar\tau_2 t)}(M)$. Let γ be a minimizer of $L(x,\bar\tau_1 t;y,\bar\tau_2 t)$. Assume that $(x,\bar\tau_1 t;y,\bar\tau_2 t)$ is not in the $\mathcal L$-cut locus. Set $\xi_i:=\sqrt{\bar\tau_1}ue_i\oplus\sqrt{\bar\tau_2}ve_i^*(u,\bar\tau_1 t;v,\bar\tau_2 t)$. Then*

$$
\begin{aligned}
\frac{\partial\Lambda}{\partial t}(t,x,y)=\frac1t\int_{\bar\tau_1 t}^{\bar\tau_2 t}\tau^{3/2}\bigg(&\frac{3}{2\tau}R_{g(\tau)}\big(\gamma(\tau)\big)-\Delta_{g(\tau)}R_{g(\tau)}\big(\gamma(\tau)\big)-2|\mathrm{Ric}_{g(\tau)}|^2_{g(\tau)}\big(\gamma(\tau)\big)\\
&-\frac{1}{2\tau}\big|\dot\gamma(\tau)\big|^2_{g(\tau)}+2\,\mathrm{Ric}_{g(\tau)}\big(\dot\gamma(\tau),\dot\gamma(\tau)\big)\bigg)\,d\tau,
\end{aligned}\tag{5}
$$

$$
\begin{aligned}
&\sum_{i=1}^d\mathrm{Hess}_{g(\tau_1)\oplus g(\tau_2)}\,\Lambda|_{(t,x,y)}(\xi_i,\xi_i)\\
&\quad\leqslant\frac{d\sqrt\tau}{t}\bigg|_{\tau=\bar\tau_1 t}^{\tau=\bar\tau_2 t}+\frac1t\int_{\bar\tau_1 t}^{\bar\tau_2 t}\tau^{3/2}\big(2|\mathrm{Ric}_{g(\tau)}|^2_{g(\tau)}\big(\gamma(\tau)\big)+\Delta_{g(\tau)}R_{g(\tau)}\big(\gamma(\tau)\big)\\
&\qquad-\frac2\tau R_{g(\tau)}\big(\gamma(\tau)\big)-2\,\mathrm{Ric}_{g(\tau)}\big(\dot\gamma(\tau),\dot\gamma(\tau)\big)\big)\,d\tau
\end{aligned}\tag{6}
$$

and consequently

$$
\frac{\partial\Lambda}{\partial t}(t,x,y)+\sum_{i=1}^d\mathrm{Hess}_{g(\tau_1)\oplus g(\tau_2)}\,\Lambda|_{(t,x,y)}(\xi_i,\xi_i)
$$

$$\leqslant \frac{d}{\sqrt{t}}(\sqrt{\bar\tau_2}-\sqrt{\bar\tau_1}) - \frac{1}{2t}\int_{\bar\tau_1 t}^{\bar\tau_2 t}\sqrt{\tau}\big(R_{g(\tau)}\big(\gamma(\tau)\big)+\big|\dot\gamma(\tau)\big|^2_{g(\tau)}\big)\,d\tau$$
$$= \frac{d}{\sqrt{t}}(\sqrt{\bar\tau_2}-\sqrt{\bar\tau_1}) - \frac{1}{2t}\Lambda(t,x,y).$$

The proof of Theorem 2 is now achieved under Assumption 1 by combining Lemma 1 and Proposition 1:

Proof of Theorem 2 under Assumption 1. Since Θ is bounded from below, it suffices to show that the bounded variation part of $\Theta_t(\tilde X_t,\tilde Y_t)$ is non-positive. By Lemma 1,

$$d\Theta_t(\tilde X_t,\tilde Y_t) = \Bigg[\partial_t\Theta_t(\tilde X_t,\tilde Y_t) + \sum_{i=1}^d \operatorname{Hess}_{g(\bar\tau_1 t)\oplus g(\bar\tau_2 t)}\Theta_t|_{(\tilde X_t,\tilde Y_t)}\big(\sqrt{\bar\tau_1}\tilde U_t e_i\oplus\sqrt{\bar\tau_2}\tilde V_t e_i^*,\sqrt{\bar\tau_1}\tilde U_t e_i\oplus\sqrt{\bar\tau_2}\tilde V_t e_i^*\big)\Bigg]dt + \sum_{i=1}^d\big[\sqrt{2\bar\tau_1}\tilde U_t e_i\oplus\sqrt{2\bar\tau_2}\tilde V_t e_i^*\big]\Theta_t(\tilde X_t,\tilde Y_t)\,dW_t^i.$$

For the bounded variation part we obtain

$$\partial_t\Theta_t(\tilde X_t,\tilde Y_t) = \frac{\sqrt{\bar\tau_2}-\sqrt{\bar\tau_1}}{\sqrt{t}}\Lambda(t,\tilde X_t,\tilde Y_t) + 2(\sqrt{\bar\tau_2 t}-\sqrt{\bar\tau_1 t})\frac{\partial\Lambda}{\partial t}(t,\tilde X_t,\tilde Y_t) - 2d(\sqrt{\bar\tau_2}-\sqrt{\bar\tau_1})^2$$

and

$$\sum_{i=1}^d \operatorname{Hess}_{g(\bar\tau_1 t)\oplus g(\bar\tau_2 t)}\Theta_t|_{(\tilde X_t,\tilde Y_t)}\big(\sqrt{\bar\tau_1}\tilde U_t e_i\oplus\sqrt{\bar\tau_2}\tilde V_t e_i^*,\sqrt{\bar\tau_1}\tilde U_t e_i\oplus\sqrt{\bar\tau_2}\tilde V_t e_i^*\big)$$
$$= 2(\sqrt{\bar\tau_2 t}-\sqrt{\bar\tau_1 t})\sum_{i=1}^d \operatorname{Hess}_{g(\bar\tau_1 t)\oplus g(\bar\tau_2 t)}\Lambda|_{(t,\tilde X_t,\tilde Y_t)}\big(\sqrt{\bar\tau_1}\tilde U_t e_i\oplus\sqrt{\bar\tau_2}\tilde V_t e_i^*,\sqrt{\bar\tau_1}\tilde U_t e_i\oplus\sqrt{\bar\tau_2}\tilde V_t e_i^*\big).$$

Thus, by Proposition 1,

$$\partial_t\Theta_t(\tilde X_t,\tilde Y_t) + \sum_{i=1}^d \operatorname{Hess}_{g(\bar\tau_1 t)\oplus g(\bar\tau_2 t)}\Theta_t|_{(\tilde X_t,\tilde Y_t)}\big(\sqrt{\bar\tau_1}\tilde U_t e_i\oplus\sqrt{\bar\tau_2}\tilde V_t e_i^*,\sqrt{\bar\tau_1}\tilde U_t e_i\oplus\sqrt{\bar\tau_2}\tilde V_t e_i^*\big)$$
$$\leqslant 2(\sqrt{\bar\tau_2 t}-\sqrt{\bar\tau_1 t})\left[\frac{d}{\sqrt{t}}(\sqrt{\bar\tau_2}-\sqrt{\bar\tau_1}) - \frac{1}{2t}\Lambda(t,\tilde X_t,\tilde Y_t)\right] + \frac{\sqrt{\bar\tau_2}-\sqrt{\bar\tau_1}}{\sqrt{t}}\Lambda(t,\tilde X_t,\tilde Y_t) - 2d(\sqrt{\bar\tau_2}-\sqrt{\bar\tau_1})^2 = 0.$$

Hence $\Theta_t(\tilde X_t,\tilde Y_t)$ is indeed a supermartingale. □

Remark 6. Unlike the case in [2], the pathwise contraction of $\Theta_t(\tilde{X}_t, \tilde{Y}_t)$ is no longer true in our case. In other words, the martingale part of $\Theta_t(\tilde{X}_t, \tilde{Y}_t)$ does not vanish. We will see it in the following. The minimal $\mathcal{L}$-geodesic $\gamma = \gamma_{xy}^{\tau_1\tau_2}$ of $L(x, \tau_1; y, \tau_2)$ satisfies the $\mathcal{L}$-geodesic equation

$$\nabla^{g(\tau)}_{\dot{\gamma}(\tau)} \dot{\gamma}(\tau) = \frac{1}{2} \nabla^{g(\tau)} R_{g(\tau)} - 2 \operatorname{Ric}^{\#}_{g(\tau)}\big(\dot{\gamma}(\tau)\big) - \frac{1}{2\tau} \dot{\gamma}(\tau) \tag{7}$$

(see [6, Corollary 7.19]). Thus the first variation formula (see [6, Lemma 7.15]) yields

$$\sqrt{2\bar{\tau}_1} \tilde{U}_t e_i \oplus \sqrt{2\bar{\tau}_2} \tilde{V}_t e_i^* \Lambda(t, \tilde{X}_t, \tilde{Y}_t) = \sqrt{2t}\,\bar{\tau}_2 \big\langle \tilde{V}_t e_i^*, \dot{\gamma}(\bar{\tau}_2 t)\big\rangle_{g(\bar{\tau}_2 t)} - \sqrt{2t}\,\bar{\tau}_1 \big\langle \tilde{U}_t e_i, \dot{\gamma}(\bar{\tau}_1 t)\big\rangle_{g(\bar{\tau}_1 t)}. \tag{8}$$

One obstruction to pathwise contraction is in the difference of time-scalings $\bar{\tau}_1$ and $\bar{\tau}_2$. In addition, by (7), $\sqrt{\tau}\dot{\gamma}(\tau)$ is *not* space–time parallel to γ in general (cf. Remark 4).

4. Proof of Proposition 1

In this section, we write $\tau_1 := \bar{\tau}_1 t$ and $\tau_2 := \bar{\tau}_2 t$. We assume $\tau_2 < T$. For simplicity of notations, we abbreviate the dependency on the metric $g(\tau)$ of several geometric quantities such as Ric, R, the inner product $\langle\cdot,\cdot\rangle$, the covariant derivative ∇, etc. when our choice of τ is obvious. For this abbreviation, we will think that $\gamma(\tau)$ is in $(M, g(\tau))$ and $\dot{\gamma}(\tau)$ is in $(T_{\gamma(\tau)}M, g(\tau))$. Note that, when $\bar{\tau}_1 = 0$, $\lim_{\tau \downarrow \bar{\tau}_1} \sqrt{\tau}\dot{\gamma}(\tau)$ exists while $\lim_{\tau\downarrow 0} |\dot{\gamma}(\tau)| = \infty$. In any case, $\sqrt{\tau}|\dot{\gamma}(\tau)|$ is locally bounded (see Lemma 3).

