J. reine angew. Math. **636** (2009), 93—122
DOI 10.1515/CRELLE.2009.083

Journal für die reine und
angewandte Mathematik

$\mathscr{L}$-optimal transportation for Ricci flow

By *Peter Topping* at Coventry

Abstract. We introduce the notion of $\mathscr{L}$-optimal transportation, and use it to construct a natural monotonic quantity for Ricci flow which includes a selection of other monotonicity results, including some key discoveries of Perelman [13] (both related to entropy and to $\mathscr{L}$-length) and a recent result of McCann and the author [11].

1. Introduction

Given a closed manifold $\mathscr{M}$, of dimension n, a smooth family $g(t)$ of Riemannian metrics is called a Ricci flow if it satisifes the nonlinear PDE

$$\frac{\partial g}{\partial t} = -2\,\mathrm{Ric}\big(g(t)\big), \tag{1.1}$$

introduced by Hamilton [8] (see [16] for further information).

In order to do analysis on Ricci flows, one has been traditionally reliant largely on the maximum principle. In particular, one does not have a Sobolev inequality; more precisely one has no *a priori* control on the evolution of the standard Sobolev constant. Instead, one can look for other quantities which *are* controlled under Ricci flow, the best-known of which is the optimal constant in a certain log-Sobolev inequality. That log-Sobolev constant is monotonic in time by virtue of the monotonicity of Perelman's $\mathscr{W}$ entropy (see [13] and [16] for details, and its application to proving "no local collapsing" for Ricci flow).

The goal of this paper is to introduce a new geometric quantity for Ricci flow which is also monotonic, and which simultaneously generalises Perelman's $\mathscr{W}$ entropy and one of Perelman's crucial monotonicity results involving his celebrated notion of $\mathscr{L}$-length. Furthermore, the monotonicity of our new quantity includes a recent result of McCann and the author [11] where Ricci flow was considered in conjunction with the theory of optimal transportation. The new quantity elucidates why these previous entropies and other quantities function the way they do, and indicate the extent to which we can hope to generalise them to other geometric flows.

To describe the new quantity, we introduce a new notion of optimal transportation of measures through space-time in Ricci flow, and an associated notion of Wasserstein-type distance between probability measures. Before we can describe this concept, we must first survey how one can make sense of a distance between two *points* in space-time.

In light of the work of Perelman, it is convenient to consider the Ricci flow backwards in time. To this end, we adopt the notation τ to represent some backwards time parameter (i.e. $\tau = C - t$ for some $C \in \mathbb{R}$) and consider the reverse Ricci flow $\frac{\partial g}{\partial \tau} = 2\,\mathrm{Ric}(g(\tau))$, defined on a time interval including $[\tau_1, \tau_2]$ where $0 \leqq \tau_1 < \tau_2$. Perelman's $\mathscr{L}$-length of a path $\gamma : [\tau_1, \tau_2] \to \mathscr{M}$ (where one should view the point $\gamma(\tau)$ as a point in the Riemannian manifold $(\mathscr{M}, g(\tau))$) is defined [13] by

$$\mathscr{L}(\gamma) := \int_{\tau_1}^{\tau_2} \sqrt{\tau}\left(R(\gamma(\tau), \tau) + |\gamma'(\tau)|^2_{g(\tau)}\right) d\tau, \tag{1.2}$$

where $R(x, \tau)$ is the scalar curvature at x in $(\mathscr{M}, g(\tau))$. One can use such a length to give rise to a distance, mirroring the classical construction of Riemannian geometry: We define the $\mathscr{L}$-distance between a point (x, τ_1) and (y, τ_2) (where $x, y \in \mathscr{M}$ and $0 \leqq \tau_1 < \tau_2$ are times) as

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

with the caveat that this distance can be negative, and one is not directly generating a metric space via this construction. When τ_1 and τ_2 are pushed together, the scalar curvature term in the definition (1.2) of $\mathscr{L}$ is dwarfed by the 'energy' term, and one recovers the classical Riemannian distance in the sense that

$$\lim_{\tau_2 \downarrow \tau_1} 2(\sqrt{\tau_2} - \sqrt{\tau_1}) Q(x, \tau_1; y, \tau_2) = d^2(x, y, \tau_1), \tag{1.3}$$

uniformly in x and y, where $d(\cdot\,, \cdot, \tau)$ is the Riemannian distance with respect to $g(\tau)$.

Equipped with Q, we can introduce the $\mathscr{L}$-Wasserstein "distance" $V(v_1, \tau_1; v_2, \tau_2)$ between two Borel probability measures v_1 and v_2, viewed at times τ_1 and τ_2 respectively:

$$V(v_1, \tau_1; v_2, \tau_2) := \inf_{\pi \in \Gamma(v_1, v_2)} \int_{\mathscr{M} \times \mathscr{M}} Q(x, \tau_1; y, \tau_2)\, d\pi(x, y) \tag{1.4}$$

where $\Gamma(v_1, v_2)$ is the space of Borel probability measures on $\mathscr{M} \times \mathscr{M}$ with marginals v_1 and v_2 (i.e. $\pi(\Omega \times \mathscr{M}) = v_1(\Omega)$ and $\pi(\mathscr{M} \times \Omega) = v_2(\Omega)$ for Borel $\Omega \subset \mathscr{M}$). By virtue of (1.3), we can recover the standard 2-Wasserstein distance W_2 from V in the limit that $\tau_2 \downarrow \tau_1$:

$$\begin{aligned} \lim_{\tau_2 \downarrow \tau_1} 2(\sqrt{\tau_2} - \sqrt{\tau_1}) V(v_1, \tau_1; v_2, \tau_2) &= W_2^2(v_1, v_2, \tau_1) \\ &:= \inf_{\pi \in \Gamma(v_1, v_2)} \int_{\mathscr{M} \times \mathscr{M}} d^2(x, y, \tau_1)\, d\pi(x, y). \end{aligned} \tag{1.5}$$

Whilst the distance V will be the main ingredient of our new result, all of the results we will discuss in this paper are phrased (or can be rephrased) in terms of the probability densities of Brownian diffusion on Ricci flows, backwards in time (that is, forwards in τ). In other words, we consider families $v(\tau)$ of Borel probability measures so that if $\tau_a < \tau_b$ and $v(\tau_a)$ represents the probability of the location of a Brownian particle at time τ_a, then $v(\tau_b)$ represents the probability of the location of the particle at time τ_b. Mathematically, if we denote the Riemannian volume measure on $(\mathscr{M}, g(\tau))$ by $\mu(\tau)$, and write $d\nu(\tau) = u(\tau)\, d\mu(\tau)$ for some evolving probability density $u : \mathscr{M} \times (\tau_a, \tau_b) \to (0, \infty)$ then u satisfies the equation

$$\frac{\partial u}{\partial \tau} = \Delta u - Ru, \tag{1.6}$$

where the scalar curvature term is arising because of the evolution of the volume element $\frac{\partial}{\partial \tau} d\mu(\tau) = \left(\frac{1}{2} \operatorname{tr} \frac{\partial g}{\partial \tau}\right) d\mu(\tau) = R\, d\mu(\tau)$—see [16], (2.5.7). By considering families $\nu(\tau)$ over open intervals, we may always assume that $\nu(\cdot)$ is a smooth family of positive measures, by which we mean that its density u is smooth and strictly positive. For brevity, throughout the paper we will refer to such families $\nu(\tau)$ satisfying (1.6) simply as *diffusions*. It is a general principle which can be extracted from Perelman's work [13] that the properties of such diffusions are related to the properties of the Ricci flow itself. This mirrors the classical connection between the geometry of fixed Riemannian manifolds and the properties of the heat kernels they support.

Our main theorem asserts the monotonicity of a renormalised version of the $\mathscr{L}$-Wasserstein distance between two diffusions, at different times. The quantity has a global space-time aspect, but localising or restricting it will reveal some more familiar monotonic quantities.

Theorem 1.1. *Suppose that $0 < \overline{\tau_1} < \overline{\tau_2}$ and $g(\tau)$ is a (reverse) Ricci flow on a closed manifold $\mathscr{M}$ of dimension n, for τ in some open interval containing $[\overline{\tau_1}, \overline{\tau_2}]$. Suppose that $\nu_1(\tau)$ and $\nu_2(\tau)$ are two diffusions (as defined above) for τ in some neighbourhoods of $\overline{\tau_1}$ and $\overline{\tau_2}$ respectively. 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 renormalised distance between the diffusions ν_1 and ν_2 at s by*

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

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

Then $\Theta(s)$ is a (weakly) decreasing function of s.

