A simple method for the optimal transportation

XIAN-TAO HUANG

In this paper, we will give a new proof of the monotonicity of Wasserstein distances of two diffusions under super Ricci flow. Our proof is based on the coupling method of B. Andrew and J. Clutterbuck (see [2]). The same method can also be applied to the contractivity of normalized $\mathscr{L}$-Wasserstein distance under backward Ricci flow.

1. Introduction

Suppose M is a compact oriented Riemannian manifold of dimension n. Let $\mu, \nu \in P(M)$ be two Borel probability measures on M. Let $c : M \times M \to \mathbb{R} \bigcup \{+\infty\}$ be a lower semi-continuous cost function. One can consider the Monge–Kantorovich minimization problem:

$$T_c(\mu, \nu) = \inf_{\pi \in \Gamma(\mu, \nu)} \int_{M \times M} c(x, y) d\pi(x, y),$$

where $\Gamma(\mu, \nu)$ is the set of Borel probability measures π on $M \times M$ which have marginals μ and ν (i.e., $\pi(A \times M) = \mu(A), \pi(M \times A) = \nu(A)$ for every Borel set $A \subset M$). In particular, when $c(x, y) = d^p(x, y)$, we have the p-Wasserstein distance ($p > 0$) between μ and ν:

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Gamma(\mu, \nu)} \int_{M \times M} d^p(x, y) d\pi(x, y) \right)^{\frac{1}{p}}.$$

Many efforts have been devoted to characterize manifolds with lower bounds for the Ricci curvature (see, e.g., [10, 13, 14] and the references therein), and some of these characterizations use optimal transportation as a key tool. For example, in [13], Sturm and von Renesse proved that $\mathrm{Ric}(M) \geq K$ is equivalent to the condition that $e^{Kt} W_p(\mu(t), \nu(t))$ is nonincreasing in t for all $p \in [1, \infty]$, where $\mu(t), \nu(t)$ are two solutions of the heat equation.

The above papers are all considered manifolds with static metrics. In the paper [11], McCann and Topping first considered the equivalent properties of super Ricci flow, i.e., a smooth family of metrics $g_\tau, \tau \in [\tau_1, \tau_2]$,

parameterized backward in time, on a compact oriented n-dimensional manifold M, satisfying

$$(1.1) \qquad -\frac{\partial}{\partial \tau} g_\tau + 2\text{Ric}(g_\tau) \geq 0.$$

Let $\mu(\tau)$, $\tau \in (\tau_1, \tau_2)$, be a family of probability measures with $d\mu(\tau) = u(\cdot, \tau)dV_\tau$ such that u satisfies the conjugate heat equation

$$(1.2) \qquad \frac{\partial u}{\partial \tau} = \Delta_\tau u - \left(\frac{1}{2} \text{tr} \frac{\partial g_\tau}{\partial \tau} \right) u,$$

where dV_τ and Δ_τ denote the volume form and Laplacian with respect to g_τ, respectively. For brevity, we will refer to such a family $\mu(\tau)$ as a diffusion throughout this paper. We will often abuse the notation of the probability μ and the volume form $d\mu$ for simplicity.

One can consider the problem of optimal transportation of two diffusions $d\mu(\tau) = u(\cdot, \tau)dV_\tau$, $d\nu(\tau) = v(\cdot, \tau)dV_\tau$. Let $\eta : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ be a smooth function with $\eta(0, \tau) = 0$. We consider a time-dependent cost function $c_\tau(x, y) = \eta(d_\tau(x, y), \tau)$, where $d_\tau(x, y)$ denotes the distance between x and y with respect to g_τ. The corresponding Monge–Kantorovich minimization problem is denoted by

$$(1.3) \qquad \text{T}_{c_\tau}(\mu(\tau), \nu(\tau)) = \inf_{\pi \in \Gamma(\mu(\tau), \nu(\tau))} \int_{M \times M} c_\tau(x, y)d\pi(x, y),$$

The corresponding p-Wasserstein distance $(p > 0)$ between $\mu(\tau)$ and $\nu(\tau)$ is denoted by $W_p(\mu(\tau), \nu(\tau))$.

In [11], the authors proved the 2-Wasserstein contractivity under super Ricci flow by calculating the derivatives of the entropy along Wasserstein geodesics. Lott [9] gave a new proof of the 2-Wasserstein contractivity under Ricci flow. The 1-Wasserstein contractivity was proved by Ilmanen, see [4]. Such monotonicity results were extended to a more general class of cost functions by Arnaudon et al. [3] using a probabilistic method.

One of the purposes of this paper is to give an alternative proof of some of the monotonicity results in [3]. Our main theorem is as follows:

Theorem 1.1. *If g_τ satisfies*

$$(1.4) \qquad -\frac{\partial}{\partial \tau} g_\tau + 2\text{Ric}(g_\tau) \geq 2K g_\tau,$$

$d\mu(\tau) = u(x,\tau)dV_\tau$, $d\nu(\tau) = v(y,\tau)dV_\tau$ *are two diffusions. Furthermore,* $\eta :$ $\mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ *satisfies*

(1.5)
$$
\begin{cases}
\eta(0,\tau) = 0, \\[4pt]
\dfrac{\partial}{\partial s}\eta(s,\tau) \geq 0, \\[6pt]
-\dfrac{\partial}{\partial \tau}\eta(s,\tau) + Ks\dfrac{\partial}{\partial s}\eta(s,\tau) - \min\left\{4\dfrac{\partial^2}{\partial s^2}\eta(s,\tau), 0\right\} \geq 0.
\end{cases}
$$

If $c_\tau(x,y) = \eta(d_\tau(x,y),\tau)$, *then* $\mathrm{T}_{c_\tau}(\mu(\tau),\nu(\tau))$ *is nonincreasing in* τ.