We first compute the time derivative of Λ. When $\bar{\tau}_1 > 0$, by [30, Formulas (A.4) and (A.5)] we have

$$\frac{\partial L}{\partial \tau_1}(x, \tau_1; y, \tau_2) = -\sqrt{\tau_1}\big(R_{g(\tau_1)}(x) - \big|\dot{\gamma}(\tau_1)\big|^2\big),$$
$$\frac{\partial L}{\partial \tau_2}(x, \tau_1; y, \tau_2) = \sqrt{\tau_2}\big(R_{g(\tau_2)}(y) - \big|\dot{\gamma}(\tau_2)\big|^2\big),$$

so that

$$\begin{aligned}\frac{\partial \Lambda}{\partial t}(t, x, y) &= \bar{\tau}_1 \frac{\partial L}{\partial \tau_1}(x, \tau_1; y, \tau_2) + \bar{\tau}_2 \frac{\partial L}{\partial \tau_2}(x, \tau_1; y, \tau_2) \\ &= \frac{1}{t}\big(\tau_2^{3/2}\big(R\big(\gamma(\tau_2)\big) - \big|\dot{\gamma}(\tau_2)\big|^2\big) - \tau_1^{3/2}\big(R\big(\gamma(\tau_1)\big) - \big|\dot{\gamma}(\tau_1)\big|^2\big)\big).\end{aligned} \tag{9}$$

Thus the integration-by-parts formula yields,

$$\begin{aligned}\frac{\partial \Lambda}{\partial t}(t, x, y) = {}& \frac{3}{2t} \int_{\tau_1}^{\tau_2} \sqrt{\tau}\big(R\big(\gamma(\tau)\big) - \big|\dot{\gamma}(\tau)\big|^2\big)\, d\tau \\ & + \frac{1}{t} \int_{\tau_1}^{\tau_2} \tau^{3/2} \bigg(\frac{\partial R}{\partial \tau}\big(\gamma(\tau)\big) + \nabla_{\dot{\gamma}(\tau)} R\big(\gamma(\tau)\big) \\ & - 2\big\langle \nabla_{\dot{\gamma}(\tau)} \dot{\gamma}(\tau), \dot{\gamma}(\tau)\big\rangle - 2 \operatorname{Ric}\big(\dot{\gamma}(\tau), \dot{\gamma}(\tau)\big)\bigg)\, d\tau.\end{aligned} \tag{10}$$

Note that we have

$$\frac{\partial R}{\partial \tau} = -\Delta R - 2|\mathrm{Ric}|^2 \tag{11}$$

(see e.g. [29, Proposition 2.5.4]). Since γ satisfies the $\mathcal{L}$-geodesic equation (7), by substituting (7) and (11) into (10), we obtain (5). Note that the derivation of (9) and (10) is still valid even when $\bar{\tau}_1 = 0$ because of the remark at the beginning of this section. Thus (5) holds when $\bar{\tau}_1 = 0$, too.

In order to estimate $\sum_{i=1}^{d} \mathrm{Hess}_{g(\tau_1)\oplus g(\tau_2)} \Lambda|_{(t,x,y)}(\xi_i, \xi_i)$ we begin with the second variation formula for the $\mathcal{L}$-functional:

Lemma 2 *(Second variation formula). (See [6, Lemma 7.37].) Let $\Gamma : (-\varepsilon, \varepsilon) \times [\tau_1, \tau_2] \to M$ be a variation of γ, $S(s,\tau) := \partial_s \Gamma(s,\tau)$, and $Z(\tau) := \partial_s \Gamma(0,\tau)$ the variation field of Γ. Then*

$$\begin{aligned}\frac{d^2}{ds^2}\bigg|_{s=0} \mathcal{L}(\Gamma_s) &= 2\sqrt{\tau}\big\langle \dot{\gamma}(\tau), \nabla_{Z(\tau)} S(0,\tau)\big\rangle\Big|_{\tau=\tau_1}^{\tau=\tau_2} - 2\sqrt{\tau}\,\mathrm{Ric}\big(Z(\tau), Z(\tau)\big)\Big|_{\tau=\tau_1}^{\tau=\tau_2} \\ &\quad + \frac{1}{\sqrt{\tau}}\big|Z(\tau)\big|^2\Big|_{\tau=\tau_1}^{\tau=\tau_2} - \int_{\tau_1}^{\tau_2} \sqrt{\tau} H\big(\dot{\gamma}(\tau), Z(\tau)\big)\, d\tau \\ &\quad + \int_{\tau_1}^{\tau_2} 2\sqrt{\tau}\left|\nabla_{\dot{\gamma}(\tau)} Z(\tau) + \mathrm{Ric}^{\#}\big(Z(\tau)\big) - \frac{1}{2\tau} Z(\tau)\right|^2 d\tau, \end{aligned} \tag{12}$$

where

$$\begin{aligned} H\big(\dot{\gamma}(\tau), Z(\tau)\big) := &-2\frac{\partial\,\mathrm{Ric}}{\partial \tau}\big(Z(\tau), Z(\tau)\big) - \mathrm{Hess}\, R\big(Z(\tau), Z(\tau)\big) + 2\big|\mathrm{Ric}^{\#}\big(Z(\tau)\big)\big|^2 \\ &- \frac{1}{\tau}\,\mathrm{Ric}\big(Z(\tau), Z(\tau)\big) - 2\,\mathrm{Rm}\big(Z(\tau), \dot{\gamma}(\tau), \dot{\gamma}(\tau), Z(\tau)\big) \\ &- 4(\nabla_{\dot{\gamma}(\tau)}\mathrm{Ric})\big(Z(\tau), Z(\tau)\big) + 4(\nabla_{Z(\tau)}\mathrm{Ric})\big(\dot{\gamma}(\tau), Z(\tau)\big). \end{aligned} \tag{13}$$

In [6] this lemma is only proved in the case $\tau_1 = 0$ and $Z(\tau_1) = 0$. However, the proof given there can be easily adapted to the slightly more general case needed here.

Corollary 1. *(See [6, Lemma 7.39] for a similar statement.) If the variation field Z is of the form*

$$Z(\tau) = \sqrt{\frac{\tau}{t}} Z^*(\tau) \tag{14}$$

with a space–time parallel field Z^* satisfying $|Z^*(\tau)| \equiv 1$, then

$$\frac{d^2}{ds^2}\bigg|_{s=0}\mathcal{L}(\Gamma_s)=2\sqrt{\tau}\langle\dot\gamma(\tau),\nabla_{Z(\tau)}S(0,\tau)\rangle_{g(\tau)}\big|_{\tau=\tau_1}^{\tau=\tau_2}-2\sqrt{\tau}\operatorname{Ric}\big(Z(\tau),Z(\tau)\big)\big|_{\tau=\tau_1}^{\tau=\tau_2}$$
$$-\int_{\tau_1}^{\tau_2}\sqrt{\tau}H\big(\dot\gamma(\tau),Z(\tau)\big)\,d\tau+\frac{\sqrt{\tau}}{t}\bigg|_{\tau=\tau_1}^{\tau=\tau_2}.$$

Proof. Since Z^* is space–time parallel, Z satisfies

$$\nabla_{\dot\gamma(\tau)}Z(\tau)=-\operatorname{Ric}^{\#}\big(Z(\tau)\big)+\frac{1}{2\tau}Z(\tau),\tag{15}$$

so that the last term in (12) vanishes. □

Corollary 2 *(Hessian of L). (See [6, Corollary 7.40] for a similar statement.) Let Z be a vector field along γ of the form* (14) *and $\xi:=Z(\tau_1)\oplus Z(\tau_2)\in T_{(x,y)}(M\times M)$. Then*

$$\operatorname{Hess}_{g(\tau_1)\oplus g(\tau_2)}L|_{(x,\tau_1;y,\tau_2)}(\xi,\xi)\leqslant-\int_{\tau_1}^{\tau_2}\sqrt{\tau}H\big(\dot\gamma(\tau),Z(\tau)\big)\,d\tau+\frac{\sqrt{\tau}}{t}\bigg|_{\tau=\tau_1}^{\tau=\tau_2}$$
$$-2\sqrt{\tau}\operatorname{Ric}_{g(\tau)}\big(Z(\tau),Z(\tau)\big)\big|_{\tau=\tau_1}^{\tau=\tau_2}.\tag{16}$$

Proof. Let $\Gamma:(-\varepsilon,\varepsilon)\times[\tau_1,\tau_2]\to M$ be any variation of γ with variation field Z and such that

$$\nabla_{Z(\tau_1)}S(0,\tau_1)\text{ and }\nabla_{Z(\tau_2)}S(0,\tau_2)\text{ vanish.}\tag{17}$$

Let $l(s):=L(\Gamma(s,\tau_1),\tau_1;\Gamma(s,\tau_2),\tau_2)$ and $\hat l(s):=\mathcal{L}(\Gamma(s,\cdot))$. Since $\hat l(0)=l(0)$ and $\hat l(s)\geqslant l(s)$ for all $s\in(-\varepsilon,\varepsilon)$, we have $l''(0)\leqslant\hat l''(0)$ so that, using (17),

$$\operatorname{Hess}_{g(\tau_1)\oplus g(\tau_2)}L|_{(x,\tau_1;y,\tau_2)}(\xi,\xi)=\frac{d^2}{ds^2}\bigg|_{s=0}L\big(\Gamma(s,\tau_1),\tau_1;\Gamma(s,\tau_2),\tau_2\big)$$
$$=l''(0)\leqslant\hat l''(0)=\frac{d^2}{ds^2}\bigg|_{s=0}\mathcal{L}(\Gamma_s).$$

The claim now follows from Corollary 1. □

Let now Z_i^* ($i=1,\dots,d$) be space–time parallel fields along γ satisfying $Z_i^*(\tau_1)=ue_i$ (and consequently $Z_i^*(\tau_2)=ve_i^*$), and $Z_i(\tau):=\sqrt{\tau/t}Z_i^*(\tau)$ (so that $\xi_i=Z_i(\tau_1)\oplus Z_i(\tau_2)$). In order to estimate $\sum_{i=1}^d\operatorname{Hess}_{g(\tau_1)\oplus g(\tau_2)}L|_{(x,\tau_1;y,\tau_2)}(\xi_i,\xi_i)$ using Corollary 2 we will compute $\sum_{i=1}^d H(\dot\gamma(\tau),Z_i(\tau))$ in the following (see [6, Section 7.5.3] for a similar argument). Set I_1, I_2 and I_3 by

$$I_1 := -2\sum_{i=1}^{d} \frac{\partial\,\mathrm{Ric}}{\partial\tau}\big(Z_i(\tau), Z_i(\tau)\big),$$

$$I_2 := \sum_{i=1}^{d}\Big[-\mathrm{Hess}\,R\big(Z_i(\tau), Z_i(\tau)\big) + 2\big|\mathrm{Ric}^{\#}\big(Z_i(\tau)\big)\big|^2 - \frac{1}{\tau}\mathrm{Ric}\big(Z_i(\tau), Z_i(\tau)\big) - 2\,\mathrm{Rm}\big(Z_i(\tau), \dot\gamma(\tau), \dot\gamma(\tau), Z_i(\tau)\big)\Big],$$

$$I_3 := 4\sum_{i=1}^{d}\big[(\nabla_{Z_i(\tau)}\mathrm{Ric})\big(Z_i(\tau), \dot\gamma(\tau)\big) - (\nabla_{\dot\gamma(\tau)}\mathrm{Ric})\big(Z_i(\tau), Z_i(\tau)\big)\big].$$