The fact that we should track the diffusions ν_1 and ν_2 with this exponential parametrisation is somewhat unconventional but is natural when one considers the invariance of Ricci flow under parabolic rescaling [16], §1.2.3.

We will prove Theorem 1.1 in Section 4. Before doing so, we will have to develop the theory of $\mathscr{L}$-optimal transportation (Section 2) in order to understand the structure of the minimiser $\pi \in \Gamma(\nu_1, \nu_2)$ which will exist for the variational problem in (1.4). This will lead us to a construction of what we will call $\mathscr{L}$-Wasserstein geodesics between two given probability measures. In Section 3 we will investigate the properties of the classical Boltzmann-Shannon entropy along these $\mathscr{L}$-Wasserstein geodesics. This will involve investigating carefully the behaviour of $\mathscr{L}$-geodesics for Ricci flow, and their $\mathscr{L}$-Jacobi fields, and making natural computations for the second derivatives of the volume element along $\mathscr{L}$-geodesics which extend the first derivative calculations which were used so successfully by Perelman [13]. Luckily, many of the optimal transportation aspects of this theory can be developed along similar lines to the development of the original rigorous theory of optimal transportation on Riemannian manifolds. In particular, we follow the work of McCann [10] and Cordero-Erausquin, McCann and Schmuckenschläger [6] wherever possible. Heuristics which motivated some of that original theory can be found in work of Otto and Villani [12]. Cedric Villani has pointed out to us that an alternative to developing the optimal transport structure theory following [10] and [6] would be to invoke the theory applicable to very general cost functions which is developed in his forthcoming lecture notes [18]. Optimal transport in this generality is also considered in [3] as pointed out to us by Robert McCann. The proof of our main result itself is closest in spirit and detail to our previous work [11] with McCann.

Before proceeding with this detail, we explain how the results of Perelman, and McCann and the author, fall naturally out of Theorem 1.1, as alluded to earlier. We are only looking for quantities which are adapted to studying shrinking solitons in Ricci flow; modifications to the theory could be made to recover corresponding quantities adapted to steady or expanding solitons if they were required.

1.1. Recovering the result of McCann-Topping. We have seen in (1.5) how the standard Wasserstein distance W_2 arises in a limit of our $\mathscr{L}$-Wasserstein distance V. In Lemma B.1 and Corollary B.3 of Appendix B, we will sharpen this relationship. Turning to Theorem 1.1, if we take $\overline{\tau_2} \downarrow \overline{\tau_1}$, then for each s, $\Theta(s) \to W_2^2\big(v_1(\tau_1), v_2(\tau_1), \tau_1\big)$, and we find:

Corollary 1.2 (McCann-Topping [11]). *Given two diffusions $v_1(\tau)$ and $v_2(\tau)$ (as defined earlier) on a reverse Ricci flow $g(\tau)$, the function*

$$\tau \to W_2\big(v_1(\tau), v_2(\tau), \tau\big)$$

is (weakly) decreasing in τ.

This result leads in [11] to a characterisation of supersolutions to the Ricci flow equation, which can be exploited to give a notion of weak solutions for Ricci flow.

1.2. Recovering Perelman's $\mathscr{W}$-entropy. Perelman's celebrated $\mathscr{W}$-entropy is used to prove "no local collapsing" for Ricci flow, and it lies behind Perelman's pseudolocality result [13]. To recover it, we need also to consider the limit as τ_1 and τ_2 approach each other. However now, we consider the case that $v_1(\tau)$ and $v_2(\tau)$ coincide. By the previous case, our renormalised distance $\Theta(s)$ will be zero in the limit $\overline{\tau_2} \downarrow \overline{\tau_1}$, so in this case, we will look at the next term in the expansion of $\Theta(s)$ in terms of $(\overline{\tau_2} - \overline{\tau_1})$ to get a new monotonic quantity.

We will need to consider the infinitesimal version of the $\mathscr{L}$-Wasserstein distance implied in the following lemma. Given a smooth family of positive probability measures $v(\tau)$ on a closed manifold $\mathscr{M}$, for τ in some neighbourhood of τ_1, we call a vector field $X \in \Gamma(T\mathscr{M})$ an *advection* field for $v(\tau)$ at $\tau = \tau_1$ if there exists a smooth family of diffeomorphisms $\psi_\tau : \mathscr{M} \to \mathscr{M}$, for τ in a neighbourhood of τ_1, with ψ_{τ_1} the identity, and such that $(\psi_\tau)_\# v(\tau_1) = v(\tau)$ and $X = \left.\dfrac{\partial \psi}{\partial \tau}\right|_{\tau=\tau_1}$.

Lemma 1.3. *Suppose $g(\tau)$ is a (reverse) Ricci flow, and $\nu(\tau)$ is a smooth family of positive probability measures on a closed manifold $\mathcal{M}$, for τ in some neighbourhood of $\tau_1 \in \mathbb{R}$. Then as $\tau_2 \downarrow \tau_1$,*

$$V\big(\nu(\tau_1), \tau_1; \nu(\tau_2), \tau_2\big) = (\tau_2 - \tau_1)\left[\inf_X \sqrt{\tau_1} \int_{\mathcal{M}} \big(R(\cdot, \tau_1) + |X|^2_{g(\tau_1)}\big)\, d\nu(\tau_1)\right] + o(\tau_2 - \tau_1), \tag{1.7}$$

where the infimum is taken over all advection fields X for $\nu(\tau)$ at $\tau = \tau_1$.

In this lemma, we are choosing the advection field X above to have the least 'kinetic energy'; the minimising X can be written explicitly as the gradient of $v : \mathcal{M} \to \mathbb{R}$ solving $-\mathrm{div}(U\nabla v) = \dfrac{dU}{d\tau}$ at $\tau = \tau_1$, where $U(\tau)$ is the one-parameter family of probability densities satisfying $d\nu(\tau) = U(\tau)\, d\mu(\tau_1)$ with $\mu(\tau_1)$ representing the Riemannian volume measure for $g(\tau_1)$. It is the coefficient of $(\tau_2 - \tau_1)$ in (1.7) (the part within square brackets) which we call the infinitesimal $\mathcal{L}$-Wasserstein distance, or $\mathcal{L}$-Wasserstein *speed* of $\nu(\tau)$, with respect to $g(\tau)$, at $\tau = \tau_1$. We delay the proof of Lemma 1.3 until Appendix B.

Let us apply this lemma in the case of Theorem 1.1 specialised to the situation that $\nu_1(\tau) = \nu_2(\tau)$; we will write this measure simply as $\nu(\tau)$. We denote its probability density with respect to Riemannian volume measure $\mu(\tau)$ by $u(\tau) := \dfrac{d\nu(\tau)}{d\mu(\tau)}$ and its probability density with respect to $\mu(\tau_1)$ by $U(\tau) := \dfrac{d\nu(\tau)}{d\mu(\tau_1)}$ as before. Then $\dfrac{\partial U}{\partial \tau} = \Delta U = \Delta u$ at $\tau = \tau_1$, so the optimal advection field is given by $X = -\nabla \ln u$. Let us write $\overline{\tau_2} = (1+\eta)\overline{\tau_1}$, so the functions $\tau_1(s)$ and $\tau_2(s)$ of the theorem satisfy $\tau_2(s) = (1+\eta)\tau_1(s)$ for all s. In this situation, Lemma 1.3 tells us that

$$\Theta(s) = \eta^2\left(\tau_1^2 \int_{\mathcal{M}} (R + |\nabla \ln u|^2)\, d\nu(\tau_1) - \frac{n\tau_1}{2}\right) + o(\eta^2) = \eta^2\left(\tau_1^2 \mathcal{F}(\tau_1) - \frac{n\tau_1}{2}\right) + o(\eta^2),$$

where for each τ,

$$\mathcal{F} = \int_{\mathcal{M}} (R + |\nabla \ln u|^2) u\, d\mu$$

is Perelman's $\mathcal{F}$-information functional [13], [16], §6.2. Theorem 1.1 then tells us that

$$\left(\tau^2 \mathcal{F}(\tau) - \frac{n\tau}{2}\right) \text{ is weakly decreasing in } \tau. \tag{1.8}$$