As a corollary, if $K = 0$ and $\frac{\partial}{\partial \tau}\eta(s,\tau) = 0$, we have

Corollary 1.2. *Suppose* g_τ *satisfies* (1.1), $d\mu(\tau)$, $d\nu(\tau)$ *are two diffusions, the function* $\eta : \mathbb{R}^+ \to \mathbb{R}^+$ *satisfies* $\eta(0) = 0, \eta' \geq 0$. *If* $c_\tau(x,y) = \eta(d_\tau(x,y))$, *then* $\mathrm{T}_{c_\tau}(\mu(\tau),\nu(\tau))$ *is nonincreasing in* τ.

For general K, if we choose $\eta(s,\tau) = e^{pK\tau}s^p$ for some $p > 0$, then the conditions (1.5) are all satisfied. Note that in this case,

$$
e^{K\tau}W_p(\mu(\tau),\nu(\tau)) = \{T_{c_\tau}(\mu(\tau),\nu(\tau))\}^{\frac{1}{p}},
$$

we have

Corollary 1.3. *If* g_τ, $d\mu(\tau)$, $d\nu(\tau)$ *are the same as in Theorem 1.1, then* $e^{K\tau}W_p(\mu(\tau),\nu(\tau))$ $(p > 0)$ *is nonincreasing in* τ.

Remark 1.4. *If the family of metrics are fixed, one can recover the corresponding monotonicity of Wasserstein distances for manifolds with Ricci curvature bounded from below, which was proved by Sturm and von Renesse [13].*

Remark 1.5. *Corollary 1.2 and Corollary 1.3 were proved originally by a probabilistic method in [3].*

The method we use originated from Andrew and Clutterbuck's papers [1, 2]. They have used this method to bound the modulus of continuity of solutions of various parabolic equations. We will construct an operator that can be interpreted as a coupling of two Laplacians, and use the parabolic maximum principle in our argument.

There is an analogous notion of $\mathscr{L}$-Wasserstein distance introduced by Topping in [15].

Suppose we have the backward Ricci flow, $\frac{\partial}{\partial\tau}g_\tau = 2\mathrm{Ric}(g_\tau)$, defined on an open interval I containing $[\bar{\tau}_1, \bar{\tau}_2]$, where $0 < \bar{\tau}_1 < \bar{\tau}_2$. Perelman's $\mathscr{L}$-length of a smooth curve $\gamma : [\tau_1, \tau_2] \to M$ ($[\tau_1, \tau_2] \subset I$) is defined to be

$$\mathscr{L}(\gamma) := \int_{\tau_1}^{\tau_2} \sqrt{\tau}(R(\gamma(\tau), \tau) + |\gamma'(\tau)|_{g_\tau}^2)d\tau,$$

where $R(x, \tau)$ denotes the scalar curvature at x in (M, g_τ). The $\mathscr{L}$-distance between a pair of points (x, τ_1) and (y, τ_2) is defined to be

$$Q(x, \tau_1; y, \tau_2)$$
$$:= \inf\left\{\mathscr{L}(\gamma)|\gamma : [\tau_1, \tau_2] \to M \text{ is smooth}, \gamma(\tau_1) = x, \gamma(\tau_2) = y\right\}.$$

One can also consider the notion of $\mathscr{L}$-Wasserstein distance $V(\nu_1(\tau_1), \tau_1; \nu_2(\tau_2), \tau_2)$ between two diffusions $\nu_1(\tau)$ and $\nu_2(\tau)$:

$$V(\nu_1(\tau_1), \tau_1; \nu_2(\tau_2), \tau_2) := \inf_{\pi \in \Gamma(\nu_1(\tau_1), \nu_2(\tau_2))} \int_{M \times M} Q(x, \tau_1; y, \tau_2)d\pi(x, y).$$

Now let $\tau_1 = \tau_1(s) := \bar{\tau}_1 e^s$, $\tau_2 = \tau_2(s) := \bar{\tau}_2 e^s$ be two exponential functions of $s \in \mathbb{R}$, and define the normalized distance between the diffusions $\nu_1(\tau_1(s))$ and $\nu_2(\tau_2(s))$ by

$$\Theta(s) := 2(\sqrt{\tau_1} - \sqrt{\tau_2})V(\nu_1(\tau_1), \tau_1; \nu_2(\tau_2), \tau_2) - 2n(\sqrt{\tau_1} - \sqrt{\tau_2})^2$$

for s in a neighborhood of 0 such that $\nu_i(\tau_i(s))$ are defined ($i = 1, 2$).

Topping proved the following theorem in [15]:

Theorem 1.6 Topping. $\Theta(s)$ *is a nonincreasing function of* s.

Kuwada and Philipowski gave a new proof based on probabilistic methods in [8]. We find that our new method also apply to this problem, see Section 4 for details.

2. A coupling method

In this section, we will define a time-dependent operator $\mathscr{D}_\tau$ which can be viewed as a coupling of two Laplacians, and calculate the action of $\mathscr{D}_\tau$ on $\eta(d_\tau(x, y), \tau)$. The definition of $\mathscr{D}_\tau$ is inspired by the paper [2].

Suppose Ω is the set $(M \times M)\backslash\{(x, x)|x \in M\}$. Denote the product metric $g_\tau \times g_\tau$ on Ω by $\tilde{g}_\tau$. Let $\tilde{\nabla}_\tau, \tilde{\nabla}_\tau^2$ be the gradient and Hessian with respect to $\tilde{g}_\tau$ respectively.