Then $\sum_{i=1}^{d} H(\dot\gamma(\tau), Z_i(\tau)) = I_1 + I_2 + I_3$ holds. By a direct computation,

$$I_2 = \frac{\tau}{t}\bigg(-\Delta R\big(\gamma(\tau)\big) + 2|\mathrm{Ric}|^2\big(\gamma(\tau)\big) - \frac{1}{\tau}R\big(\gamma(\tau)\big) + 2\,\mathrm{Ric}\big(\dot\gamma(\tau), \dot\gamma(\tau)\big)\bigg). \tag{18}$$

The contracted Bianchi identity $\mathrm{div}\,\mathrm{Ric} = \frac{1}{2}\nabla R$ [18, Lemma 7.7] yields

$$I_3 = \frac{4\tau}{t}\big((\mathrm{div}\,\mathrm{Ric})\big(\dot\gamma(\tau)\big) - (\nabla_{\dot\gamma(\tau)}R)\big(\gamma(\tau)\big)\big) = -\frac{2\tau}{t}(\nabla_{\dot\gamma(\tau)}R)\big(\gamma(\tau)\big). \tag{19}$$

For I_1, we have

$$\begin{aligned} I_1 &= -2\sum_{i=1}^{d}\bigg[\frac{d}{d\tau}\big(\mathrm{Ric}\big(Z_i(\tau), Z_i(\tau)\big)\big) - (\nabla_{\dot\gamma(\tau)}\mathrm{Ric})\big(Z_i(\tau), Z_i(\tau)\big) - 2\,\mathrm{Ric}\big(\nabla_{\dot\gamma(\tau)}Z_i(\tau), Z_i(\tau)\big)\bigg] \\ &= -2\frac{d}{d\tau}\bigg(\frac{\tau}{t}R\big(\gamma(\tau)\big)\bigg) + 2\frac{\tau}{t}\nabla_{\dot\gamma(\tau)}R\big(\gamma(\tau)\big) + 4\sum_{i=1}^{d}\mathrm{Ric}\big(\nabla_{\dot\gamma(\tau)}Z_i(\tau), Z_i(\tau)\big) \\ &= -\frac{2\tau}{t}\bigg(\frac{1}{\tau}R\big(\gamma(\tau)\big) + \frac{\partial R}{\partial\tau}\big(\gamma(\tau)\big)\bigg) + 4\sum_{i=1}^{d}\mathrm{Ric}\big(\nabla_{\dot\gamma(\tau)}Z_i(\tau), Z_i(\tau)\big). \end{aligned} \tag{20}$$

Since Z_i satisfies (15),

$$\begin{aligned} 4\sum_{i=1}^{d}\mathrm{Ric}\big(\nabla_{\dot\gamma(\tau)}Z_i(\tau), Z_i(\tau)\big) &= 4\sum_{i=1}^{d}\mathrm{Ric}\bigg(-\mathrm{Ric}^{\#}\big(Z_i(\tau)\big) + \frac{1}{2\tau}Z_i(\tau), Z_i(\tau)\bigg) \\ &= -\frac{2\tau}{t}\bigg(2|\mathrm{Ric}|^2\big(\gamma(\tau)\big) - \frac{1}{\tau}R\big(\gamma(\tau)\big)\bigg). \end{aligned} \tag{21}$$

By substituting (21) into (20),

$$I_1 = -\frac{2\tau}{t}\bigg(\frac{\partial R}{\partial\tau}\big(\gamma(\tau)\big) + 2|\mathrm{Ric}|^2\big(\gamma(\tau)\big)\bigg). \tag{22}$$

Hence, by combining (22), (19) and (18),

$$\sum_{i=1}^{d} H\big(\dot\gamma(\tau), Z_i(\tau)\big) = \frac{\tau}{t}\bigg(-2\frac{\partial R}{\partial \tau}\big(\gamma(\tau)\big) - 2|\mathrm{Ric}|^2\big(\gamma(\tau)\big) - \Delta R\big(\gamma(\tau)\big) - \frac{1}{\tau}R\big(\gamma(\tau)\big) + 2\,\mathrm{Ric}\big(\dot\gamma(\tau), \dot\gamma(\tau)\big) - 2(\nabla_{\dot\gamma(\tau)}R)\big(\gamma(\tau)\big)\bigg).$$

Inserting this into (16) we obtain

$$\begin{aligned}
&\sum_{i=1}^{d} \mathrm{Hess}_{g(\tau_1)\oplus g(\tau_2)} L|_{(x,\tau_1;y,\tau_2)}(\xi_i,\xi_i) \\
&\quad\leqslant \frac{1}{t}\int_{\tau_1}^{\tau_2} \tau^{3/2}\bigg(2\frac{\partial R}{\partial \tau}\big(\gamma(\tau)\big) + 2|\mathrm{Ric}|^2\big(\gamma(\tau)\big) + \Delta R\big(\gamma(\tau)\big) \\
&\qquad + \frac{1}{\tau}R\big(\gamma(\tau)\big) - 2\,\mathrm{Ric}\big(\dot\gamma(\tau), \dot\gamma(\tau)\big) + 2(\nabla_{\dot\gamma(\tau)}R)\big(\gamma(\tau)\big)\bigg)\,d\tau \\
&\qquad + \frac{d\sqrt{\tau}}{t}\bigg|_{\tau=\tau_1}^{\tau=\tau_2} - \frac{2\tau^{3/2}}{t}R\big(\gamma(\tau)\big)\Big|_{\tau=\tau_1}^{\tau=\tau_2} \\
&\quad= \frac{d\sqrt{\tau}}{t}\bigg|_{\tau=\tau_1}^{\tau=\tau_2} + \frac{1}{t}\int_{\tau_1}^{\tau_2} \tau^{3/2}\bigg(2|\mathrm{Ric}|^2\big(\gamma(\tau)\big) + \Delta R\big(\gamma(\tau)\big) \\
&\qquad - \frac{2}{\tau}R\big(\gamma(\tau)\big) - 2\,\mathrm{Ric}\big(\dot\gamma(\tau), \dot\gamma(\tau)\big)\bigg)\,d\tau
\end{aligned}$$

which completes the proof of Proposition 1.

5. Coupling via approximation by geodesic random walks

To avoid a technical difficulty coming from singularity of L on the $\mathcal{L}$-cut locus, we provide an alternative way to constructing a coupling of Brownian motions by space–time parallel transport. In this section, we first define a coupling of geodesic random walks which approximate $g(\tau)$-Brownian motion. Next, we introduce some estimates on geometric quantities in Section 5.1. Those are obtained as a small modification of existing arguments in [6,30,35]. The $\mathcal{L}$-cut locus is also reviewed and studied there. We use those estimates in Section 5.2 to study the behavior of the $\mathcal{L}$-distance between coupled random walks. The argument there includes a discrete analogue of the Itô formula as well as a local uniform control of error terms. Finally, we will complete the proof of Theorems 2 and 3 in Section 5.3.

Let us take a family of minimal $\mathcal{L}$-geodesics $\{\gamma_{xy}^{\tau_1\tau_2} \mid \bar\tau_1 \leqslant \tau_1 < \tau_2 \leqslant \bar\tau_2,\ x, y \in M\}$ so that a map $(x, \tau_1; y, \tau_2) \mapsto \gamma_{xy}^{\tau_1\tau_2}$ is measurable. The existence of such a family of minimal $\mathcal{L}$-geodesics can be shown in a similar way as discussed in the proof of [20, Proposition 2.6] since the family of minimal $\mathcal{L}$-geodesics with fixed endpoints is compact (cf. [6, the proof of Lemma 7.27]). For each $\tau \in [\bar\tau_1, T]$, take a measurable section $\Phi^{(\tau)}$ of $g(\tau)$-orthonormal frame bundle $\mathcal{O}^{g(\tau)}(M)$

of M. For $x, y \in M$ and $\tau_1, \tau_2 \in [\bar{\tau}_1, T]$ with $\tau_1 < \tau_2$, let us define $\Phi_i(x, \tau_1; y, \tau_2) \in \mathcal{F}(M)$ for $i = 1, 2$ by

$$\Phi_1(x, \tau_1; y, \tau_2) := \Phi^{(\tau_1)}(x),$$
$$\Phi_2(x, \tau_1; y, \tau_2) := m_{xy}^{\tau_1\tau_2} \circ \Phi^{(\tau_1)}(x),$$

where $m_{xy}^{\tau_1\tau_2}$ is as given in Definition 2. Let us take a family of $\mathbb{R}^d$-valued i.i.d. random variables $(\lambda_n)_{n\in\mathbb{N}}$ which are uniformly distributed on a unit ball centered at origin. We denote the (Riemannian) exponential map with respect to $g(\tau)$ at $x \in M$ by $\exp_x^{(\tau)}$. In what follows, we define a coupled geodesic random walk $\mathbf{X}_t^\varepsilon = (X_{\bar{\tau}_1 t}^\varepsilon, Y_{\bar{\tau}_2 t}^\varepsilon)$ with scale parameter $\varepsilon > 0$ and initial condition $\mathbf{X}_s^\varepsilon = (x_1, y_1)$ inductively. First we set $(X_{\bar{\tau}_1 s}^\varepsilon, Y_{\bar{\tau}_2 s}^\varepsilon) := (x_1, y_1)$. For simplicity of notations, we set $t_n := (s + \varepsilon^2 n) \wedge (T/\bar{\tau}_2)$. After we defined $(\mathbf{X}_t^\varepsilon)_{t\in[s,t_n]}$, we extend it to $(\mathbf{X}_t^\varepsilon)_{t\in[s,t_{n+1}]}$ by

$$\hat{\lambda}_{n+1}^{(i)} := \sqrt{d+2}\,\Phi_i\big(X_{\bar{\tau}_1 t_n}^\varepsilon, \bar{\tau}_1 t_n; Y_{\bar{\tau}_2 t_n}^\varepsilon, \bar{\tau}_2 t_n\big)\lambda_{n+1}, \quad i = 1, 2,$$
$$X_{\bar{\tau}_1 t}^\varepsilon := \exp_{X_{\bar{\tau}_1 t_n}^\varepsilon}^{(\bar{\tau}_1 t_n)}\left(\frac{t - t_n}{\varepsilon}\sqrt{2\bar{\tau}_1}\hat{\lambda}_{n+1}^{(1)}\right),$$
$$Y_{\bar{\tau}_2 t}^\varepsilon := \exp_{Y_{\bar{\tau}_2 t_n}^\varepsilon}^{(\bar{\tau}_2 t_n)}\left(\frac{t - t_n}{\varepsilon}\sqrt{2\bar{\tau}_2}\hat{\lambda}_{n+1}^{(2)}\right)$$