We are interested in understanding Perelman's $\mathcal{W}$-entropy which is normally written (for given τ) as

$$\mathcal{W} = \int_{\mathcal{M}} [\tau(|\nabla f|^2 + R) + f - n] u\, d\mu,$$

where $f : \mathscr{M} \to \mathbb{R}$ is defined by $u = \frac{e^{-f}}{(4\pi\tau)^{\frac{n}{2}}}$. (See [13] and [16] for more information and applications to proving "no local collapsing".) Now a short calculation shows that

$$\frac{d\mathscr{W}}{d\tau} = \frac{1}{\tau}\frac{d}{d\tau}\left(\tau^2\mathscr{F}(\tau) - \frac{n\tau}{2}\right)$$

so by (1.8) we recover the monotonicity of $\mathscr{W}$:

$$\frac{d\mathscr{W}}{d\tau} \leqq 0.$$

1.3. Recovering Perelman's enlarged length monotonicity. Whereas we have considered distances $Q(x, \tau_1; y, \tau_2)$ so far, most of Perelman's constructions involve the special case $L(y, \tau) := Q(x, 0; y, \tau)$ for fixed $x \in \mathscr{M}$, or variants thereof. In particular, he defines the *enlarged distance* $\overline{L}(y, \tau) := 2\sqrt{\tau}L(y, \tau)$, and proves that the minimum over $\mathscr{M}$ of $\overline{L}(\cdot, \tau) - 2n\tau$ is a weakly *decreasing* function of τ. Because the minimum is zero in the limit $\tau \downarrow 0$, this implies that for any τ, one can always find a point $y \in \mathscr{M}$ for which $\overline{L}(y, \tau) \leqq 2n\tau$, and that fact turns out to be essential in Perelman's arguments to extract asymptotic solitons for κ-solutions, and also to prove "no local collapsing" estimates when one is studying Ricci flows with surgery. (See [13] and [14] for more details.)

Here we point out that the above monotonicity is also encoded in our Theorem 1.1. To see this, we would like to set $\overline{\tau_1} = 0$. (Strictly speaking, we have assumed that $\tau_1 > 0$ to avoid dealing with a host of special cases and technical issues in the proofs; we leave the reader either to extend the theory, or take a limit $\tau_1 \downarrow 0$.) The exponential function $\tau_1(s)$ will then be zero for all s. For $v_1(\tau)$, we take the diffusion which at $\tau = 0$ is the point unit mass δ_x centred at x. Therefore $v_1\big(\tau_1(s)\big)$ is that same measure for all s, and because the minimising π in the transportation problem defining $V\big(\delta_x, 0, v_2(\tau_2), \tau_2\big)$ will be $\delta_x \times v_2(\tau_2)$, we have $V\big(\delta_x, 0, v_2(\tau_2), \tau_2\big) = \int_{\mathscr{M}} L(\cdot, \tau_2)\, dv(\tau_2)$, and hence

$$\Theta(s) = \int_{\mathscr{M}} \big(\overline{L}(\cdot, \tau_2) - 2n\tau_2\big)\, dv_2(\tau_2).$$

Theorem 1.1 then shows that the function

$$\tau \to \int_{\mathscr{M}} \big(\overline{L}(\cdot, \tau) - 2n\tau\big)\, dv_2(\tau)$$

is weakly decreasing, which because $v_2(\tau)$ is an arbitrary diffusion, tells us that the minimum of the integrand is also (weakly) decreasing.

1.4. Fixed manifolds. A further precursor to Theorem 1.1 is the work of Sturm and von Renesse [15]. They showed that on a *fixed* Riemannian manifold of (weakly) positive Ricci curvature, the Wasserstein distance between two diffusions is decreasing. Our results intersect in the special case that one considers a Ricci flat Riemannian manifold.

Acknowledgements. I would like to thank Robert McCann for many useful conversations about optimal transportation. Following our previous work [11] with McCann, I would like to thank John Lott and Ben Chow for encouraging the search for links between the results in [11] and the entropy monotonicity and $\mathscr{L}$-length theory in [13], respectively. This work was partly supported by The Leverhulme Trust.

2. Overview of $\mathscr{L}$-optimal transportation

Throughout this section, we will be considering a smooth (reverse) Ricci flow $g(\tau)$ defined on an open time interval including some interval $[\tau_1, \tau_2]$ with $0 < \tau_1 < \tau_2$. Our goal is to understand the variational problem from (1.4). To begin, we note that $\Gamma(\nu_1, \nu_2)$ from (1.4) is a weak-$*$ compact subset of the dual to the Banach space of continuous functions on $\mathcal{M} \times \mathcal{M}$ equipped with the C^0 norm, and so we can be sure of the existence of a minimiser $\pi \in \Gamma(\nu_1, \nu_1)$ for the variational problem in (1.4) by the Banach-Alaoglu theorem. We call this minimising π the optimal transference plan, reusing the standard terminology from standard mass transportation theory.

However, in order to rigorously prove anything about $\mathscr{L}$-optimal transportation, we must understand the *structure* of the minimising π in some detail, and that is what we address now.

In order to discuss these issues, we need some basic theory of Perelman's $\mathscr{L}$-length. The more elaborate theory we require along these lines will be relegated to Appendix A. We have already introduced Perelman's notion of $\mathscr{L}$-distance; he also introduced [13] a notion of $\mathscr{L}$-geodesic $\gamma : [\tau_1, \tau_2] \to \mathcal{M}$ analogous to the usual Riemannian notion, which satisfies the equation $D_\tau X = \frac{1}{2}\nabla R - 2\,\mathrm{Rc}(X) - \frac{1}{2\tau}X$, where $X = \gamma'(\tau)$, Rc is the Ricci curvature viewed as an endomorphism, and D_τ represents the pull-back under γ of the Levi-Civita connection on $(\mathcal{M}, g(\tau))$, acting in the direction $\frac{\partial}{\partial \tau}$. This notion of geodesic then gives rise to an $\mathscr{L}$-exponential map $\mathscr{L}_{\tau_1,\tau_2}\exp_x : T_x\mathcal{M} \to \mathcal{M}$ which maps a vector $Z \in T_x\mathcal{M}$ to the point $\gamma(\tau_2) \in \mathcal{M}$, where $\gamma : [\tau_1, \tau_2] \to \mathcal{M}$ is the unique $\mathscr{L}$-geodesic such that $\gamma(\tau_1) = x$ and $\sqrt{\tau_1}\gamma'(\tau_1) = Z$.

Consider, for the moment, the optimal transference plan π in the case that the measures ν_1 and ν_2 in (1.4) are absolutely continuous with respect to volume measure. (Consider volume measure to be Riemannian volume measure here and in the sequel; the notion of absolute continuity is independent of the smooth Riemannian metric one chooses.) We'll show that the π arises as the push-forward of ν_1 under a map $\mathcal{M} \to \mathcal{M} \times \mathcal{M}$ defined by $x \to (x, F(x))$ where $F : \mathcal{M} \to \mathcal{M}$ is a Borel map defined in terms of a potential function $\varphi : \mathcal{M} \to \mathbb{R}$ and the $\mathscr{L}$-exponential map (see Remark 2.8). The potential φ will arise via a 'Kantorovich' dual formulation of the variational problem. Using this structure, we will be able to control π effectively. For example, π will be seen to give zero measure to the $\mathscr{L}$-cut locus $\mathscr{L}Cut_{\tau_1,\tau_2}$ which could be defined as the smallest subset of $\mathcal{M} \times \mathcal{M}$ off which $Q(\cdot, \tau_1; \cdot, \tau_2)$ is smooth.

Developing this structure theory yields a Jacobian change of variables formula (via Theorem 2.14) which will allow us later to effectively compute entropies of measures along $\mathscr{L}$-Wasserstein geodesics, which are certain optimal paths of Borel probability measures on $\mathcal{M}$ defined in terms of the $\mathscr{L}$-exponential map and the potential φ mentioned above.

Ultimately, the entropy calculations will be phrased in terms of $\mathscr{L}$-Jacobi fields, which are analogues of Riemannian Jacobi fields in this setting. The necessary $\mathscr{L}$-Jacobi field computations will be made in the next section.

Virtually all of the material in this section is in one to one correspondence with the development of the standard theory of optimal transportation on manifolds by McCann [10] and Cordero-Erausquin, McCann and Schmuckenschläger [6]. (As mentioned earlier, one could also appeal to [18].) We follow their route as closely as possible, and only give brief sketches of proofs where little adaptation is necessary. Our main goal here is to point out the exact analogues of their results in our setting. The first deviation of presentation—of the Legendre-Fenchel-type transform used in the classical theory—is motivated by the asymmetry of our cost function.