The set $\mathscr{F}_\tau \subset \mathrm{Sym}_2(T^*\Omega)$ is defined to be

$$\mathscr{F}_\tau := \big\{ A \in \mathrm{Sym}_2(T^*\Omega) \big| A \geq 0, A_{(x,y)}|_{T_x M \otimes T_x M} = g_\tau(x),$$
$$A_{(x,y)}|_{T_y M \otimes T_y M} = g_\tau(y) \big\}.$$

Obviously, $\tilde{g}_\tau \in \mathscr{F}_\tau$. Suppose $(x,y) \in \Omega$, $\{E_i\}_{1 \leq i \leq n}$ and $\{F_i\}_{1 \leq i \leq n}$ are two orthonormal bases (with respect to g_τ) defined in small open neighborhoods $U \ni x$ and $V \ni y$, respectively. Denote $\{E_*^i\}_{1 \leq i \leq n}$ and $\{F_*^i\}_{1 \leq i \leq n}$ the dual coframes of $\{E_i\}_{1 \leq i \leq n}$ and $\{F_i\}_{1 \leq i \leq n}$, respectively. Denote $\tilde{A} = \sum_{i=1}^n (E_*^i, F_*^i) \bigotimes (E_*^i, F_*^i)$ on $U \times V$. Suppose U', V' are open neighborhoods of x and y, respectively, such that $\overline{U'} \subset U$, $\overline{V'} \subset V$. Let $\alpha : \Omega \to [0,1]$ be a smooth cutoff function satisfying $\alpha(z) \equiv 1$ on $U' \times V'$, $\alpha(z) \equiv 0$ on $\Omega \setminus (U \times V)$. Define A_1 to be

$$A_1 = \alpha \tilde{A} + (1-\alpha)\tilde{g}_\tau,$$

it is easy to see $A_1 \in \mathscr{F}_\tau$. Hence we can extend $\sum_{i=1}^n (E_*^i, F_*^i) \bigotimes (E_*^i, F_*^i)|_{(x,y)}$ to an element of $\mathscr{F}_\tau$.

We define the operator $\mathscr{D}_\tau : C^2(\Omega) \to C^0(\Omega)$ to be

$$\mathscr{D}_\tau(f(x,y)) := \inf \big\{ \mathrm{tr}(A(\tilde{\nabla}_\tau^2 f)) \big| A \in \mathscr{F}_\tau \big\}$$

for $f \in C^2(\Omega)$. When f is independent of x, $\mathscr{D}_\tau(f) = \Delta_\tau f(y)$. When f is independent of y, $\mathscr{D}_\tau(f) = \Delta_\tau f(x)$. Hence $\mathscr{D}_\tau$ is a coupling of two Laplacians $\Delta_\tau|_x$ and $\Delta_\tau|_y$.

Let $x, y \in M, x \neq y$, and $d = d_\tau(x,y)$. Let $\gamma : [-\frac{d}{2}, \frac{d}{2}] \to M$ be a minimizing geodesic from x to y, parameterized by arc length in (M, g_τ). Choose an orthonormal basis $\{E_i\}_{1 \leq i \leq n}$ for $T_x M$, such that $E_n = \gamma'(-\frac{d}{2})$. Parallel transportation along γ gives an orthonormal basis $\{E_i(s)\}_{1 \leq i \leq n}$ for $T_{\gamma(s)} M$ with $E_n(s) = \gamma'(s)$.

From the definition of $\mathscr{D}_\tau$,

(2.1)

$$\mathscr{D}_\tau f \leq \min \left\{ \sum_{i=1}^n \tilde{\nabla}_\tau^2 f \left(\left(E_i\left(-\frac{d}{2}\right), E_i\left(\frac{d}{2}\right) \right), \left(E_i\left(-\frac{d}{2}\right), E_i\left(\frac{d}{2}\right) \right) \right), \right.$$
$$\sum_{i=1}^{n-1} \tilde{\nabla}_\tau^2 f \left(\left(E_i\left(-\frac{d}{2}\right), E_i\left(\frac{d}{2}\right) \right), \left(E_i\left(-\frac{d}{2}\right), E_i\left(\frac{d}{2}\right) \right) \right)$$
$$\left. + \tilde{\nabla}_\tau^2 f \left(\left(E_n\left(-\frac{d}{2}\right), -E_n\left(\frac{d}{2}\right) \right), \left(E_n\left(-\frac{d}{2}\right), -E_n\left(\frac{d}{2}\right) \right) \right) \right\},$$

where the inequality holds in support sense.

Throughout this paper, the derivative of a Lipschitz function $\frac{d^+}{d\tau}f(\tau)$ is interpreted as

$$\frac{d^+}{d\tau}f(\tau) = \limsup_{h\downarrow 0}\frac{f(\tau+h)-f(\tau)}{h}.$$

We will denote the s, τ derivatives of $\eta(s,\tau)$ by $\eta', \dot\eta$ respectively for brevity.

Lemma 2.1. *Suppose (M, g_τ) satisfies (1.4) and $\eta : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ satisfies $\eta'(s,\tau) \geq 0$. If $c_\tau(x,y) = \eta(d_\tau(x,y),\tau)$, then*

$$\left(-\frac{d^+}{d\tau} - \mathscr{D}_\tau\right)c_\tau(x,y) \geq -\dot\eta + K\eta'd_\tau - \min\{4\eta'',0\}$$

for $(x,y) \in \Omega$.