for $t \in [t_n, t_{n+1}]$. We can (and we will) extend the definition of X_τ^ε for $\tau \in [T\bar{\tau}_1/\bar{\tau}_2, T]$ in the same way. As in Section 3, $X_{\bar{\tau}_1 t}^\varepsilon$ does not move when $\bar{\tau}_1 = 0$. Note that the term $\sqrt{d+2}$ in the definition of $\hat{\lambda}_{n+1}^{(i)}$ is a normalization factor in the sense $\mathrm{Cov}(\sqrt{d+2}\lambda_n) = \mathrm{Id}$. Let us equip path spaces $C([a, b] \to M)$ or $C([a, b] \to M \times M)$ with the uniform convergence topology induced from $g(T)$. Here the interval $[a, b]$ will be chosen appropriately in each context. As shown in [14], $(X_\tau^\varepsilon)_{\tau\in[\bar{\tau}_1 s,T]}$ and $(Y_\tau^\varepsilon)_{\tau\in[\bar{\tau}_2 s,T]}$ converge in law to $g(\tau)$-Brownian motions $(X_\tau)_{\tau\in[\bar{\tau}_1 s,T]}$ and $(Y_\tau)_{\tau\in[\bar{\tau}_2 s,T]}$ on M with initial conditions $X_{\bar{\tau}_1 s} = x_1$, $Y_{\bar{\tau}_2 s} = y_1$ respectively as $\varepsilon \to 0$ (when $\bar{\tau}_1 > 0$). As a result, $\mathbf{X}^\varepsilon$ is tight and hence there is a convergent subsequence of $\mathbf{X}^\varepsilon$. We fix such a subsequence and use the same symbol $(\mathbf{X}^\varepsilon)_\varepsilon$ for simplicity of notations. We denote the limit in law of $\mathbf{X}^\varepsilon$ as $\varepsilon \to 0$ by $\mathbf{X}_t = (X_{\bar{\tau}_1 t}, Y_{\bar{\tau}_2 t})$. Recall that, in this paper, $g(\tau)$-Brownian motion means a time-inhomogeneous diffusion process associated with $\Delta_{g(\tau)}$ instead of $\Delta_{g(\tau)}/2$.

Remark 7. We explain the reason why our alternative construction works efficiently to avoid the obstruction arising from singularity of L. To make it clear, we begin with observing the essence of difficulties in the SDE approach we used in Section 3. Recall that our argument is based on the Itô formula. Hence the non-differentiability of L at the $\mathcal{L}$-cut locus causes the technical difficulty. One possible strategy is to extend the Itô formula for $\mathcal{L}$-distance. Since $\mathcal{L}$-cut locus is sufficiently thin, we can expect that the totality of times when our coupled particles stay there has measure zero. In addition, as that of Riemannian cut locus, the presence of $\mathcal{L}$-cut locus would work to decrease the $\mathcal{L}$-distance between coupled particles. Thus one might think it possible to extend Itô formula for $\mathcal{L}$-distance to the one involving a "local time at the $\mathcal{L}$-cut locus". If we succeed in doing so, we will obtain a differential inequality which implies the supermartingale property by neglecting this additional term since it would be non-positive.

Instead of completing the above strategy, our alternative approach in this section directly provides a difference inequality without extracting the additional "local time" term. When the endpoint of minimal $\mathcal{L}$-geodesic is in the $\mathcal{L}$-cut locus, we divide it into two pieces. Then the pair of endpoints of each piece is not in the $\mathcal{L}$-cut locus. As a result, we obtain the desired difference inequality of $\mathcal{L}$-distance even in such a case (see the proof of Lemma 4 for more details). In order to follow such a procedure, it is more suitable to work with discrete time processes.

5.1. Preliminaries on properties of $\mathcal{L}$-functional

Recall that we assumed the uniform lower Ricci curvature bound (2). On the basis of it, we can compare Riemannian metrics at different times. That is, for $\tau_1 < \tau_2$,

$$g(\tau_1) \leqslant e^{2K(\tau_2-\tau_1)} g(\tau_2) \tag{23}$$

(see [29, Lemma 5.3.2], for instance). Recall that $\rho_{g(\tau)}$ is the distance function on M at time τ. Note that a similar comparison between $\rho_{g(\tau_1)}$ and $\rho_{g(\tau_2)}$ follows from (2). By neglecting the term involving $\dot{\gamma}$ in the definition of $L(\gamma)$, the condition (2) implies

$$\inf_{x,y\in M} L(x,\tau_1;y,\tau_2) \geqslant -\frac{2dK}{3}\big(\tau_2^{3/2}-\tau_1^{3/2}\big). \tag{24}$$

We also obtain the following bounds for L from (2) and (23). Let $\gamma : [\tau_1,\tau_2] \to M$ be a minimal $\mathcal{L}$-geodesic. Then, for $\tau \in [\tau_1,\tau_2]$,

$$\frac{e^{-2KT}}{2(\sqrt{\tau_2}-\sqrt{\tau_1})}\rho_{g(\bar{\tau}_1)}\big(\gamma(\tau_1),\gamma(\tau)\big)^2 - \frac{2}{3}dK\big(\tau_2^{3/2}-\tau_1^{3/2}\big) \leqslant L\big(\gamma(\tau_1),\tau_1;\gamma(\tau_2),\tau_2\big) \tag{25}$$

(see [6, Lemma 7.13] or [30, Proposition B.2]). The same estimate holds for $\rho_{g(\bar{\tau}_1)}(\gamma(\tau),\gamma(\tau_2))^2$ instead of $\rho_{g(\bar{\tau}_1)}(\gamma(\tau_1),\gamma(\tau))^2$. Taking the fact that $\mathcal{L}$-functional is *not* invariant under reparametrization of curves into account, we will introduce a local estimate on the velocity of the minimal $\mathcal{L}$-geodesic γ.

Lemma 3. *Let $\tau_1,\tau_2 \in [\bar{\tau}_1,T]$ and suppose that $\tau_2-\tau_1 \geqslant \delta$ for some $\delta > 0$. Then, for any compact set $M_0 \subset M$, there exist constants $C_1 > 0$ depending on K, M_0 and δ such that, for any $\gamma : [\tau_1,\tau_2] \to M$ with $\gamma(\tau_1),\gamma(\tau_2) \in M_0$ and $\tau_1 \leqslant \tau \leqslant \tau_2$,*

$$\tau\big|\dot{\gamma}(\tau)\big|^2_{g(\tau)} \leqslant C_1. \tag{26}$$

Proof. Though the conclusion follows by combining arguments in [6, Lemma 7.24] and [35, Proposition 2.12], we give a proof for completeness. Let $o \in M$ be a reference point and take $r_0 > 0$ so large that $B^{g(T)}_{r_0/2}(o)$ contains M_0. Take $K_0 > 0$ so that $\sup_\tau |R_{g(\tau)}| \leqslant K_0$ holds on $B^{g(T)}_{r_0}(o)$. We claim that there exists a constant $C_0 > 0$ such that

$$L(x,\tau_1;y,\tau_2) \leqslant C_0 \tag{27}$$

for any $x,y \in M_0$. Take a constant speed $g(T)$-minimal geodesic $\gamma_0 : [\tau_1,\tau_2] \to M$ joining x and y. Note that γ_0 is contained in $B^{g(T)}_{r_0}(o)$. Thus, by virtue of (23), we have

$$
\begin{aligned}
L(x,\tau_1;y,\tau_2) &\leqslant \int_{\tau_1}^{\tau_2} \sqrt{\tau}\big(\big|\gamma_0(\tau)\big|_{g(\tau)}^2 + R_{g(\tau)}\big(\gamma_0(\tau)\big)\big)\,d\tau \\
&\leqslant \frac{2\mathrm{e}^{2KT}}{3}\frac{\tau_2^{3/2}-\tau_1^{3/2}}{(\tau_2-\tau_1)^2} d_{g(T)}(x,y)^2 + \frac{2K_0}{3}\big(\tau_2^{3/2}-\tau_1^{3/2}\big) \\
&\leqslant \frac{2T^{3/2}\mathrm{e}^{2KT}}{3\delta^2} d_{g(T)}(x,y)^2 + \frac{2K_0}{3}T^{3/2}.
\end{aligned}
$$

Thus the claim follows since $x, y \in M_0$.

By combining the above claim with (25), we can show that there exists $r_1 > r_0$ which is independent of γ such that $\gamma(\tau) \in B_{r_1}^{g(T)}(o)$ for any $\tau \in [\tau_1,\tau_2]$. Take $K_1 > 0$ so that $|\mathrm{Ric}_{g(\tau)}|_{g(\tau)} \leqslant K_1$ and $|\nabla R_{g(\tau)}|_{g(\tau)} \leqslant K_1$ hold on $B_{r_1}^{g(T)}(o)$ for any $\tau \in [\bar{\tau}_1, T]$. By a similar argument as in [6, Lemma 7.13(ii)], there exists $\tau^* \in [\tau_1,\tau_2]$ such that

$$
\tau^*\big|\dot{\gamma}\big(\tau^*\big)\big|_{g(\tau^*)}^2 \leqslant \frac{1}{2(\sqrt{\tau_2}-\sqrt{\tau_1})}\bigg(L\big(\gamma(\tau_1),\tau_1;\gamma(\tau_2),\tau_2\big) + \frac{2dK}{3}\big(\tau_2^{3/2}-\tau_1^{3/2}\big)\bigg). \tag{28}
$$

By virtue of (2), there exist constants $c_1', C_1' > 0$ which depends on K, K_1 and T such that for all $\tau_1', \tau_2' \in [\tau_1,\tau_2]$ with $\tau_1' < \tau_2'$,

$$
\tau_2'\big|\dot{\gamma}\big(\tau_2'\big)\big|_{g(\tau_2')}^2 \leqslant c_1'\tau_1'\big|\dot{\gamma}\big(\tau_1'\big)\big|_{g(\tau_1')}^2 + C_1', \tag{29}
$$

$$
\tau_1'\big|\dot{\gamma}\big(\tau_1'\big)\big|_{g(\tau_1')}^2 \leqslant c_1'\tau_2'\big|\dot{\gamma}\big(\tau_2'\big)\big|_{g(\tau_2')}^2 + C_1'. \tag{30}
$$

The first inequality in (29) can be shown similarly as [6, Lemma 7.24]. It is due to a differential inequality based on the $\mathcal{L}$-geodesic equation (7) which provides an upper bound of $\partial_\tau(\tau|\dot{\gamma}(\tau)|_{g(\tau)}^2)$. By considering a lower bound of the same quantity instead, we obtain the second inequality (30) in a similar way. Hence the proof is completed by combining (29) and (30) with (28) and (27). □

Let us recall that the $\mathcal{L}$-cut locus, denoted by $\mathcal{L}\mathrm{Cut}$, is defined as a union of two different kinds of sets (see [35]; see [6,30] also). The first one consists of $(x,\tau_1;y,\tau_2)$ such that there exist more than one minimal $\mathcal{L}$-geodesics joining (x,τ_1) and (y,τ_2). The second consists of $(x,\tau_1;y,\tau_2)$ such that (y,τ_2) is conjugate to (x,τ_1) along a minimal $\mathcal{L}$-geodesic with respect to $\mathcal{L}$-Jacobi field. Note that L is smooth on $M \setminus \mathcal{L}\mathrm{Cut}$ (see [35, Lemma 2.9]) and that $\mathcal{L}\mathrm{Cut}$ is closed (see [30]; though they assumed M to be compact, an extension to the non-compact case is straightforward).