Given a continuous function $\varphi : \mathscr{M} \to \mathbb{R}$, we define the function $\hat{\varphi} : \mathscr{M} \to \mathbb{R}$ by

$$\hat{\varphi}(y) = \inf_{x \in \mathscr{M}} [Q(x, \tau_1; y, \tau_2) - \varphi(x)]. \tag{2.1}$$

Likewise, if $\psi : \mathscr{M} \to \mathbb{R}$ is continuous, then $\check{\psi} : \mathscr{M} \to \mathbb{R}$ is defined by

$$\check{\psi}(x) = \inf_{y \in \mathscr{M}} [Q(x, \tau_1; y, \tau_2) - \psi(y)]. \tag{2.2}$$

These transforms depend on τ_1 and τ_2, but those parameters can be viewed as fixed for now. Indeed, let us abbreviate $Q(x, y) := Q(x, \tau_1; y, \tau_2)$ where no confusion will arise. As mentioned above, these transforms are the analogues of the c-transform in classical mass transportation (see [10], for example) with slightly different notation to emphasise the asymmetry of $Q(\cdot\,,\cdot)$. It is straightforward to check that taking one transform and then the other can only increase the original function:

$$\check{\hat{\varphi}} \geqq \varphi; \quad \hat{\check{\psi}} \geqq \psi. \tag{2.3}$$

We have equality in, say, the first of these inequalities if $\varphi = \check{\psi}$ for some ψ, because then

$$\check{\hat{\varphi}} = \check{\hat{\check{\psi}}} \leqq \check{\psi} = \varphi.$$

Definition 2.1. Given a Ricci flow $g(\tau)$, we call a function $\varphi : \mathscr{M} \to \mathbb{R}$ *reflexive* (with respect to the interval $[\tau_1, \tau_2]$) if it is continuous, and satisfies $\check{\hat{\varphi}} = \varphi$.

This concept is called c-concavity in the classical theory of optimal transportation. Taking these transforms improves regularity in the following sense. (The proof can be adapted from [10].)

Lemma 2.2 (cf. [10], Lemma 2). *Suppose that there exists $K < \infty$ such that for all $x \in \mathscr{M}$, the Lipschitz constant of $Q(x, \cdot)$ is no more than K. Then for all continuous $\varphi : \mathscr{M} \to \mathbb{R}$, the function $\hat{\varphi}$ is also Lipschitz with Lipschitz constant no more than K. Here, Lipschitz is with respect to $g(\tau_2)$.*

Similarly, Lipschitz control on $Q(\cdot, y)$ gives Lipschitz control on $\check{\psi}$. (Generally, we will not state similar results obtained by switching x and y.) As we recall in Appendix A, *Q is* Lipschitz in both its variables. Throughout this section, we will be implicitly using the consequence of this lemma, via Rademacher's theorem, that any reflexive φ is differentiable almost-everywhere.

We wish to work towards a *Kantorovich* dual formulation of the $\mathscr{L}$-optimal transportation problem. Define

$$\Sigma = \{(\varphi, \psi) \mid \varphi, \psi : \mathscr{M} \to \mathbb{R} \text{ continuous and } \varphi(x) + \psi(y) \leqq Q(x, y)\ \forall x, y \in \mathscr{M}\},$$

and, given Borel probability measures ν_1 and ν_2 on $\mathscr{M}$, define $J : \Sigma \to \mathbb{R}$ by

$$J(\varphi, \psi) = \int_{\mathscr{M}} \varphi \, d\nu_1 + \int_{\mathscr{M}} \psi \, d\nu_2.$$

Lemma 2.3 (cf. [10], Proposition 3). *There exists a reflexive φ such that the supremum of J over Σ is attained at $(\varphi, \hat{\varphi})$.*

The proof (following [10]) is based on showing that if $(\varphi, \psi) \in \Sigma$, then $(\check{\hat{\varphi}}, \hat{\varphi}) \in \Sigma$ and $J(\varphi, \psi) \leqq J(\check{\hat{\varphi}}, \hat{\varphi})$. By virtue of Lemma 2.2, this allows one to alter any maximising sequence (φ_i, ψ_i) to one with controlled Lipschitz continuity, which enables us to pass to a limit via the Ascoli-Arzelà theorem to get a maximum.

By definition of the transform (2.1), we have $\varphi(x) + \hat{\varphi}(y) \leqq Q(x, y)$ for all $x, y \in \mathscr{M}$. The case of equality is special:

Lemma 2.4 (cf. [10], Lemma 7). *Suppose that φ is reflexive, and is differentiable at $x \in \mathscr{M}$. Then $\varphi(x) + \hat{\varphi}(y) = Q(x, y)$ if and only if*

$$y = \mathscr{L}_{\tau_1, \tau_2} \exp_x\left(-\frac{\nabla\varphi(x)}{2}\right). \tag{2.4}$$

In this case, $Q(\cdot, y)$ is differentiable at x, and $\nabla\varphi(x) = \nabla\big(Q(\cdot, y)\big)(x)$.

The gradient here is with respect to $g(\tau_1)$. There is an analogous result in the case of differentiability of $\hat{\varphi}$ at y.

Remark 2.5. Whenever we have $x, y \in \mathscr{M}$ such that $\varphi(x) + \hat{\varphi}(y) = Q(x, y)$, the function φ must be a support function (or 'lower barrier') for $Q(\cdot, y) - \hat{\varphi}(y)$ near x. That is, the former function lies below the latter near x, with equality *at* x. This will repeatedly allow us to relate differentiability and convexity properties of φ and $Q(\cdot, y)$ at such points x.

Concerning the proof of the lemma (analogous to that in [10]), for the *only if* part, note that by Remark 2.5, and the differentiability of φ at x, the function $Q(\cdot, y)$ admits $\nabla\varphi(x)$ as a subgradient at x. Moreover, by Lemma A.3 in Appendix A, $-2Z$ is a supergradient of it at x, where $Z \in \Omega(x, \tau_1; \tau_2) \subset T_x\mathscr{M}$ satisfies $y = \mathscr{L}_{\tau_1, \tau_2} \exp_x(Z)$. (See Appendix A for notation.) The supergradient and subgradient must then coincide as a genuine gradient, $-2Z = \nabla\varphi(x)$. This is enough to establish (2.4). The *if* part is easier; by definition of $\hat{\varphi}$, there always exists at least one point $z \in \mathscr{M}$ at which $\varphi(x) + \hat{\varphi}(z) = Q(x, z)$, and by what we have seen, this z must coincide with any y satisfying (2.4).

These considerations put us in a position to construct maps F which transport certain measures in an optimal way.

Theorem 2.6 (cf. [10], Theorem 8). *Suppose σ is a Borel probability measure which is absolutely continuous with respect to (any) volume measure on $\mathscr{M}$. Suppose that φ is a reflexive function. Then $F : \mathscr{M} \to \mathscr{M}$, a Borel map defined at points of differentiability of φ by*

$$F(x) := \mathscr{L}_{\tau_1,\tau_2} \exp_x\left(-\frac{\nabla\varphi(x)}{2}\right), \tag{2.5}$$

minimises the functional

$$\int_{\mathscr{M}} Q\big(x, F(x)\big)\, d\sigma(x)$$

amongst all Borel maps $\tilde{F}$ such that $\tilde{F}_\#\sigma = F_\#\sigma$. Any other minimiser must agree with F σ-a.e.

The proof follows exactly as in [10]: For any $\tilde{F}$ as in the theorem, and any $(u, v) \in \Sigma$, then with J defined with respect to $v_1 = \sigma$ and $v_2 = F_\#\sigma$, we have

$$\begin{aligned} J(u,v) &= \int_{\mathscr{M}} u\, d\sigma + \int_{\mathscr{M}} v\, d(F_\#\sigma) \\ &= \int_{\mathscr{M}} u\, d\sigma + \int_{\mathscr{M}} v\big(\tilde{F}(x)\big)\, d\sigma(x) \leqq \int_{\mathscr{M}} Q\big(x, \tilde{F}(x)\big)\, d\sigma(x). \end{aligned} \tag{2.6}$$

But by Lemma 2.4 and the almost-everywhere differentiability of φ, we have $\varphi(x) + \hat{\varphi}\big(F(x)\big) = Q\big(x, F(x)\big)$ for almost all x (with respect to any volume measure) and hence

$$J(\varphi, \hat{\varphi}) = \int_{\mathscr{M}} Q\big(x, F(x)\big)\, d\sigma(x).$$

Combining with (2.6), we find that

$$J(\varphi, \hat{\varphi}) = \sup_{\Sigma} J = \inf_{\tilde{F}} \int_{\mathscr{M}} Q\big(x, \tilde{F}(x)\big)\, d\sigma(x) = \int_{\mathscr{M}} Q\big(x, F(x)\big)\, d\sigma(x),$$

and in particular, that F is the sought minimiser. If $\tilde{F}$ is any other minimiser, we must still have $\varphi(x) + \hat{\varphi}\big(\tilde{F}(x)\big) = Q\big(x, \tilde{F}(x)\big)$ for σ-almost all x, and by Lemma 2.4, we then know that $\tilde{F}(x) = F(x)$ for σ-almost all x.