Proof. Let $\gamma : [-\frac{d}{2}, \frac{d}{2}] \to M$ be any minimizing geodesic connecting x and y parameterized by arc length in (M, g_τ). Let $\{E_i(s)\}_{1\leq i\leq n}$ be a parallel orthonormal basis for $T_{\gamma(s)}M$ with $E_n(s) = \gamma'(s)$. For $i \in \{1, 2, \ldots, n-1\}$, let $\gamma_i : (-\epsilon, \epsilon) \times [-\frac{d}{2}, \frac{d}{2}] \to M$ be the variation $\gamma_i(r,s) = \exp_{\gamma(s)}(rE_i)$, then

$$d_\tau(\exp_x(rE_i), \exp_y(rE_i)) \leq L_\tau[\gamma_i(r,\cdot)],$$

with equality at $r = 0$, where $L_\tau[\gamma]$ means the length of γ with respect to g_τ. Since $\eta(s,\tau)$ is nondecreasing in the first variable, we have

$$c_\tau(\exp_x(rE_i), \exp_y(rE_i)) \leq \eta(L_\tau[\gamma_i(r,\cdot)],\tau),$$

with equality at $r = 0$. By the first variation formula,

$$\left.\frac{d}{dr}\right|_{r=0} L_\tau[\gamma_i(r,\cdot)] = 0.$$

By the second variation formula,

$$\left.\frac{d^2}{dr^2}\right|_{r=0} L_\tau[\gamma_i(r,\cdot)] = \int_{-\frac{d}{2}}^{\frac{d}{2}}\left\{|\nabla_{\gamma'}E_i|^2 - R(\gamma', E_i, E_i, \gamma')\right.$$
$$\left. - \left(\frac{\partial}{\partial s}\langle E_i,\gamma'\rangle\right)^2\right\}ds + \langle\nabla_{E_i}E_i,\gamma'\rangle\Big|_{-\frac{d}{2}}^{\frac{d}{2}}$$
$$= -\int_{-\frac{d}{2}}^{\frac{d}{2}} R(\gamma', E_i, E_i, \gamma')ds.$$

Therefore,

$$(2.2) \quad \sum_{i=1}^{n-1} \tilde{\nabla}_\tau^2 c((E_i, E_i), (E_i, E_i)) \le \sum_{i=1}^{n-1} \frac{d^2}{dr^2}\Big|_{r=0} \eta(L_\tau[\gamma_i(r, \cdot)], \tau)$$

$$= \sum_{i=1}^{n-1} \left[\eta'' \left(\frac{d}{dr}\Big|_{r=0} L_\tau \right)^2 + \eta' \frac{d^2}{dr^2}\Big|_{r=0} L_\tau \right]$$

$$= -\eta' \int_{-\frac{d}{2}}^{\frac{d}{2}} \operatorname{Ric}(\gamma', \gamma') ds.$$

Similarly, if we extend γ a little to get a longer geodesic and we will still denote it by $\gamma : [-\frac{d}{2}, \frac{d}{2} + \delta] \to M$, then

$$d_\tau \left(\gamma \left(-\frac{d}{2} + r \right), \gamma \left(\frac{d}{2} + r \right) \right) \le L_\tau[\gamma|_{(-\frac{d}{2}+r, \frac{d}{2}+r)}] \equiv d,$$

with equality at $r = 0$. Therefore,

$$c_\tau \left(\gamma \left(-\frac{d}{2} + r \right), \ \gamma \left(\frac{d}{2} + r \right) \right) \le \eta(d, \tau).$$

Hence

$$(2.3) \quad \tilde{\nabla}_\tau^2 c((E_n, E_n), (E_n, E_n)) \le 0.$$

It is easy to see that

$$(2.4) \quad \tilde{\nabla}_\tau^2 c((E_n, -E_n), (E_n, -E_n))$$

$$= \frac{d^2}{dr^2}\Big|_{r=0} \eta(d - 2r, \tau) = 4\eta''.$$

Combining (2.1) to (2.4), we get

$$(2.5) \quad \mathscr{D}_\tau c_\tau(x, y) \le -\eta' \int_{-\frac{d}{2}}^{\frac{d}{2}} \operatorname{Ric}(\gamma', \gamma') ds + \min\{4\eta'', 0\}.$$

On the other hand, we get

$$(2.6) \qquad \frac{d^+}{d\tau} d_\tau(x,y) = \limsup_{h \downarrow 0} \frac{d_{\tau+h}(x,y) - d_\tau(x,y)}{h}$$

$$\leq \limsup_{h \downarrow 0} \frac{L_{\tau+h}[\gamma] - L_\tau[\gamma]}{h}$$

$$= \frac{d}{d\tau} L_\tau[\gamma] = \frac{1}{2} \int_{-\frac{d}{2}}^{\frac{d}{2}} \frac{\partial}{\partial \tau} g(\gamma', \gamma') ds.$$

Therefore,

$$(2.7) \qquad -\frac{d^+}{d\tau} c_\tau(x,y) = \eta'\left(-\frac{d^+}{d\tau} d_\tau(x,y)\right) - \dot{\eta}$$

$$\geq \frac{\eta'}{2}\left(\int_{-\frac{d}{2}}^{\frac{d}{2}} -\frac{\partial}{\partial \tau} g(\gamma', \gamma') ds\right) - \dot{\eta}.$$

Combining (2.5) and (2.7), we get

$$\left(-\frac{d^+}{d\tau} - \mathscr{D}_\tau\right) c_\tau(x,y) \geq \frac{\eta'}{2} \int_{-\frac{d}{2}}^{\frac{d}{2}} \left(-\frac{\partial}{\partial \tau} g + 2\operatorname{Ric}\right)(\gamma', \gamma') ds - \dot{\eta} - \min\{4\eta'', 0\}$$

$$\geq K d\eta' - \dot{\eta} - \min\{4\eta'', 0\},$$

where we have used (1.4) in the last inequality. So we have completed the proof. □

3. Proof of Theorem 1.1

The following lemma of Kantorovich duality is well known to the experts.