5.2. *Variations of the $\mathcal{L}$-distance of coupled random walks*

For proving Theorem 2, our first task is to show a difference inequality of $\Lambda(t,\mathbf{X}_t^\varepsilon)$ in Lemma 4. We begin with introducing some notations. Set $\gamma_n := \gamma_{\mathbf{X}_{t_n}^\varepsilon}^{\bar{\tau}_1 t_n, \bar{\tau}_2 t_n}$ and let us define a vector field $\hat{\lambda}_{n+1}^\dagger$ along γ_n by $\hat{\lambda}_{n+1}(\tau) = \sqrt{\tau/t_n}\lambda_{n+1}^*(\tau)$, where λ_{n+1}^* is a space–time parallel vector field along γ_n with initial condition $\hat{\lambda}_{n+1}^*(\bar{\tau}_1 t_n) = \hat{\lambda}_{n+1}^{(1)}$. Let us define random variables ζ_n and Σ_n as follows:

$$\zeta_{n+1} := \sqrt{2\tau}\big\langle \hat{\lambda}^{\dagger}_{n+1}(\tau), \dot{\gamma}_n(\tau)\big\rangle_{g(\tau)}\Big|_{\tau=\bar{\tau}_1 t_n}^{\bar{\tau}_2 t_n},$$

$$\begin{aligned}\Sigma_{n+1} := {} & \frac{1}{t_n}\tau^{3/2}\big(R_{g(\tau)}\big(\gamma_n(\tau)\big) - \big|\dot{\gamma}_n(\tau)\big|^2_{g(\tau)}\big)\Big|_{\tau=\bar{\tau}_1 t_n}^{\bar{\tau}_2 t_n} \\ & + \left(\left(\frac{\sqrt{\tau}}{t_n} - 2\sqrt{\tau}\,\mathrm{Ric}_{g(\tau)}\big(\hat{\lambda}^{\dagger}_{n+1}(\tau), \hat{\lambda}^{\dagger}_{n+1}(\tau)\big)\right)\right|_{\tau=\bar{\tau}_1 t_n}^{\bar{\tau}_2 t_n} \\ & \left. - \int_{\bar{\tau}_1 t_n}^{\bar{\tau}_2 t_n} \sqrt{\tau} H\big(\dot{\gamma}(\tau), \hat{\lambda}^{\dagger}_{n+1}(\tau)\big)\,d\tau\right).\end{aligned}$$

Here H is as given in (13). The term ζ_{n+1} corresponds to the martingale part of $\Lambda(t, \mathbf{X}_t)$ and Σ_n does to the one dominating the bounded variation part of $\Lambda(t, \tilde{X}_t, \tilde{Y}_t)$ in Section 3. As we will see in Lemma 4 below, there is a discrete analogue of the Itô formula (and the corresponding difference inequality) involving ζ_n and Σ_n. As a result of our discretization, we are no longer able to apply Proposition 1 directly to estimate Σ_n itself. In this case, we can do it to the conditional expectation of Σ_n instead. Set $\mathcal{G}_n := \sigma(\lambda_1, \ldots, \lambda_n)$. Then, since each Φ_i is isometry and $(d + 2)\mathbb{E}[\langle \lambda_n, e_i\rangle\langle \lambda_n, e_j\rangle] = \delta_{ij}$, Proposition 1 yields

$$\mathbb{E}[\Sigma_{n+1} \mid \mathcal{G}_n] \leqslant \frac{d}{\sqrt{t_n}}(\sqrt{\bar{\tau}_2} - \sqrt{\bar{\tau}_1}) - \frac{1}{2t_n}\Lambda\big(t_n, \mathbf{X}^{\varepsilon}_{t_n}\big). \tag{31}$$

For $M_0 \subset M$, we define $\sigma_{M_0} : C([s, T/\bar{\tau}_2] \to M \times M) \to [0, \infty)$ by

$$\sigma_{M_0}(w, \tilde{w}) := \inf\{t \geqslant s \mid w_{\bar{\tau}_1 t} \notin M_0 \text{ or } w_{\bar{\tau}_2 t} \notin M_0\}.$$

For simplicity of notations, we denote $\sigma_{M_0}(\mathbf{X}^{\varepsilon})$ and $\sigma_{M_0}(\mathbf{X})$ by $\sigma^{\varepsilon}_{M_0}$ and $\sigma^{0}_{M_0}$ respectively. As shown in [14], for any $\eta > 0$, we can take a compact set $M_0 \subset M$ such that $\lim_{\varepsilon\to 0}\mathbb{P}[\sigma^{\varepsilon}_{M_0} \leqslant T] \leqslant \eta$ holds (cf. [16]).

Lemma 4. *Let $M_0 \subset M$ be a compact set. Then there exist a family of random variables $(Q^{\varepsilon}_n)_{n\in\mathbb{N},\,\varepsilon>0}$ and a family of deterministic constants $(\delta(\varepsilon))_{\varepsilon>0}$ with $\lim_{\varepsilon\to 0}\delta(\varepsilon) = 0$ satisfying*

$$\sum_{n;\, t_n < \sigma^{\varepsilon}_{M_0} \wedge (T/\bar{\tau}_2)} Q^{\varepsilon}_n \leqslant \delta(\varepsilon) \tag{32}$$

such that

$$\Lambda\big(t_{n+1}, \mathbf{X}^{\varepsilon}_{t_{n+1}}\big) \leqslant \Lambda\big(t_n, \mathbf{X}^{\varepsilon}_{t_n}\big) + \varepsilon\zeta_{n+1} + \varepsilon^2\Sigma_{n+1} + Q^{\varepsilon}_{n+1}. \tag{33}$$

Proof. When $(X^{\varepsilon}(\bar{\tau}_1 t_n), \bar{\tau}_1 t_n; Y^{\varepsilon}(\bar{\tau}_2 t_n), \bar{\tau}_2 t_n) \notin \mathcal{L}\mathrm{Cut}$, the inequality (33) follows from the Taylor expansion with the error term $Q^{\varepsilon}_{n+1} = o(\varepsilon^2)$. Indeed, the first variation formula ([6, Lemma 7.15], cf. (8)) produces $\varepsilon\zeta_{n+1}$ and Corollary 2 together with (9) implies the bound $\varepsilon^2\Sigma_{n+1}$ of the second-order term. To include the case $(X^{\varepsilon}(\bar{\tau}_1 t_n), \bar{\tau}_1 t_n; Y^{\varepsilon}(\bar{\tau}_2 t_n), \bar{\tau}_2 t_n) \in \mathcal{L}\mathrm{Cut}$ and to obtain a uniform bound (32), we extend this argument. Set $\tau^*_n := (\bar{\tau}_1 + \bar{\tau}_2)t_n/2$. Then we can show

$$\big(X^{\varepsilon}_{\bar\tau_1 t_n}, \bar\tau_1 t_n; \gamma_n\big(\tau^*_n\big), \tau^*_n\big) \notin \mathcal{L}\mathrm{Cut},$$
$$\big(\gamma_n\big(\tau^*_n\big), \bar\tau^*_n; X^{\varepsilon}_{\bar\tau_2 t_n}, \bar\tau_2 t_n\big) \notin \mathcal{L}\mathrm{Cut}$$

since minimal $\mathcal{L}$-geodesics with these pair of endpoints can be extended with keeping its minimality (cf. see [6, Section 7.8] and [35]). Set $x^*_{n+1} = \exp^{(\tau^*_n)}_{\gamma_n(\tau^*_n)}(\sqrt{\bar\tau_1 + \bar\tau_2}\lambda^{\dagger}_{n+1}(\tau^*_n))$. The triangle inequality for L yields

$$\Lambda\big(t_n, \mathbf{X}^{\varepsilon}_{t_n}\big) = L\big(X^{\varepsilon}_{\bar\tau_1 t_n}, \bar\tau_1 t_n; \gamma_n\big(\tau^*_n\big), \tau^*_n\big) + L\big(\gamma_n\big(\tau^*_n\big), \tau^*_n; X^{\varepsilon}_{\bar\tau_2 t_n}, \bar\tau_2 t_n\big),$$
$$\Lambda\big(t_{n+1}, \mathbf{X}^{\varepsilon}_{t_{n+1}}\big) \leqslant L\big(X^{\varepsilon}_{\bar\tau_1 t_{n+1}}, \bar\tau_1 t_{n+1}; x^*_{n+1}, \tau^*_{n+1}\big) + L\big(x^*_{n+1}, \tau^*_{n+1}; X^{\varepsilon}_{\bar\tau_2 t_{n+1}}, \bar\tau_2 t_{n+1}\big).$$

Hence

$$\Lambda\big(t_{n+1}, \mathbf{X}^{\varepsilon}_{t_{n+1}}\big) - \Lambda\big(t_n, \mathbf{X}^{\varepsilon}_{t_n}\big) \leqslant \big(L\big(X^{\varepsilon}_{\bar\tau_1 t_{n+1}}, \bar\tau_1 t_{n+1}; x^*_{n+1}, \tau^*_{n+1}\big) - L\big(X^{\varepsilon}_{\bar\tau_1 t_n}, \bar\tau_1 t_n; \gamma_n\big(\tau^*_n\big), \tau^*_n\big)\big)$$
$$+ \big(L\big(x^*_{n+1}, \tau^*_{n+1}; X^{\varepsilon}_{\bar\tau_2 t_{n+1}}, \bar\tau_2 t_{n+1}\big) - L\big(\gamma_n\big(\tau^*_n\big), \tau^*_n; X^{\varepsilon}_{\bar\tau_2 t_n}, \bar\tau_2 t_n\big)\big)$$

and the desired inequality with $Q^{\varepsilon}_n = o(\varepsilon^2)$ holds by applying the Taylor expansion to each term on the right-hand side of the above inequality.