Given the previous theorem, one would like to be able to find a reflexive φ (and hence F) to make the measure $F_\#\sigma$ coincide with a measure of our choice:

Theorem 2.7 (cf. [10], Theorem 9). *Suppose that v_1 and v_2 are Borel probability measures, with v_1 absolutely continuous with respect to (any) volume measure on $\mathscr{M}$. Then there exists a reflexive function $\varphi : \mathscr{M} \to \mathbb{R}$ such that Borel $F : \mathscr{M} \to \mathscr{M}$ defined at points of differentiability of φ by* (2.5) *satisfies $F_\# v_1 = v_2$.*

The proof mimics that of [10], Theorem 9. The function φ is that given by Lemma 2.3.

Remark 2.8. Theorem 2.7 can be extended using Theorem 2.6 to assert that the optimal transference plan π in the definition (1.4) of $V(v_1, \tau_1; v_2, \tau_2)$ is given by the push-forward of v_1 under the map $x \to (x, F(x))$.

Returning to Lemma 2.4 and the definition of F from Theorem 2.6, we see that the image of F at a point x of differentiability for φ, is the unique point y at which $\varphi(x) + \hat{\varphi}(y) = Q(x, y)$. Following [6], we now view F as the multi-valued function which assigns to an arbitrary point x the *set* of points y at which $\varphi(x) + \hat{\varphi}(y) = Q(x, y)$. (We will tend to abuse notation by occasionally retaining the old viewpoint for F at points of differentiability of φ.)

The following is merely a fragment of the proof of Lemma 2.4, but is included as the analogue of [6], Lemma 3.7, and is needed to prove Lemma 2.13 below. Again, the terminology $\Omega(x, \tau_1; \tau_2)$ comes from Appendix A.

Lemma 2.9. *If φ is reflexive, $y \in F(x)$ and we pick $Z \in \Omega(x, \tau_1; \tau_2) \subset T_x\mathscr{M}$ such that $y = \mathscr{L}_{\tau_1,\tau_2} \exp_x(Z)$, then $-2Z$ is a supergradient of φ at x.*

We now turn to study second derivatives of Q and potentials φ. We are particularly interested in semiconcavity properties. (If necessary, see Appendix A for the definition of semiconcave.) Given any reflexive function φ, we can pick arbitrary points $x \in \mathscr{M}$ and $y \in F(x)$ and consider φ as a support function for $Q(\cdot, y) - \hat{\varphi}(y)$ at x as in Remark 2.5. This implies that for $u \in T_x\mathscr{M}$ sufficiently small,

$$\begin{aligned} (2.7) \qquad & \frac{\varphi(\exp_x u) + \varphi(\exp_x(-u)) - 2\varphi(x)}{|u|^2} \\ & \leqq \frac{Q(\exp_x u, y) + Q(\exp_x(-u), y) - 2Q(x, y)}{|u|^2}, \end{aligned}$$

where we are using the exponential map with respect to $g(\tau_1)$. We then see that φ inherits the uniform semiconcavity of Q from Lemma A.4 in Appendix A, and we may deduce semiconcavity of φ from [6], Lemma 3.11:

Lemma 2.10. *A reflexive function is semiconcave.*

We have already seen that a reflexive function φ is differentiable almost everywhere, because it is Lipschitz. By virtue of the semiconcavity of φ, we can be sure also that a Hessian in the sense of Alexandrov exists almost everywhere (see [6], [1]). The following lemmata obtain refined control at points where this Hessian exists. (See Appendix A for a discussion of the $\mathscr{L}$-cut locus $\mathscr{L}Cut$ and its subset $\mathscr{L}Cut_{\tau_1,\tau_2}$.)

Lemma 2.11 (cf. [6], Proposition 4.1(a)). *Suppose that $\varphi : \mathscr{M} \to \mathbb{R}$ is a reflexive function which admits a Hessian at $x \in \mathscr{M}$. With $F(x)$ still defined by (2.5), we have $(x, \tau_1; F(x), \tau_2) \notin \mathscr{L}Cut$—hence $Q(\cdot, F(x))$ is smooth near x—and at x there holds*

$$(2.8) \qquad \nabla[Q(\cdot, F(x)) - \varphi] = 0; \quad \mathrm{Hess}[Q(\cdot, F(x)) - \varphi] \geqq 0.$$

To prove this, we have, similarly to (2.7), that for $u \in T_x\mathscr{M}$ sufficiently small,

$$\frac{\varphi(\exp_x u) + \varphi(\exp_x(-u)) - 2\varphi(x)}{|u|^2} \leqq \frac{Q(\exp_x u, F(x)) + Q(\exp_x(-u), F(x)) - 2Q(x, F(x))}{|u|^2},$$

and since the left-hand side is controlled from below in the limit $u \to 0$ (because the Hessian of φ exists) the right-hand side must be also. By Lemma A.5 in Appendix A, this implies that $(x, \tau_1, F(x), \tau_2) \notin \mathscr{L}Cut$ (hence the local smoothness of Q by Lemma A.2) and (2.8) follows by returning to Remark 2.5 and using the fact that $Q(\cdot, F(x)) - \varphi$ has a minimum at x. (The first part of (2.8) is already contained in Lemma 2.4.)

Combining Theorem 2.7 and Remark 2.8 with this lemma, we obtain:

Corollary 2.12. *Suppose that ν_1 and ν_2 are Borel probability measures, with ν_1 absolutely continuous with respect to (any) volume measure on $\mathscr{M}$. If we denote by π the optimal transference plan in the definition* (1.4) *of $V(\nu_1, \tau_1; \nu_2, \tau_2)$, then $\pi(\mathscr{L}Cut_{\tau_1,\tau_2}) = 0$.*

We next want to define a differential dF for the map F, where such a notion makes sense, and confirm that it has the properties one would expect given the name and notation.

Lemma 2.13 (cf. [6], Proposition 4.1(b)). *Suppose that $\varphi : \mathscr{M} \to \mathbb{R}$ is a reflexive function which admits a Hessian at $x \in \mathscr{M}$. Define F again by* (2.5), *and a map $dF(x) : T_x\mathscr{M} \to T_{F(x)}\mathscr{M}$ by*

$$(2.9) \qquad dF(x) := \frac{1}{2} d(\mathscr{L}_{\tau_1,\tau_2}\exp_x)\left(-\frac{\nabla\varphi(x)}{2}\right) \circ \left[\mathrm{hess}\big(Q(\cdot, F(x)) - \varphi\big)(x)\right],$$

where $\mathrm{hess}(f)(x)$ is the Hessian of a function $f : \mathscr{M} \to \mathbb{R}$ viewed as an endomorphism of $T_x\mathscr{M}$ (i.e. the covariant derivative of the gradient of f). Then for $u \in T_x\mathscr{M}$, we have

$$(2.10) \qquad \sup|v - dF(x)(u)| = o(|u|),$$

where the supremum is taken over all $v \in T_{F(x)}\mathscr{M}$ such that $\exp_{F(x)}^{g(\tau_2)}(v) \in F\big(\exp_x^{g(\tau_1)}(u)\big)$ and $|v|_{g(\tau_2)} = d\big(F(x), \exp_{F(x)}^{g(\tau_2)}(v), \tau_2\big)$.

It is worth pointing out that when φ is smooth in a neighbourhood of x (making F a smooth single-valued map in a neighbourhood of x) then this formula for $dF(x)$ coincides with the differential of F as classically defined.

As usual, the lemma above follows by adapting the corresponding proof from [6]. The same is true for the following result which uses the differential we have just defined to give a Jacobian identity.