Lemma 3.1. *(Kantorovich duality, see [16]) Suppose M is a manifold. Let $\mu, \nu \in P(M)$, and $c : M \times M \to \mathbb{R} \bigcup \{+\infty\}$ be a lower semi-continuous cost function. Define*

$$(3.1) \qquad J(\varphi, \psi) = \int_M \varphi d\mu + \int_M \psi d\nu$$

for $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$. Let Φ_c be the set of all $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$ satisfying

$$(3.2) \qquad \phi(x) + \psi(y) \leq c(x,y)$$

for dµ-almost all $x \in M$, dν-almost all $y \in M$. Then

$$(3.3) \qquad T_c(\mu, \nu) = \sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi).$$

A pair of functions $(\varphi, \psi) \in \Phi_c$ is said to be competitive.

Now we begin the proof of Theorem 1.1, using Lemma 3.1 together with the ideas of Section 3 in [2].

Proof of Theorem 1.1. Suppose $b \in (\tau_1, \tau_2)$. For any $\varepsilon > 0$, there exists (α_b, β_b) such that $\alpha_b(x) + \beta_b(y) \le c_b(x, y)$ and $J_b(\alpha_b, \beta_b) > T_{c_b}(\mu, \nu) - \epsilon$. We solve the equations

$$(3.4) \qquad \begin{cases} -\dfrac{\partial \varphi}{\partial \tau} = \Delta_\tau \varphi, & \tau \in (\tau_1, b) \\ \varphi(x, b) = \alpha_b(x), \end{cases}$$

$$(3.5) \qquad \begin{cases} -\dfrac{\partial \psi}{\partial \tau} = \Delta_\tau \psi, & \tau \in (\tau_1, b) \\ \psi(x, b) = \beta_b(x). \end{cases}$$

Since

$$\frac{d}{d\tau} \int_M \varphi u \, dV_\tau = \int_M \left[\frac{\partial \varphi}{\partial \tau} u + \varphi \left(\frac{\partial u}{\partial \tau} \right) + \frac{1}{2} \operatorname{tr} \left(\frac{\partial g}{\partial \tau} \right) \varphi u \right] dV_\tau = 0$$

and $\frac{d}{d\tau} \int_M \psi v \, dV_\tau = 0$, we have

$$\frac{d}{d\tau} J_\tau(\varphi(\cdot, \tau), \psi(\cdot, \tau)) = 0.$$

If we can prove $\varphi(x, \tau) + \psi(y, \tau) \le c_\tau(x, y)$ for every $\tau \in (\tau_1, b)$, then

$$\begin{aligned} T_{c_b}(\mu, \nu) &\le J_b(\alpha_b, \beta_b) + \epsilon \\ &= J_\tau(\varphi_\tau, \psi_\tau) + \epsilon \\ &\le T_{c_\tau}(\mu, \nu) + \epsilon. \end{aligned}$$

By the arbitrariness of ϵ, we have $T_{c_b}(\mu, \nu) \le T_{c_\tau}(\mu, \nu), \forall \tau \in (\tau_1, b)$, i.e., $T_{c_\tau}(\mu, \nu)$ is nonincreasing in τ.

In the following, we will prove $\varphi(x,\tau) + \psi(y,\tau) \le c_\tau(x,y) = \eta(d_\tau(x,y),\tau)$ for $\tau \in (\tau_1, b)$. Define an evolving quantity Z on $M \times M \times (\tau_1, b)$:

$$Z(x,y,\tau) = \varphi(x,\tau) + \psi(y,\tau) - c_\tau(x,y) + \epsilon(\tau - b - 1).$$

Note that since $\eta(s,\tau)$ satisfies (1.5), we have $(-\frac{d^+}{d\tau} - \mathscr{D}_\tau)c_\tau(x,y) \ge 0$ by Lemma 2.1.

We now apply the operator $-\frac{d^+}{d\tau} - \mathscr{D}_\tau$ to Z:

$$\left(-\frac{d^+}{d\tau} - \mathscr{D}_\tau\right)Z(x,y,\tau) = -\frac{\partial}{\partial\tau}\varphi(x,\tau) - \frac{\partial}{\partial\tau}\psi(y,\tau) + \frac{d^+}{d\tau}c_\tau(x,y) - \epsilon$$
$$- \Delta_\tau\varphi(x,\tau) - \Delta_\tau\psi(y,\tau) + \mathscr{D}_\tau c_\tau(x,y)$$
$$= -\epsilon + \left(\frac{d^+}{d\tau} + \mathscr{D}_\tau\right)c_\tau(x,y)$$
$$\le -\epsilon < 0.$$

By assumption, we have $Z(x,y,b) \le -\epsilon < 0$. On the other hand, let $f(x,\tau) = \varphi(x,\tau) + \psi(x,\tau)$, then f satisfies $-\frac{\partial f}{\partial\tau} = \Delta_\tau f$, with $f(x,b) = \varphi(x,b) + \psi(x,b) \le \eta(0,b) = 0$. By maximum principle, we have $f(x,\tau) \le 0 = \eta(0,\tau), \forall\tau \in (\tau_1, b)$. So $Z(x,x,\tau) < 0, \forall\tau \in (\tau_1, b)$. It follows that if Z ever becomes positive, then there exists a maximal $\tau_0 < b$ and $x_0 \ne y_0$ in M such that $Z(x_0, y_0, \tau_0) = 0$. Then at (x_0, y_0, τ_0),

$$\frac{d^+}{d\tau}Z \le 0, \quad \mathscr{D}_\tau Z \le 0.$$

Therefore, $(-\frac{d^+}{d\tau} - \mathscr{D}_\tau)Z(x_0, y_0, \tau_0) \ge 0$. We get a contradiction. Hence $Z(x,y,\tau) \le 0$. By the arbitrariness of ϵ, we get $\varphi(x,\tau) + \psi(y,\tau) \le c_\tau(x,y)$ and finish the proof. $\square$