We turn to showing the claimed control (32) of the error term Q^{ε}_n. Take a compact set $M_1 \supset M_0$ such that every minimal $\mathcal{L}$-geodesic joining $(x, \bar\tau_1 t)$ and $(y, \bar\tau_2 t)$ is included in M_1 if $x, y \in M_0$ and $t \in [s, T/\bar\tau_2]$. Indeed, such M_1 exists since we have the lower bound of L in (25) and L is continuous. Let us define a set A by

$$A := \big\{\big((\tau_1, x), (\tau_3, z), (\tau_2, y)\big) \in \big([\bar\tau_1, T] \times M_1\big)^3 \bigm| x, y \in M_0,\ \tau_2 - \tau_1 \geqslant (\bar\tau_2 - \bar\tau_1)s,$$
$$\tau_3 = (\tau_1 + \tau_2)/2,\ L(x, \tau_1; z, \tau_3) + L(z, \tau_3; y, \tau_2) = L(x, \tau_1; y, \tau_2)\big\}.$$

Note that A is compact. Let $\pi_1, \pi_2 : A \to ([\bar\tau_1, T] \times M_1)^2$ be defined by

$$\pi_1\big((\tau_1, x), (\tau_3, z), (\tau_2, y)\big) := \big((\tau_1, x), (\tau_3, z)\big),$$
$$\pi_2\big((\tau_1, x), (\tau_3, z), (\tau_2, y)\big) := \big((\tau_3, z), (\tau_2, y)\big).$$

Then $\pi_1(A)$ and $\pi_2(A)$ are compact and $\pi_i(A) \cap \mathcal{L}\mathrm{Cut} = \emptyset$ for $i = 1, 2$. The second assertion comes from the fact that (z, τ_3) is on a minimal $\mathcal{L}$-geodesic joining (x, τ_1) and (y, τ_2) for $((x, \tau_1), (z, \tau_3), (y, \tau_2)) \in A$. Recall that $\mathcal{L}\mathrm{Cut}$ is closed. Thus we can take relatively compact open sets $G_1, G_2 \subset [\bar\tau_1, T] \times M$ such that $\pi_i(A) \subset G_i$ and $\bar G_i \cap \mathcal{L}\mathrm{Cut} = \emptyset$ for $i = 1, 2$. Then the Taylor expansion we discussed above can be done on G_1 or G_2 for sufficiently small ε. Recall that L is smooth outside of $\mathcal{L}\mathrm{Cut}$ (see [6]). Thus the convergence $\varepsilon^{-2} Q_n(\varepsilon) \to 0$ as $\varepsilon \to 0$ is uniform in n and independent of $\mathbf{X}^{\varepsilon}_{t_n}$ as long as $t_n < \sigma^{\varepsilon}_{M_0} \wedge (T/\bar\tau_2)$. Since the cardinality of $\{n \mid t_n < \sigma^{\varepsilon}_{M_0} \wedge (T/\bar\tau_2)\}$ is of order at most ε^{-2}, the assertion (32) holds. □

We next establish the corresponding difference inequality for $\Theta_t(\mathbf{X}^{\varepsilon}_t)$ (Corollary 3). For that, we show the following auxiliary lemma.

Lemma 5. *Let $M_0 \subset M$ be a compact set. Then there exists a deterministic constant $C_2 > 0$ depending on M_0 such that* $\max\{|\zeta_n|, |\Lambda(t_n, \mathbf{X}^{\varepsilon}_{t_n})|, |\Sigma_n|\} \leqslant C_2$ *holds if* $t_n \leqslant \sigma^{\varepsilon}_{M_0} \wedge (T/\bar\tau_2)$.

Proof. By the definition of ζ_n, we have

$$|\zeta_n| \leqslant \sqrt{2(d+2)t_{n-1}}\big(\bar{\tau}_1\big|\dot{\gamma}_{n-1}(\bar{\tau}_1 t_{n-1})\big|_{g(\bar{\tau}_1 t_{n-1})} + \bar{\tau}_2\big|\dot{\gamma}_{n-1}(\tau_2 t_{n-1})\big|_{g(\bar{\tau}_2 t_{n-1})}\big).$$

Thus the asserted bound for $|\zeta_n|$ follows from (26). Similarly, the estimate for $\Lambda(t_n, \mathbf{X}^\varepsilon_{t_n})$ follows from (24) and (27). For estimating Σ_n, we deal with the integral involving H in the definition of Σ_n. Note that every tensor field appearing in the definition of H is continuous. As in the proof of Lemma 4, take a compact set $M_1 \supset M_0$ such that every minimal $\mathcal{L}$-geodesic joining $(x, \bar{\tau}_1 t)$ and $(y, \bar{\tau}_2 t)$ is included in M_1 if $x, y \in M_0$ and $t \in [s, T/\bar{\tau}_2]$. Since $\mathbf{X}^\varepsilon_{t_{n-1}} \in M_0 \times M_0$ holds on the event $\{t_n < \sigma^\varepsilon_{M_0} \wedge (T/\bar{\tau}_2)\}$, the upper bound (26) of $\sqrt{\tau}|\dot{\gamma}(\tau)|$ implies that $H(\dot{\gamma}_n(\tau), Z(\tau))$ is uniformly bounded for any vector field $Z(\tau)$ along γ_n of the form $Z(\tau) = \sqrt{\tau/t_n}Z^*(\tau)$ with a space–time parallel vector field $Z^*(\tau)$ satisfying $|Z^*(\tau)|_{g(\tau)} \leqslant 1$. This fact yields a required bound for the integral. For any other terms in the definition of Σ_n, we can estimate them as we did for ζ_n and $\Lambda(t_n, \mathbf{X}^\varepsilon_{t_n})$. $\square$

By virtue of Lemma 5, Lemma 4 yields the following:

Corollary 3. *Let $M_0 \subset M$ be a compact set. Then there exist a family of random variables $(\tilde{Q}^\varepsilon_n)_{n\in\mathbb{N},\,\varepsilon>0}$ and a family of deterministic constants $(\tilde{\delta}(\varepsilon))_{\varepsilon>0}$ with $\lim_{\varepsilon\to 0}\tilde{\delta}(\varepsilon) = 0$ satisfying*

$$\sum_{n;\, t_n < \sigma^\varepsilon_{M_0} \wedge (T/\bar{\tau}_2)} \tilde{Q}^\varepsilon_n \leqslant \tilde{\delta}(\varepsilon)$$

such that

$$\begin{aligned}
\Theta_{t_{n+1}}\big(\mathbf{X}^\varepsilon_{t_{n+1}}\big) \leqslant{}& \Theta_{t_n}\big(\mathbf{X}^\varepsilon_{t_n}\big) + \frac{\varepsilon^2}{\sqrt{t_n}}(\sqrt{\bar{\tau}_2} - \sqrt{\bar{\tau}_1})\Lambda\big(t_n, \mathbf{X}^\varepsilon_{t_n}\big) - 2\varepsilon^2 d(\sqrt{\bar{\tau}_2} - \sqrt{\bar{\tau}_1})^2 \\
&+ 2\varepsilon(\sqrt{\bar{\tau}_2 t_{n+1}} - \sqrt{\bar{\tau}_1 t_{n+1}})\zeta_{n+1} + 2\varepsilon^2(\sqrt{\bar{\tau}_2 t_{n+1}} - \sqrt{\bar{\tau}_1 t_{n+1}})\Sigma_{n+1} \\
&+ \tilde{Q}^\varepsilon_{n+1}.
\end{aligned} \tag{34}$$

For $u \in [s, T/\bar{\tau}_2]$, let us define $\lfloor u \rfloor_\varepsilon$ by

$$\lfloor u \rfloor_\varepsilon := \sup\big\{s + \varepsilon^2 n \;\big|\; n \in \mathbb{N} \cup \{0\},\ 1 + \varepsilon^2 n < u\big\}.$$

Set $\hat{\sigma}^\varepsilon_{M_0} := \lfloor \sigma^\varepsilon_{M_0} \rfloor_\varepsilon + \varepsilon^2$. Note that $\{\hat{\sigma}^\varepsilon_{M_0} = t_n\} \in \mathcal{G}_n$ for all $n \in \mathbb{N} \cup \{0\}$. We finally prepare the following moment bound of $\Theta_t(\mathbf{X}_t)$ before entering the proof of Theorem 2.

Lemma 6. *There exist $c_3, C_3 > 0$ such that*

$$\mathbb{E}\Big[\sup_{s \leqslant t \leqslant T/\bar{\tau}_2} \Theta_t(\mathbf{X}_t)^2\Big] < c_3 \Theta_s(x_1, y_1)^2 + C_3.$$

Proof. Recall that Θ is uniformly bounded from below by (24). Take $b \geqslant 0$ so large that $\Theta_t(x, y) + b \geqslant 0$ for $x, y \in M$ and $t \in [s, T/\bar{\tau}_2]$ and set $\hat{\Theta}_t(x, y) := \Theta_t(x, y) + b$. It suffices to show

$$\mathbb{E}\Big[\sup_{s\leqslant t\leqslant T/\bar{\tau}_2}\hat{\Theta}_t(\mathbf{X}_t)^2\Big]\leqslant 4\hat{\Theta}_s(x_1,y_1)^2+C$$

for some $C>0$. Take a relatively compact open set $M_0\subset M$ and consider $\hat{\Theta}_{t_n\wedge\hat{\sigma}^\varepsilon_{M_0}}(\mathbf{X}^\varepsilon_{t_n\wedge\hat{\sigma}^\varepsilon_{M_0}})$. Lemma 5 ensures that the term appearing in (34) is integrable on the event $t_n<\hat{\sigma}^\varepsilon_{M_0}$. Thus Corollary 3 and (31) yield that

$$\mathbb{E}\big[\hat{\Theta}_{t_m\wedge\hat{\sigma}^\varepsilon_{M_0}}\big(\mathbf{X}^\varepsilon_{t_m\wedge\hat{\sigma}^\varepsilon_{M_0}}\big)-\hat{\Theta}_{t_n\wedge\hat{\sigma}^\varepsilon_{M_0}}\big(\mathbf{X}^\varepsilon_{t_n\wedge\hat{\sigma}^\varepsilon_{M_0}}\big)\bigm|\mathcal{G}_n\big]\leqslant\tilde{\delta}(\varepsilon).$$

By imitating the proof of the maximal inequality (cf. [27, Chapter 2, Exercise 1.15]), we obtain

$$\mathbb{P}\Big[\sup_{n;\,s\leqslant t_n\leqslant\hat{\sigma}^\varepsilon_{M_0}\wedge(T/\bar{\tau}_2)}\hat{\Theta}_{t_n}\big(\mathbf{X}^\varepsilon_{t_n}\big)\geqslant r\Big]\leqslant\frac{1}{r}\big(\hat{\Theta}_s(x_1,y_1)+\tilde{\delta}(\varepsilon)\big)$$

for $r>0$. Then, by following the proof of the Doob inequality in [27],

$$\mathbb{E}\Big[\sup_{s\leqslant t_n\leqslant\hat{\sigma}^\varepsilon_{M_0}\wedge(T/\bar{\tau}_2)}\big(\hat{\Theta}_{t_n}\big(\mathbf{X}^\varepsilon_{t_n}\big)\wedge R\big)^2\Big]\leqslant 4\big(\hat{\Theta}_s(x_1,y_1)+\tilde{\delta}(\varepsilon)\big)^2 \tag{35}$$

holds for each $R>0$. By (23) and the definition of X^ε, there exist $C_{M_0}>0$ such that

$$\big(\hat{\Theta}_{t\wedge\sigma^\varepsilon_{M_0}}\big(\mathbf{X}^\varepsilon_t\big)\wedge R\big)^2\leqslant\big(\hat{\Theta}_{t_n}\big(\mathbf{X}^\varepsilon_{t_n\wedge\hat{\sigma}^\varepsilon_{M_0}}\big)\wedge R\big)^2+C_{M_0}\varepsilon$$

for $t\in[t_n,t_{n+1}]$. Thus (35) yields

$$\mathbb{E}\Big[\sup_{s\leqslant t\leqslant\sigma^\varepsilon_{M_0}\wedge(T/\bar{\tau}_2)}\big(\hat{\Theta}_t\big(\mathbf{X}^\varepsilon_t\big)\wedge R\big)^2\Big]\leqslant 4\big(\hat{\Theta}_s(x_1,y_1)+\tilde{\delta}(\varepsilon)\big)^2+C_{M_0}\varepsilon. \tag{36}$$