Theorem 2.14 (cf. [6], Theorem 4.2). *Suppose that ν_1 and ν_2 are Borel probability measures on $\mathscr{M}$, which are absolutely continuous with respect to (any) volume measure. Let f_{τ_1} and f_{τ_2} be the densities defined by $d\nu_1 = f_{\tau_1}\, d\mu(\tau_1)$ and $d\nu_2 = f_{\tau_2}\, d\mu(\tau_2)$. If $\varphi : \mathscr{M} \to \mathbb{R}$ is a reflexive function for which $F_\#\nu_1 = \nu_2$ (where F is from* (2.5)) *as provided by Theorem* 2.7, *then there exists a Borel set $K \subset \mathscr{M}$ with $\nu_1(K) = 1$ such that*

- *φ admits a Hessian at each $x \in K$;*
- *for all $x \in K$, we have $f_{\tau_1}(x) = f_{\tau_2}(F(x)) \det dF(x) \neq 0$.*

We now have enough technology to construct $\mathscr{L}$-Wasserstein geodesics, along the lines of [6], Section 5. We define these to be one-parameter families of measures $\mathscr{V}_\tau$, with $\tau \in [\tau_1, \tau_2]$, which arise as the push-forwards $(F_\tau)_\# \nu_1$ by Borel maps $F_\tau : \mathscr{M} \to \mathscr{M}$ defined at points of differentiability of φ by

$$F_\tau(x) := \mathscr{L}_{\tau_1,\tau} \exp_x\left(-\frac{\nabla\varphi(x)}{2}\right). \tag{2.11}$$

The theory above involving F has always required that the function φ in the definition of F is reflexive, or more precisely, reflexive with respect to $[\tau_1, \tau_2]$. In order to apply the theory we have developed to F_τ as well as $F = F_{\tau_2}$, we must check that such a function φ is also reflexive with respect to $[\tau_1, \tau]$.

Lemma 2.15. *If $\varphi : \mathscr{M} \to \mathbb{R}$ is a reflexive function with respect to $[\tau_1, \tau_2]$, then for any $\tau \in (\tau_1, \tau_2)$, it is also reflexive with respect to $[\tau_1, \tau]$.*

Proof. By definition of Q, we have $Q(a, \tau_1; y, \tau_2) \leqq Q(a, \tau_1; z, \tau) + Q(z, \tau; y, \tau_2)$, and so with respect to $[\tau_1, \tau]$,

$$\begin{aligned} \check{\hat{\varphi}}(x) &= \inf_z \left[Q(x, \tau_1; z, \tau) - \inf_a [Q(a, \tau_1; z, \tau) - \varphi(a)]\right] \\ &\leqq \inf_z \left[Q(x, \tau_1; z, \tau) + Q(z, \tau; y, \tau_2) - \inf_a [Q(a, \tau_1; y, \tau_2) - \varphi(a)]\right] \\ &= Q(x, \tau_1; y, \tau_2) - \inf_a [Q(a, \tau_1; y, \tau_2) - \varphi(a)]. \end{aligned} \tag{2.12}$$

If we now minimise over $y \in \mathscr{M}$, the right-hand side becomes precisely $\check{\hat{\varphi}}(x)$ with respect to $[\tau_1, \tau_2]$, which is $\varphi(x)$ by hypothesis. Keeping in mind the first inequality of (2.3), the proof is complete. □

Note in particular, that in the context of Theorem 2.6, the maps F_τ will all map σ optimally to $\mathscr{V}_\tau := (F_\tau)_\# \sigma$.

Lemma 2.16 (cf. [6], Lemma 5.3). *If $\varphi : \mathscr{M} \to \mathbb{R}$ is a reflexive function and F_τ is defined as in* (2.11), *for x in the subset of $\mathscr{M}$ (of full measure) on which φ is differentiable, then F_τ is injective.*

In practice, we need a quantified version of this:

Lemma 2.17 (cf. [6], Proposition 5.4). *Suppose that ν_1 and ν_2 are Borel probability measures on $\mathscr{M}$ which are both absolutely continuous with respect to (any) volume measure. If $\varphi : \mathscr{M} \to \mathbb{R}$ is a reflexive function such that $(F_{\tau_2})_\# \nu_1 = \nu_2$, then for all $\tau \in (\tau_1, \tau_2]$, the interpolant measure $\mathscr{V}_\tau := (F_\tau)_\# \nu_1$ is also absolutely continuous with respect to (any) volume measure.*

Again, the proof is a translation of that in [6]. The analogue of the condition $\mathrm{Hess}\left(\frac{d^2_{F_t(x)}}{2} - t\phi\right) > 0$ in that proof in [6] translates to $\mathrm{Hess}\big(Q(\cdot,\tau_1;F_\tau(x),\tau) - \varphi\big) > 0$ in our setting, whilst the inequality $\mathrm{Hess}\left(\frac{d^2_{F_t(x)}}{2} - t\frac{d^2_{F(x)}}{2}\right) \geqq 0$ from [6] is simply $\mathrm{Hess}\big(Q(\cdot,\tau_1;F_\tau(x),\tau) - Q(\cdot,\tau_1;F_{\tau_2}(x),\tau_2)\big) \geqq 0$ for us (which follows immediately from the analogue of the triangle inequality

$$Q\big(a,\tau_1;F_{\tau_2}(x),\tau_2\big) \leqq Q\big(a,\tau_1;F_\tau(x),\tau\big) + Q\big(F_\tau(x),\tau;F_{\tau_2}(x),\tau_2\big).$$

In the next section, we want to analyse the behaviour of the classical entropy (to be defined in (3.15)) along $\mathscr{L}$-Wasserstein geodesics (see also [11], [6] and the references therein). We will compute this functional using the second part of Theorem 2.14 applied to F_τ, and hence we need to compute $\det dF_\tau(x)$ for x at certain points where φ admits a Hessian, where

$$dF_\tau(x) := \frac{1}{2} d(\mathscr{L}_{\tau_1,\tau}\exp_x)\left(-\frac{\nabla\varphi(x)}{2}\right) \circ \big[\mathrm{hess}\big(Q(\cdot,\tau_1;F_\tau(x),\tau) - \varphi\big)(x)\big],$$

is the generalisation of (2.9). In practice, we will do that with the following observation (cf. [6]) involving $\mathscr{L}$-Jacobi fields (which will be discussed further in Section 3).

Lemma 2.18. *Suppose that $\varphi : \mathscr{M} \to \mathbb{R}$ is a reflexive function which admits a Hessian at $x \in \mathscr{M}$, and that $\hat{Y} \in T_x\mathscr{M}$. Let $\gamma : [\tau_1,\tau_2] \to \mathscr{M}$ be the $\mathscr{L}$-geodesic $\gamma(\tau) = F_\tau(x)$, and define $Y \in \Gamma\big(\gamma^*(T\mathscr{M})\big)$ by $Y(\tau) := dF_\tau(x)(\hat{Y})$ for $\tau \in (\tau_1,\tau_2]$, and $Y(\tau_1) := \lim_{\tau\downarrow\tau_1} Y(\tau)$. Then $Y(\tau)$ is the $\mathscr{L}$-Jacobi field along γ, with initial data*

$$Y(\tau_1) = \hat{Y} \quad \textit{and} \quad D_\tau Y(\tau_1) = -\frac{1}{2\sqrt{\tau_1}}\,\mathrm{hess}(\varphi)(\hat{Y}).$$

Remark 2.19. Because of this lemma, we find that if we choose any orthonormal basis $\{\hat{Y}_i\}_{i=1,\dots,n}$ for $T_x\mathscr{M}$ (with respect to $g(\tau_1)$) and consider the $\mathscr{L}$-Jacobi fields $Y_i \in \Gamma\big(\gamma^*(T\mathscr{M})\big)$ determined by $Y_i(\tau_1) = \hat{Y}_i$ and $D_\tau Y_i(\tau_1) = -\frac{1}{2\sqrt{\tau_1}}\,\mathrm{hess}(\varphi)(\hat{Y}_i)$, then

$$\det dF_\tau(x) = \det\langle Y_i(\tau), Y_j(\tau)\rangle^{\frac{1}{2}}_{g(\tau)}.$$

3. Behaviour of Boltzmann-Shannon entropy along $\mathscr{L}$-Wasserstein geodesics

In this section, we perform the computations for $\mathscr{L}$-Jacobi fields which allow us to understand the behaviour of the entropy along an $\mathscr{L}$-Wasserstein geodesic. The discussion at the end of the last section motivates the inequalities of the following lemma. (We continue to consider a smooth (reverse) Ricci flow $g(\tau)$ defined on an open time interval including some interval $[\tau_1,\tau_2]$ with $0 < \tau_1 < \tau_2$.)