4. L-optimal transportation

Before starting our new proof of Theorem 1.6, we recall some basic theory of Perelman's $\mathscr{L}$-length. The readers can refer to [5–7, 12] for more details and further results. Suppose we have the backward Ricci flow, $\frac{\partial}{\partial\tau}g_\tau = 2\text{Ric}(g_\tau)$. For a curve $\gamma : [\tau_1, \tau_2] \to M$ with $\tau_2 > \tau_1 > 0$, denote $X(\tau) = \gamma'(\tau)$, and let $Y(\tau)$ be a smooth vector field along $\gamma(\tau)$. The first variation formula of

$\mathscr{L}$-length is:

$$\delta_Y[\mathscr{L}] = 2\sqrt{\tau}\langle X, Y\rangle\big|_{\tau_1}^{\tau_2} + \int_{\tau_1}^{\tau_2} \sqrt{\tau} \left\langle Y, \nabla R - 2\nabla_X X - 4\mathrm{Ric}\,(\cdot, X) - \frac{1}{\tau}X \right\rangle d\tau,$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product with respect to g_τ.

A smooth curve $\gamma(\tau)$ in M is called an $\mathscr{L}$-geodesic if it satisfies the following $\mathscr{L}$-geodesic equation:

$$(4.1) \qquad 2\nabla_X X - \nabla R + 4\mathrm{Ric}\,(\cdot, X) + \frac{1}{\tau}X = 0.$$

Along an $\mathscr{L}$-geodesic γ, we have the second variation formula

$$(4.2) \quad \delta_Y^2[\mathscr{L}] = 2\sqrt{\tau}\langle\nabla_Y Y, X\rangle\big|_{\tau_1}^{\tau_2} + \int_{\tau_1}^{\tau_2} \sqrt{\tau}\big[2|\nabla_X Y|^2 + 2\langle R(Y, X)Y, X\rangle$$
$$+ \mathrm{Hess}R(Y, Y) + 2\nabla_X\mathrm{Ric}\,(Y, Y) - 4\nabla_Y\mathrm{Ric}\,(Y, X)\big]d\tau.$$

Suppose $\gamma : [\tau_1, \tau_2] \to M$ is a shortest $\mathscr{L}$-geodesic connecting (x, τ_1) and (y, τ_2). For $(\hat{x}, \hat{\tau}_1)$, $(\hat{y}, \hat{\tau}_2)$ near (x, τ_1) and (y, τ_2), respectively, denote $\mathscr{L}(\hat{x}, \hat{\tau}_1; \hat{y}, \hat{\tau}_2)$ the $\mathscr{L}$-length of the $\mathscr{L}$-geodesic $\gamma_{(\hat{x}, \hat{\tau}_1; \hat{y}, \hat{\tau}_2)}$ connecting $(\hat{x}, \hat{\tau}_1)$ and $(\hat{y}, \hat{\tau}_2)$ near γ, then $\mathscr{L}(\hat{x}, \hat{\tau}_1; \hat{y}, \hat{\tau}_2) \geq Q(\hat{x}, \hat{\tau}_1; \hat{y}, \hat{\tau}_2)$, with equality when $(\hat{x}, \hat{\tau}_1; \hat{y}, \hat{\tau}_2) = (x, \tau_1; y, \tau_2)$. By the computations similar to Perelman's [12] (see also Lemma A.6 in [15]), we can derive

$$(4.3) \qquad \tau_1\frac{\partial}{\partial\tau_1}\mathscr{L}(x, \tau_1; y, \tau_2) + \tau_2\frac{\partial}{\partial\tau_2}\mathscr{L}(x, \tau_1; y, \tau_2)$$
$$= 2\tau_2^{\frac{3}{2}}R(y, \tau_2) - 2\tau_1^{\frac{3}{2}}R(x, \tau_1) + \mathscr{K} - \frac{1}{2}\mathscr{L}(x, \tau_1; y, \tau_2),$$

where $\mathscr{K} := \int_{\tau_1}^{\tau_2} \tau^{\frac{3}{2}}H(X(\tau))d\tau$, and $H(X)$ is Hamilton's trace Harnack quantity (with $t = -\tau$)

$$H(X) := -\frac{\partial R}{\partial\tau} - \frac{1}{\tau}R - 2\langle\nabla R, X\rangle + 2\mathrm{Ric}(X, X).$$

Now we define an operator $\mathscr{D}$, which is a coupling of $\tau_1\Delta_{\tau_1}|_x$ and $\tau_2\Delta_{\tau_2}|_y$ as in Section 2:

$$\mathscr{D}(f(x, \tau_1; y, \tau_2)) := \inf\big\{\mathrm{tr}(A(\tilde{\nabla}^2 f))\big| A \in \mathrm{Sym}_2(T^*(M \times M)), A \geq 0,$$
$$A_{(x,y)}|_{T_x M \otimes T_x M} = \tau_1 g_{\tau_1}(x), A_{(x,y)}|_{T_y M \otimes T_y M} = \tau_2 g_{\tau_2}(y)\big\}$$

for $f : M \times I \times M \times I \to \mathbb{R}$, where $\tilde{\nabla}^2$ means the Hessian with respect to the product metric of g_{τ_1} and g_{τ_2}.