Let us turn to estimate the second moment of $\sup_t\hat{\Theta}_t(\mathbf{X}_t)$. Note that we have

$$\mathbb{E}\Big[\sup_{s\leqslant t\leqslant T/\bar{\tau}_2}\big(\hat{\Theta}_t(\mathbf{X}_t)\wedge R\big)^2\Big]\leqslant\mathbb{E}\Big[\sup_{s\leqslant t\leqslant T/\bar{\tau}_2}\big(\hat{\Theta}_t(\mathbf{X}_t)\wedge R\big)^2;\sigma^0_{M_0}>t\Big]+R^2\mathbb{P}\big[\sigma^0_{M_0}\leqslant t\big]. \tag{37}$$

Since $\{\mathbf{w}\mid\sigma_{M_0}(\mathbf{w})>t\}$ is open and the map $\mathbf{w}\mapsto\sup_{s\leqslant t\leqslant T/\bar{\tau}_2}(\hat{\Theta}_t(\mathbf{w}_t)\wedge R)^2$ is bounded and continuous on $C([s,T/\bar{\tau}_2]\to M\times M)$, (36) yields

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\leqslant t\leqslant T/\bar{\tau}_2}\big(\hat{\Theta}_t(\mathbf{X}_t)\wedge R\big)^2;\sigma^0_{M_0}>t\Big]&\leqslant\liminf_{\varepsilon\to0}\mathbb{E}\Big[\sup_{s\leqslant t\leqslant T/\bar{\tau}_2}\big(\hat{\Theta}_t\big(\mathbf{X}^\varepsilon_t\big)\wedge R\big)^2;\sigma^\varepsilon_{M_0}>t\Big]\\&\leqslant\liminf_{\varepsilon\to0}\mathbb{E}\Big[\sup_{s\leqslant t\leqslant\sigma^\varepsilon_{M_0}\wedge(T/\bar{\tau}_2)}\big(\hat{\Theta}_t\big(\mathbf{X}^\varepsilon_t\big)\wedge R\big)^2\Big]\\&\leqslant 4\hat{\Theta}_s(x_1,y_1).\end{aligned}$$

Thus the conclusion follows by combining the last inequality with (37) and by letting $M_0\uparrow M$ and $R\to\infty$. □

5.3. Proof of Theorems 2 and 3

Proof of Theorem 2. First we remark that the map $(x, y) \mapsto (X^{\varepsilon}_{\bar\tau_1 \cdot}, Y^{\varepsilon}_{\bar\tau_2 \cdot})$ is obviously measurable. Thus, we obtain the same measurability for the law of $(X_{\bar\tau_1 \cdot}, Y_{\bar\tau_2 \cdot})$. The integrability of $\Theta_t(\mathbf{X}_t)$ follows from Lemma 6. We will show the supermartingale property in the sequel. For $s \leqslant s_1 < \cdots < s_m < t' < t < T$ and $f_1, \ldots, f_m \in C_c(M \times M \to \mathbb{R})$ with $0 \leqslant f_j \leqslant 1$, set $F(\mathbf{w}) := \prod_{j=1}^m f_j(\mathbf{w}_{s_j})$ for $\mathbf{w} \in C([s, T/\bar\tau_2] \to M \times M)$. Take $\eta > 0$ arbitrarily and choose a relatively compact open set $M_0 \subset M$ so that $\mathbb{P}[\sigma^0_{M_0} \leqslant t] \leqslant \eta$ holds. Note that $\limsup_{\varepsilon \to 0} \mathbb{P}[\sigma^{\varepsilon}_{M_0} \leqslant t] \leqslant \eta$ also holds since $\{w \mid \sigma_{M_0}(w) \leqslant t\}$ is closed. It suffices to show that there is a constant $C > 0$ which is independent of η and M_0 such that,

$$\mathbb{E}\big[\big(\Theta_{t\wedge\sigma^0_{M_0}}(\mathbf{X}_{t\wedge\sigma^0_{M_0}}) - \Theta_{t'\wedge\sigma^0_{M_0}}(\mathbf{X}_{t'\wedge\sigma^0_{M_0}})\big)F(\mathbf{X}_{\cdot\wedge\sigma^0_{M_0}})\big] \leqslant C\sqrt{\eta} \tag{38}$$

holds. In fact, once we have shown (38), then Lemma 6 yields

$$\mathbb{E}\big[\big(\Theta_t(\mathbf{X}_t) - \Theta_s(\mathbf{X}_s)\big)F(\mathbf{X})\big] \leqslant 0$$

since $\sigma^0_{M_0} \to \infty$ almost surely as $M_0 \uparrow M$.

Take $f \in C_c(M \times M)$ such that $0 \leqslant f \leqslant 1$ and $f|_U \equiv 1$, where $U \subset M \times M$ is an open set containing $\bar M_0 \times \bar M_0$. Then, by virtue of Lemma 6 and the choice of M_0,

$$\begin{aligned}
&\mathbb{E}\big[\big(\Theta_{t\wedge\sigma^0_{M_0}}(\mathbf{X}_{t\wedge\sigma^0_{M_0}}) - \Theta_{t'\wedge\sigma^0_{M_0}}(\mathbf{X}_{t'\wedge\sigma^0_{M_0}})\big)F(\mathbf{X}_{\cdot\wedge\sigma^0_{M_0}})\big] \\
&\quad \leqslant \mathbb{E}\big[\big(\Theta_t(\mathbf{X}_t) - \Theta_{t'}(\mathbf{X}_{t'})\big)f(\mathbf{X}_t)f(\mathbf{X}_{t'})F(\mathbf{X}); \sigma^0_{M_0} > t\big] + 2C_4^{1/2}\sqrt{\eta},
\end{aligned} \tag{39}$$

where $C_4 := c_3\Theta_s(x_1, y_1)^2 + C_3$. Since $\{\mathbf{w} \mid \sigma_{M_0}(\mathbf{w}) > t\}$ is open,

$$\begin{aligned}
&\mathbb{E}\big[\big(\Theta_t(\mathbf{X}_t) - \Theta_{t'}(\mathbf{X}_{t'})\big)f(\mathbf{X}_t)f(\mathbf{X}_{t'})F(\mathbf{X}); \sigma^0_{M_0} > t\big] \\
&\quad \leqslant \liminf_{\varepsilon\to 0}\mathbb{E}\big[\big(\Theta_t\big(\mathbf{X}^{\varepsilon}_t\big) - \Theta_{t'}\big(\mathbf{X}^{\varepsilon}_{t'}\big)\big)f\big(\mathbf{X}^{\varepsilon}_t\big)f\big(\mathbf{X}^{\varepsilon}_{t'}\big)F\big(\mathbf{X}^{\varepsilon}\big); \sigma^{\varepsilon}_{M_0} > t\big] \\
&\quad = \liminf_{\varepsilon\to 0}\mathbb{E}\big[\big(\Theta_{\lfloor t\rfloor_\varepsilon}\big(\mathbf{X}^{\varepsilon}_{\lfloor t\rfloor_\varepsilon}\big) - \Theta_{\lfloor t'\rfloor_\varepsilon}\big(\mathbf{X}^{\varepsilon}_{\lfloor t'\rfloor_\varepsilon}\big)\big)f\big(\mathbf{X}^{\varepsilon}_{\lfloor t\rfloor_\varepsilon}\big)f\big(\mathbf{X}^{\varepsilon}_{\lfloor t'\rfloor_\varepsilon}\big)F\big(\mathbf{X}^{\varepsilon}\big); \sigma^{\varepsilon}_{M_0} > t\big].
\end{aligned} \tag{40}$$

Here the last equality follows from the continuity of Θ and f. Then

$$\begin{aligned}
&\mathbb{E}\big[\big(\Theta_{\lfloor t\rfloor_\varepsilon}\big(\mathbf{X}^{\varepsilon}_{\lfloor t\rfloor_\varepsilon}\big) - \Theta_{\lfloor t'\rfloor_\varepsilon}\big(\mathbf{X}^{\varepsilon}_{\lfloor t'\rfloor_\varepsilon}\big)\big)f\big(\mathbf{X}^{\varepsilon}_{\lfloor t\rfloor_\varepsilon}\big)f\big(\mathbf{X}^{\varepsilon}_{\lfloor t'\rfloor_\varepsilon}\big)F\big(\mathbf{X}^{\varepsilon}\big); \sigma^{\varepsilon}_{M_0} > t\big] \\
&\quad \leqslant \mathbb{E}\big[\big(\Theta_{\lfloor t\rfloor_\varepsilon\wedge\hat\sigma^{\varepsilon}_{M_0}}\big(\mathbf{X}^{\varepsilon}_{\lfloor t\rfloor_\varepsilon\wedge\hat\sigma^{\varepsilon}_{M_0}}\big) - \Theta_{\lfloor t'\rfloor_\varepsilon\wedge\hat\sigma^{\varepsilon}_{M_0}}\big(\mathbf{X}^{\varepsilon}_{\lfloor t'\rfloor_\varepsilon\wedge\hat\sigma^{\varepsilon}_{M_0}}\big)\big)F\big(\mathbf{X}^{\varepsilon}_{\cdot\wedge\hat\sigma^{\varepsilon}_{M_0}}\big)\big] \\
&\qquad + 2\mathbb{E}\Big[\sup_{s\leqslant u\leqslant T/\bar\tau_2}\big|\Theta_u\big(\mathbf{X}^{\varepsilon}_u\big)f\big(\mathbf{X}^{\varepsilon}_u\big)\big|^2\Big]^{1/2}\mathbb{P}\big[\sigma^{\varepsilon}_{M_0} \leqslant t\big]^{1/2}.
\end{aligned} \tag{41}$$

Since a function $\mathbf{w} \mapsto \sup_{1\leqslant u\leqslant T/\bar\tau_2}|\Theta_u(\mathbf{w}_u)f(\mathbf{w}_u)|$ on $C([s, T/\bar\tau_2] \to M \times M)$ is bounded and continuous, we have

$$\limsup_{\varepsilon\to 0}\mathbb{E}\Big[\sup_{s\leqslant u\leqslant T/\bar\tau_2}\big|\Theta_u\big(\mathbf{X}^{\varepsilon}_u\big)f\big(\mathbf{X}^{\varepsilon}_u\big)\big|^2\Big]^{1/2}\mathbb{P}\big[\sigma^{\varepsilon}_{M_0} \leqslant t\big]^{1/2} \leqslant C_4^{1/2}\sqrt{\eta}. \tag{42}$$