Lemma 3.1. *Suppose that $\gamma : [\tau_1,\tau_2] \to \mathscr{M}$ is an $\mathscr{L}$-geodesic, and $\{Y_i(\tau)\}_{i=1,\dots,n}$ is a set of $\mathscr{L}$-Jacobi fields along γ which form a basis of $T_{\gamma(\tau)}\mathscr{M}$ for each $\tau \in [\tau_1,\tau_2]$, with*

$\{Y_i(\tau_1)\}$ orthonormal and $\langle D_\tau Y_i, Y_j\rangle$ symmetric in i and j at $\tau = \tau_1$. Then defining $\alpha : [\tau_1, \tau_2] \to \mathbb{R}$ by $\alpha(\tau) = -\frac{1}{2}\ln\det\langle Y_i(\tau), Y_j(\tau)\rangle_{g(\tau)}$, and writing $\sigma = \tau^{\frac{1}{2}}$, we have

$$\frac{d^2\alpha}{d\sigma^2} = 4\tau^{\frac{1}{2}}\frac{d}{d\tau}\left(\tau^{\frac{1}{2}}\frac{d\alpha}{d\tau}\right) \geqq 2\tau H(X), \tag{3.1}$$

and

$$\frac{d^2(\sigma\alpha)}{d\sigma^2} = 4\frac{d}{d\tau}\left(\tau^{\frac{3}{2}}\frac{d\alpha}{d\tau}\right) \geqq 2\tau^{\frac{3}{2}}H(X) - n\tau^{-\frac{1}{2}}, \tag{3.2}$$

where $X = \gamma'(\tau)$ as before, and

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

is the Hamilton Harnack quantity [9], [13].

We clarify that Ric denotes the Ricci curvature of $g(\tau)$ viewed as a bilinear form, while Rc refers to that tensor viewed as an endomorphism, using $g(\tau)$.

Proof. The starting point for proving this is the equation for an $\mathscr{L}$-Jacobi field $Y(\tau)$

$$\begin{aligned} D_\tau^2 Y := D_\tau\big(D_\tau(Y)\big) = &- R(X, Y) + \frac{1}{2}\nabla_Y(\nabla R) - \nabla_Y \mathrm{Rc}(X) - 2\,\mathrm{Rc}(D_\tau Y) \\ &- \frac{1}{2\tau}D_\tau Y + \nabla_X \mathrm{Rc}(Y) - [\nabla\,\mathrm{Ric}(\cdot, X, Y)]^{\#}, \end{aligned} \tag{3.3}$$

where we are using the sign convention $R(X, Y)Z = -\nabla_X\nabla_Y Z + \nabla_Y\nabla_X Z + \nabla_{[X,Y]}Z$, and other conventions from [16].

This equation looks at first glance somewhat different to the $\mathscr{L}$-Jacobi equation elsewhere in the literature (e.g. [5], (7.121)) but our second derivative term is a little different to the conventional one, as we now clarify. Given a curve $\gamma : [\tau_1, \tau_2] \to \mathscr{M}$, and a metric g on $\mathscr{M}$, we denote by D_τ^g the pull-back of the Levi-Civita connection of g by γ, acting in the direction $\frac{\partial}{\partial\tau}$. Given a flow of metrics (e.g. a Ricci flow) our previous notation D_τ then coincides with $D_\tau^{g(\tau)}$ in this more general notation. At $\tau = \hat{\tau}$, the second derivative $D_\tau\big(D_\tau(Y)\big) = D_\tau^{g(\tau)}\big(D_\tau^{g(\tau)}(Y)\big) = D_\tau^{g(\hat{\tau})}\big(D_\tau^{g(\tau)}(Y)\big)$ is then *not* equal to $D_\tau^{g(\hat{\tau})}\big(D_\tau^{g(\hat{\tau})}(Y)\big)$ in general since the connection itself needs to be differentiated. Considering [16], Proposition 2.3.1, we have, at $\tau = \hat{\tau}$ that

$$\begin{aligned} D_\tau\big(D_\tau(Y)\big) = D_\tau^{g(\hat{\tau})}\big(D_\tau^{g(\hat{\tau})}(Y)\big) + (\nabla_Y \mathrm{Rc})(X) + (\nabla_X \mathrm{Rc})(Y) \\ - [(\nabla\,\mathrm{Ric})(\cdot, X, Y)]^{\#}, \end{aligned} \tag{3.4}$$

which accounts for the extra terms.

Consider the frame field $e_i \in \Gamma(\gamma^*(T\mathscr{M}))$, $i = 1, \dots, n$, with $e_i(\tau_1) := Y_i(\tau_1)$ orthonormal, satisfying the ODE

$$D_\tau e_i + \mathrm{Rc}(e_i) = 0. \tag{3.5}$$

Then $\{e_i(\tau)\}$ is an orthonormal frame for all $\tau \in [\tau_1, \tau_2]$. We write $Y_j(\tau) = A_{kj}(\tau)e_k(\tau)$ for a τ-dependent $n \times n$ matrix A. By applying D_τ both once and twice, and taking the inner product $(g(\tau))$ with e_i, we find that

$$A'_{ij} = \langle D_\tau Y_j, e_i \rangle + A_{kj} \operatorname{Ric}(e_k, e_i), \tag{3.6}$$

and

$$A''_{ij} = \langle D_\tau^2 Y_j, e_i \rangle + 2A'_{kj} \operatorname{Ric}(e_k, e_i) + A_{kj} \langle D_\tau(\mathrm{Rc}(e_k)), e_i \rangle. \tag{3.7}$$

The first term on the right-hand side of (3.7) can be dealt with using (3.3) and the definition of A_{ij}. We find that

$$\begin{aligned} \langle D_\tau^2 Y_j, e_i \rangle = A_{kj} \Big[&-\mathrm{Rm}(X, e_k, X, e_i) + \frac{1}{2} \operatorname{Hess}(R)(e_i, e_k) + \nabla_X \operatorname{Ric}(e_i, e_k) \\ &- \langle \nabla_{e_k} \mathrm{Rc}(X), e_i \rangle - \langle \nabla_{e_i} \mathrm{Rc}(X), e_k \rangle \\ &+ 2\langle \mathrm{Rc}^2(e_k), e_i \rangle + \frac{1}{2\tau} \operatorname{Ric}(e_i, e_k) \Big] \\ &- 2A'_{kj} \operatorname{Ric}(e_i, e_k) - \frac{1}{2\tau} A'_{ij}. \end{aligned} \tag{3.8}$$

The inner product of the third term on the right-hand side of (3.7) can be expanded out, using the definition of $\{e_i\}$, to give

$$\langle D_\tau(\mathrm{Rc}(e_k)), e_i \rangle = \frac{\partial \operatorname{Ric}}{\partial \tau}(e_i, e_k) + \nabla_X \operatorname{Ric}(e_i, e_k) - 3\langle \mathrm{Rc}^2(e_k), e_i \rangle. \tag{3.9}$$

(One pitfall to avoid here is that while Ric and Rc differ only by "raising/lowering an index", the tensors $\frac{\partial \operatorname{Ric}}{\partial \tau}$ and $\frac{\partial \mathrm{Rc}}{\partial \tau}$ do not, because raising/lowering an index involves using the metric $g(\tau)$ which depends on τ.) Combining (3.7), (3.8) and (3.9), we find that

$$A'' + \frac{1}{2\tau} A' = MA, \tag{3.10}$$

where $M(\tau)$ is the τ-dependent $n \times n$ symmetric matrix given by

$$\begin{aligned} M_{ik} = &-\mathrm{Rm}(X, e_i, X, e_k) + \frac{1}{2} \operatorname{Hess}(R)(e_i, e_k) \\ &- \langle \nabla_{e_k} \mathrm{Rc}(X), e_i \rangle - \langle \nabla_{e_i} \mathrm{Rc}(X), e_k \rangle \\ &+ 2\nabla_X \operatorname{Ric}(e_i, e_k) - \langle \mathrm{Rc}^2(e_k), e_i \rangle + \frac{1}{2\tau} \operatorname{Ric}(e_i, e_k) + \frac{\partial \operatorname{Ric}}{\partial \tau}(e_i, e_k). \end{aligned} \tag{3.11}$$