Suppose $\gamma : [\tau_1, \tau_2] \to M$ is a shortest $\mathscr{L}$-geodesic connecting (x, τ_1) and (y, τ_2). Let $\{Y_i\}_{1 \le i \le n}$ be a basis at $\gamma(\tau_1)$ with $\langle Y_i, Y_j \rangle_{g_{\tau_1}} = \tau_1 \delta_{ij}$. We extend this basis along γ to get a family of bases $\{Y_i(\tau)\}_{1 \le i \le n}$ by solving the ODEs

$$\nabla_X Y_i = -\mathrm{Ric}\,(Y_i, \cdot) + \frac{1}{2\tau} Y_i.$$

From

$$\frac{d}{d\tau}\langle Y_i, Y_j \rangle = 2\mathrm{Ric}(Y_i, Y_j) + \langle \nabla_X Y_i, Y_j \rangle + \langle Y_i, \nabla_X Y_j \rangle = \frac{1}{\tau}\langle Y_i, Y_j \rangle,$$

it follows

$$\langle Y_i(\tau), Y_j(\tau) \rangle_{g_\tau} = \tau \delta_{ij}.$$

By the definition of $\mathscr{D}$, we have

$$\mathscr{D}f \le \sum_{i=1}^{n} \tilde{\nabla}^2 f((Y_i(\tau_1), Y_i(\tau_2)), (Y_i(\tau_1), Y_i(\tau_2))),$$

where the above inequality holds in support sense.

For any $i \in \{1, \ldots, n\}$, let $\gamma_i : (-\epsilon, \epsilon) \times [\tau_1, \tau_2] \to M$, $\gamma_i(r, s) = \exp_{\gamma(s)}(rY_i)$ be a variation of γ, then

$$Q(\exp_x(rY_i), \tau_1; \exp_y(rY_i), \tau_2) \le \mathscr{L}[\gamma_i(r, \cdot)],$$

with equality at $r = 0$. Hence

$$\tilde{\nabla}^2 Q((Y_i(\tau_1), Y_i(\tau_2)), (Y_i(\tau_1), Y_i(\tau_2)))$$
$$\le \left. \frac{d^2}{dr^2} \right|_{r=0} \mathscr{L}[\gamma_i(r, \cdot)]$$
$$= \int_{\tau_1}^{\tau_2} \sqrt{\tau} \big[2|\nabla_X Y_i|^2 + 2\langle R(Y_i, X)Y_i, X \rangle + \mathrm{Hess}R(Y_i, Y_i)$$
$$+ 2\nabla_X \mathrm{Ric}\,(Y_i, Y_i) - 4\nabla_{Y_i}\mathrm{Ric}\,(Y_i, X) \big] d\tau$$
$$= -2\sqrt{\tau}\mathrm{Ric}\,(Y_i, Y_i)\big|_{\tau_1}^{\tau_2} + \sqrt{\tau}\big|_{\tau_1}^{\tau_2} - \int_{\tau_1}^{\tau_2} \sqrt{\tau}\big[-2\langle R(Y_i, X)Y_i, X \rangle$$
$$- \mathrm{Hess}R(Y_i, Y_i) - 4\nabla_X \mathrm{Ric}\,(Y_i, Y_i) + 4\nabla_{Y_i}\mathrm{Ric}\,(X, Y_i) - \frac{1}{\tau}\mathrm{Ric}\,(Y_i, Y_i)$$
$$- 2\frac{\partial}{\partial\tau}\mathrm{Ric}\,(Y_i, Y_i) + 2|\mathrm{Ric}\,(Y_i, \cdot)|^2 \big] d\tau,$$

where in the last equality we have left out the well-known computations due to Perelman [12]. The readers can also consult [5–7] for details of the computations.

Summing over i, we get

$$(4.4) \qquad \mathscr{D}Q \leq n\sqrt{\tau}\big|_{\tau_1}^{\tau_2} - 2\tau^{\frac{3}{2}} R\big|_{\tau_1}^{\tau_2} - \int_{\tau_1}^{\tau_2} \tau^{\frac{3}{2}} H(X) d\tau$$

$$= n(\sqrt{\tau_2} - \sqrt{\tau_1}) - (2\tau_2^{\frac{3}{2}} R(y, \tau_2) - 2\tau_1^{\frac{3}{2}} R(x, \tau_1)) - \mathscr{K}.$$

Now we begin our new proof of Theorem 1.6.

Proof of Theorem 1.6. Suppose the backward Ricci flow is defined on an open interval I containing $[\bar{\tau}_1, \bar{\tau}_2]$, where $0 < \bar{\tau}_1 < \bar{\tau}_2$. Let $\tau_1 = \tau_1(s) := \bar{\tau}_1 e^s$, $\tau_2 = \tau_2(s) := \bar{\tau}_2 e^s$, and

$$P(x, y, s) := 2(\sqrt{\tau_2} - \sqrt{\tau_1}) Q(x, \tau_1; y, \tau_2) - 2n(\sqrt{\tau_2} - \sqrt{\tau_1})^2.$$

If $\gamma(\tau)$ is a shortest $\mathscr{L}$-geodesic connecting (x, τ_1) and (y, τ_2), then by (4.4),

$$(4.5) \qquad \mathscr{D}P(x, y, s) \leq 2(\sqrt{\tau_2} - \sqrt{\tau_1}) \left[n(\sqrt{\tau_2} - \sqrt{\tau_1}) - \left(2\tau_2^{\frac{3}{2}} R(y, \tau_2)\right.\right.$$

$$\left.\left. - 2\tau_1^{\frac{3}{2}} R(x, \tau_1)\right) - \mathscr{K}\right].$$