Now, with the aid of Lemma 5, the iteration of (34) together with (31) yields

$$\mathbb{E}\big[\big(\Theta_{\lfloor t\rfloor_\varepsilon\wedge\hat{\sigma}^\varepsilon_{M_0}}\big(\mathbf{X}^\varepsilon_{\lfloor t\rfloor_\varepsilon\wedge\hat{\sigma}^\varepsilon_{M_0}}\big)-\Theta_{\lfloor t'\rfloor_\varepsilon\wedge\hat{\sigma}^\varepsilon_{M_0}}\big(\mathbf{X}^\varepsilon_{\lfloor t'\rfloor_\varepsilon\wedge\hat{\sigma}^\varepsilon_{M_0}}\big)\big)F\big(\mathbf{X}^\varepsilon_{\cdot\wedge\hat{\sigma}^\varepsilon_{M_0}}\big)\big]\leqslant\tilde{\delta}(\varepsilon). \tag{43}$$

Here $\tilde{\delta}(\varepsilon)$ is what appeared in Corollary 3. Hence we complete the proof by combining (40), (41), (42) and (43) with (39). □

Proof of Theorem 3. Fix $1\leqslant s<t\leqslant T/\bar{\tau}_2$. We may assume

$$\mathcal{T}_{\varphi\circ\Theta_s}\big(p(\bar{\tau}_1 s,\cdot)\operatorname{vol}_{g(\bar{\tau}_1 t)},q(\bar{\tau}_2 s,\cdot)\operatorname{vol}_{g(\bar{\tau}_2 t)}\big)<\infty$$

without loss of generality. Let π be a minimizer of $\mathcal{T}_{\varphi\circ\Theta_s}(p(\bar{\tau}_1 s,\cdot)\operatorname{vol}_{g(\bar{\tau}_1 t)},q(\bar{\tau}_2 s,\cdot)\operatorname{vol}_{g(\bar{\tau}_2 t)})$, where the existence of π follows from [31, Theorem 4.1], using the lower bound (24). For each $(x,y)\in M\times M$, we take coupled Brownian motions $(X^x_\tau)_{\tau\in[\bar{\tau}_1 s,T]}$ and $(Y^y_\tau)_{\tau\in[\bar{\tau}_2 s,T]}$ with initial values $X^x_{\bar{\tau}_1 s}=x$ and $Y^y_{\bar{\tau}_2 s}=y$ as in Theorem 2. Since the law of (X^x,Y^y) is a measurable function of (x,y), we can construct a coupling of two Brownian motions $(X_{\bar{\tau}_1\cdot},Y_{\bar{\tau}_2\cdot})$ with initial distribution π from $((X^x_{\bar{\tau}_1\cdot},Y^y_{\bar{\tau}_2\cdot}))_{x,y\in M}$ as a coordinate process on $C([s,T/\bar{\tau}_2]\to M\times M)$ by following a usual manner. By Theorem 2, $\varphi(\Theta_t(X^x_{\bar{\tau}_1 t},Y^y_{\bar{\tau}_2 t}))$ is a supermartingale. Hence we have

$$\begin{aligned}\mathbb{E}\big[\varphi\big(\Theta_t(X_{\bar{\tau}_1 t},Y_{\bar{\tau}_2 t})\big)\big]&=\int_{M\times M}\mathbb{E}\big[\varphi\big(\Theta_t\big(X^x_{\bar{\tau}_1 t},Y^y_{\bar{\tau}_2 t}\big)\big)\big]\pi(dx,dy)\\&\leqslant\int_{M\times M}\varphi\big(\Theta_s(x,y)\big)\pi(dx,dy)\\&=\mathcal{T}_{\varphi\circ\Theta_s}\big(p(\bar{\tau}_1 s,\cdot)\operatorname{vol}_{g(\bar{\tau}_1 s)},q(\bar{\tau}_2 s,\cdot)\operatorname{vol}_{g(\bar{\tau}_2 s)}\big).\end{aligned}$$

Since the law of $(X_{\bar{\tau}_1 t},Y_{\bar{\tau}_2 t})$ is a coupling of $p(\bar{\tau}_1 t,\cdot)\operatorname{vol}_{g(\bar{\tau}_1 t)}$ and $q(\bar{\tau}_2 t,\cdot)\operatorname{vol}_{g(\bar{\tau}_2 t)}$, we have

$$\mathcal{T}_{\varphi\circ\Theta_t}\big(p(\bar{\tau}_1 t,\cdot)\operatorname{vol}_{g(\bar{\tau}_1 t)},q(\bar{\tau}_2 t,\cdot)\operatorname{vol}_{g(\bar{\tau}_2 t)}\big)\leqslant\mathbb{E}\big[\varphi\big(\Theta_t(X_{\bar{\tau}_1 t},Y_{\bar{\tau}_2 t})\big)\big]$$

and hence the conclusion follows. □

Acknowledgments

We would like to thank Anton Thalmaier and an anonymous referee for very useful comments.

References

[1] M. Arnaudon, K.A. Coulibaly, A. Thalmaier, Brownian motion with respect to a metric depending on time; definition, existence and application to Ricci flow, C. R. Acad. Sci. Paris, Ser. I 346 (2008) 773–778.

[2] M. Arnaudon, K.A. Coulibaly, A. Thalmaier, Horizontal diffusion in C^1 path space, in: Sém. Prob. XLIII, in: Lecture Notes in Math., vol. 2006, Springer-Verlag, Berlin, 2010, pp. 73–94.

[3] D. Bakry, On Sobolev and logarithmic Sobolev inequalities for Markov semigroups, in: New Trends in Stochastic Analysis, Charingworth, 1994, World Sci. Publ., River Edge, NJ, 1997, pp. 43–75.

[4] H.-D. Cao, X.-P. Zhu, A complete proof of the Poincaré and geometrization conjectures—application of the Hamilton–Perelman theory of the Ricci flow, Asian J. Math. 10 (2006) 165–492.

[5] B.-L. Chen, X.-P. Zhu, Uniqueness of the Ricci flow on complete noncompact manifolds, J. Differential Geom. 74 (2006) 119–154.
[6] B. Chow, S.-C. Chu, D. Glickenstein, C. Guenther, J. Isenberg, T. Ivey, D. Knopf, P. Lu, F. Luo, L. Ni, The Ricci Flow: Techniques and Applications, Part I: Geometric Aspects, American Mathematical Society, Providence, RI, 2007.
[7] K.A. Coulibaly, Brownian motion with respect to time-changing Riemannian metrics, applications to Ricci flow, Ann. Inst. Henri Poincaré Probab. Stat., in press.
[8] M. Cranston, Gradient estimates on manifolds using coupling, J. Funct. Anal. 99 (1991) 110–124.
[9] R.S. Hamilton, Three-manifolds with positive Ricci curvature, J. Differential Geom. 17 (1982) 255–306.
[10] E.P. Hsu, Stochastic Analysis on Manifolds, American Mathematical Society, Providence, RI, 2002.
[11] W.S. Kendall, Stochastic differential geometry, a coupling property, and harmonic maps, J. Lond. Math. Soc. 33 (1986) 554–566.
[12] B. Kleiner, J. Lott, Notes on Perelman's papers, Geom. Topol. 12 (2008) 2587–2855.
[13] K. Kuwada, Coupling of the Brownian motion via discrete approximation under lower Ricci curvature bounds, in: Probabilistic Approach to Geometry, Kyoto, 2008, in: Adv. Stud. Pure Math., vol. 57, Math. Soc. Japan, Tokyo, 2010, pp. 273–292.
[14] K. Kuwada, Coupling by reflection via discrete approximation under a backward Ricci flow, preprint, arXiv: 1007.0275v1, 2010.
[15] K. Kuwada, Duality on gradient estimates and Wasserstein controls, J. Funct. Anal. 258 (11) (2010) 3758–3774.
[16] K. Kuwada, R. Philipowski, Non-explosion of diffusion processes on manifolds with time-dependent metric, Math. Z., doi:10.1007/s00209-010-0704-7, in press.
[17] M. Ledoux, The geometry of Markov diffusion generators, Ann. Fac. Sci. Toulouse Math. (6) 9 (2) (2000) 305–366.
[18] J.M. Lee, Riemannian Manifolds: An Introduction to Curvature, Springer-Verlag, New York, 1997.
[19] J. Lott, Optimal transport and Perelman's reduced volume, Calc. Var. 36 (2009) 49–84.
[20] J. Lott, C. Villani, Ricci curvature for metric-measure spaces via optimal transport, Ann. Math. 169 (2009) 903–991.
[21] R.J. McCann, P. Topping, Ricci flow, entropy and optimal transportation, Amer. J. Math. 132 (2010) 711–730.
[22] J. Morgan, G. Tian, Ricci Flow and the Poincaré Conjecture, American Mathematical Society, Providence, RI, 2007.
[23] G. Perelman, The entropy formula for the Ricci flow and its geometric applications, preprint, arXiv:math/0211159v1, 2002.
[24] G. Perelman, Ricci flow with surgery on three-manifolds, preprint, arXiv:math/0303109v1, 2003.
[25] G. Perelman, Finite extinction time for the solutions to the Ricci flow on certain three-manifolds, preprint, arXiv: math/0307245v1, 2003.
[26] M.-K. von Renesse, K.-T. Sturm, Transport inequalities, gradient estimates, entropy, and Ricci curvature, Comm. Pure Appl. Math. 58 (2005) 923–940.
[27] D. Revuz, M. Yor, Continuous Martingales and Brownian Motion, 3rd edition, Springer-Verlag, Berlin, Heidelberg, 1999.
[28] W.-X. Shi, Deforming the metric on complete Riemannian manifolds, J. Differential Geom. 30 (1989) 223–301.
[29] P. Topping, Lectures on the Ricci Flow, Cambridge University Press, 2006.
[30] P. Topping, $\mathcal{L}$-optimal transportation for Ricci flow, J. Reine Angew. Math. 636 (2009) 93–122.
[31] C. Villani, Optimal Transport, Old and New, Grundlehren Math. Wiss., vol. 338, Springer-Verlag, 2009.
[32] F.-Y. Wang, Successful couplings of nondegenerate diffusion processes on compact manifolds, Acta Math. Sin. 37 (1) (1994) 116–121.
[33] F.-Y. Wang, On estimation of the logarithmic Sobolev constant and gradient estimates of heat semigroups, Probab. Theory Related Fields 108 (1) (1997) 87–101.
[34] F.-Y. Wang, Functional Inequalities Markov Semigroups and Spectral Theory, Science Press, Beijing/New York, 2005.
[35] R.G. Ye, On the l-function and the reduced volume of Perelman I, Trans. Amer. Math. Soc. 360 (1) (2008) 507–531.
[36] Qi S. Zhang, Sobolev Inequalities, Heat Kernels under Ricci Flow, and the Poincaré Conjecture, CRC Press, Boca Raton, FL, 2011.