On the other hand, by (4.3),

$$\frac{d^+}{ds} Q(x, \tau_1; y, \tau_2)$$

$$:= \limsup_{h \downarrow 0} \frac{Q(x, \tau_1(s+h); y, \tau_2(s+h)) - Q(x, \tau_1(s); y, \tau_2(s))}{h}$$

$$\leq \frac{d}{ds} \mathscr{L}(x, \tau_1(s); y, \tau_2(s))$$

$$= \tau_1 \frac{\partial}{\partial \tau_1} \mathscr{L}(x, \tau_1; y, \tau_2) + \tau_2 \frac{\partial}{\partial \tau_2} \mathscr{L}(x, \tau_1; y, \tau_2)$$

$$= 2\tau_2^{\frac{3}{2}} R(y, \tau_2) - 2\tau_1^{\frac{3}{2}} R(x, \tau_1) + \mathscr{K} - \frac{1}{2} Q(x, \tau_1; y, \tau_2).$$

Therefore, we have

$$(4.6) \qquad \frac{d^+}{ds} P(x, y, s) = 2(\sqrt{\tau_2} - \sqrt{\tau_1}) \left(\frac{d^+}{ds} Q(x, \tau_1; y, \tau_2)\right)$$

$$+ (\sqrt{\tau_2} - \sqrt{\tau_1}) Q(x, \tau_1; y, \tau_2) - 2n(\sqrt{\tau_2} - \sqrt{\tau_1})^2$$

$$\leq 2(\sqrt{\tau_2} - \sqrt{\tau_1})[2\tau_2^{\frac{3}{2}} R(y, \tau_2) - 2\tau_1^{\frac{3}{2}} R(x, \tau_1)$$

$$+ \mathscr{K} - n(\sqrt{\tau_2} - \sqrt{\tau_1})].$$

Combining $(4.5), (4.6)$, we get

(4.7) $$\left(-\frac{d^+}{ds} - \mathscr{D}\right) P(x, y, s) \geq 0.$$

It is easy to see

$$\Theta(s) = \inf_{\pi \in \Gamma(\nu_1(\tau_1), \nu_2(\tau_2))} \int_{M \times M} P(x, y, s) d\pi(x, y).$$

Now we can use Lemma 3.1 and the maximum principle to prove the theorem just as what we have done in the proof of Theorem 1.1. The only difference is that the price functions $\varphi(s, x)$ and $\psi(s, y)$ solve the equation

$$-\frac{\partial}{\partial s} f = \tau(s) \Delta_{\tau(s)} f,$$

which is conjugate to the equation

$$\frac{\partial}{\partial s} u = \tau(s)[\Delta_{\tau(s)} u + R(x, \tau(s)) u].$$

Finally, $(\varphi(s, x), \psi(s, y))$ will remain competitive because of (4.7) together with the maximum principle. $\qquad\square$

Acknowledgment

References

[1] B. Andrews and J. Clutterbuck, *Lipschitz bounds for solutions of quasilinear parabolic equations in one space variable*, J. Diff. Equ., **246**(11) (2009), 4268–4283.

[2] B. Andrews and J. Clutterbuck, *Sharp modulus of continuity for parabolic equations on manifolds and lower bounds for the first eigenvalue*, Anal. PDE, **6**(5) (2013), 1013–1024.

[3] M. Arnaudon, K.A. Coulibaly and A. Thalmaier, *Horizontal diffusion in C^1 path space*, Séminaire de Probabilités XLIII, Lecture Notes in Mathematics, 2011, 73–94.

[4] E. Cabezas-Rivas and P.M. Topping, *The canonical shrinking soliton associated to a Ricci flow*, Calc. Var., **43**(1–2) (2012), 173–184.

[5] H.-D. Cao and 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**(2) (2006), 165–492.

[6] B. Chow, S.-C. Chu, D. Glickenstein, C. Guenther, J. Isenberg, T. Ivey, D. Knopf, P. Lu, F. Luo and L. Ni, *The Ricci flow: techniques and applications. Part I: Geometric aspects.*, Math. Surv. Monograph, **135** (2007), xxiv+536 pp.

[7] B. Kleiner and J. Lott, *Notes on Perelman's Papers*, Geom. Topol., **12**(5) (2008), 2587–2855.

[8] K. Kuwada and R. Philipowski, *Coupling of Brownian motions and Perelman's L-functional*, J. Funct. Anal., **260**(9) (2011), 2742–2766.

[9] J. Lott, *Optimal transport and Perelman's reduced volume*, Calc. Var., **36** (2009), 49–84.

[10] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. Math., **169**(3) (2009), 903–991.

[11] R.J. McCann and P.M. Topping, *Ricci flow, entropy and optimal transportation*, Amer. J. Math., **132**(3) (2010), 711–730.

[12] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, http://arXiv.org/abs/math/0211159v1, (2002).

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

[14] K.-T. Sturm, *On the geometry of metric measure spaces I and II*, Acta. Math., **196** (2006), 65–131 and 133–177.

[15] P.M. Topping, *L-optimal transportation for Ricci flow*, J. Reine Angew. Math., **636** (2009), 93–122.

[16] C. Villani, *Topics in optimal transportation*, Grad. Stud. Math., **58** (2003), xvi+370 pp.

School of Mathematics and Computational Science
Sun Yat-sen University
Guangzhou 510275
People's Republic of China
E-mail address: hxiant@mail2.sysu.edu.cn