The trace of M is then

$$\operatorname{tr} M = -\operatorname{Ric}(X, X) + \frac{1}{2}\Delta R + 2\delta \operatorname{Ric}(X) + 2X(R) - |\operatorname{Ric}|^2 + \frac{R}{2\tau} + \operatorname{tr}\frac{\partial \operatorname{Ric}}{\partial \tau},$$

but exploiting the contracted second Bianchi identity $2\delta \operatorname{Ric} + dR = 0$ (using the notation and conventions of [16], (2.1.9)) along with the evolution equation $-\frac{\partial R}{\partial \tau} = \Delta R + 2|\operatorname{Ric}|^2$ for the scalar curvature [16], Proposition 2.5.4, and the fact that $\operatorname{tr}\frac{\partial \operatorname{Ric}}{\partial \tau} = \frac{\partial R}{\partial \tau} + 2|\operatorname{Ric}|^2$ from [16], Proposition 2.3.6, this simplifies to

$$\operatorname{tr} M = -\frac{1}{2} H(X). \tag{3.12}$$

We are now in a position to compute the volume element $\alpha(\tau)$ of the lemma, in the spirit of classical comparison geometry, and following the analogous [7], Lemma 6. By definition, $\alpha(\tau) = -\ln \det A$, and so $\frac{d\alpha}{d\tau} = -\operatorname{tr}\left(\frac{dA}{d\tau}A^{-1}\right)$ and

$$\frac{d^2\alpha}{d\tau^2} = \operatorname{tr}\left(\frac{dA}{d\tau}A^{-1}\frac{dA}{d\tau}A^{-1}\right) - \operatorname{tr}\left(\frac{d^2A}{d\tau^2}A^{-1}\right).$$

If we define $B := \frac{dA}{d\tau}A^{-1}$, then this may be combined with (3.10) and (3.12) to give

$$\begin{aligned} \tau^{-\frac{1}{2}}\frac{d}{d\tau}\left(\tau^{\frac{1}{2}}\frac{d\alpha}{d\tau}\right) &= \operatorname{tr}\left(\frac{dA}{d\tau}A^{-1}\frac{dA}{d\tau}A^{-1}\right) - \operatorname{tr}\left(\left(\frac{d^2A}{d\tau^2} + \frac{1}{2\tau}\frac{dA}{d\tau}\right)A^{-1}\right) \\ &= \operatorname{tr} B^2 + \frac{1}{2}H(X), \end{aligned} \tag{3.13}$$

and

$$\tau^{-\frac{3}{2}}\frac{d}{d\tau}\left(\tau^{\frac{3}{2}}\frac{d\alpha}{d\tau}\right) = \operatorname{tr}\left(\left(B - \frac{1}{2\tau}I\right)^2\right) + \frac{1}{2}H(X) - \frac{n}{4\tau^2}. \tag{3.14}$$

It remains to show that B (and hence also $B - \frac{1}{2\tau}I$) is symmetric, so that the first terms on the right-hand sides of (3.13) and (3.14) are (weakly) positive. But following [7], Lemma 6 by writing $B^T - B = (A^{-1})^T H A^{-1}$, where $H := \frac{dA^T}{d\tau}A - A^T\frac{dA}{d\tau}$, and noting that $\frac{d}{d\tau}(\tau^{\frac{1}{2}}H) = 0$ and that $H(\tau_1)$ is the zero matrix (because at $\tau = \tau_1$, $A = I$, and—using (3.6) and the hypothesis of the lemma—$\frac{dA}{d\tau}(\tau_1)$ is symmetric) we see that $H(\tau)$ is zero for any τ, and hence B is symmetric for any τ. □

We now turn to consider the Boltzmann-Shannon entropy of probability measures $f\,d\mu$, where μ is Riemannian volume measure, and f is a suitably regular weakly positive function on $\mathscr{M}$, defined by

$$E(f\,d\mu) = \int_{\mathscr{M}} f \ln f \, d\mu. \tag{3.15}$$

We are now in a position to investigate the behaviour of this entropy along $\mathscr{L}$-Wasserstein geodesics (as defined in the previous section) with a result analogous to [11], Lemma 8. We re-use the alternative variable $\sigma = \tau^{\frac{1}{2}}$.

Lemma 3.2. *Suppose that $\mathscr{V}_\tau$ is an $\mathscr{L}$-Wasserstein geodesic, for $\tau \in [\tau_1, \tau_2]$, induced by a potential $\varphi : \mathscr{M} \to \mathbb{R}$, with $\mathscr{V}_{\tau_1}$ and $\mathscr{V}_{\tau_2}$ both smooth, and write $d\mathscr{V}_\tau = f_\tau\, d\mu(\tau)$ where $\mu(\tau)$ is the volume measure of $g(\tau)$. Then for all $\tau \in [\tau_1, \tau_2]$, we have $f_\tau \in L \ln L\big(\mu(\tau)\big)$, and the function $\tau \to E(\mathscr{V}_\tau)$ is semiconvex and satisfies, for almost all $\tau \in [\tau_1, \tau_2]$ (where $\sigma \to E(\mathscr{V}_\tau)$ admits a second derivative in the sense of Alexandrov)*

$$4\tau^{\frac{1}{2}} \frac{d}{d\tau}\left(\tau^{\frac{1}{2}} \frac{dE(\mathscr{V}_\tau)}{d\tau}\right) = \frac{d^2}{d\sigma^2} E(\mathscr{V}_\tau) \geqq 2\tau \int_{\mathscr{M}} H\big(X(\tau)\big)\, d\mathscr{V}_{\tau_1}, \tag{3.16}$$

$$4\frac{d}{d\tau}\left(\tau^{\frac{3}{2}} \frac{dE(\mathscr{V}_\tau)}{d\tau}\right) = \frac{d^2}{d\sigma^2}\big(\sigma E(\mathscr{V}_\tau)\big) \geqq 2\tau^{\frac{3}{2}} \int_{\mathscr{M}} H\big(X(\tau)\big)\, d\mathscr{V}_{\tau_1} - n\tau^{-\frac{1}{2}}, \tag{3.17}$$

where $X(\tau)$, at a point $x \in \mathscr{M}$ where φ admits a Hessian, is $\gamma'(\tau)$, for $\gamma : [\tau_1, \tau_2] \to \mathscr{M}$ the minimising $\mathscr{L}$-geodesic from x to $F(x)$. Moreover, the one-sided derivatives of $E(\mathscr{V}_\tau)$ at τ_1 and τ_2 exist, with

$$\frac{d}{d\tau}\bigg|_{\tau_1^+} E(\mathscr{V}_\tau) \geqq -\int_{\mathscr{M}} \left(R(\cdot, \tau_1) + \left\langle \frac{\nabla\varphi}{2\sqrt{\tau_1}}, \nabla \ln f_{\tau_1} \right\rangle\right) d\mathscr{V}_{\tau_1}. \tag{3.18}$$

The φ of the lemma is the φ which induces the $\mathscr{L}$-Wasserstein geodesic under consideration. This also induces a map F via (2.5) which we use below.

Proof. The main ingredient in the proof is Lemma 3.1, applied to $\mathscr{L}$-geodesics and $\mathscr{L}$-Jacobi fields arising in Remark 2.19. Note that our volume density $\alpha(\tau)$ will now have a (suppressed) x-dependency. At the core of the proof of Lemma 3.2 is the fact that we can relate the entropy at different values of τ in terms of the volume density α. With K_τ the set provided by Theorem 2.14 with τ in place of τ_2, we have (by that theorem)

$$\begin{aligned} E(\mathscr{V}_\tau) &= \int_{\mathscr{M}} \ln f_\tau \, d\mathscr{V}_\tau = \int_{\mathscr{M}} \ln f_\tau \, d\big((F_\tau)_\# \mathscr{V}_{\tau_1}\big) = \int_{K_\tau} \ln f_\tau \circ F_\tau \, d\mathscr{V}_{\tau_1} \\ &= \int_{K_\tau} \ln \frac{f_{\tau_1}}{\det dF_\tau} d\mathscr{V}_{\tau_1} = E(\mathscr{V}_{\tau_1}) + \int_{K_\tau} \alpha(\tau)\, d\mathscr{V}_{\tau_1}. \end{aligned} \tag{3.19}$$

We will combine this with Lemma 3.1 to yield the result. Indeed, that lemma gives immediately a lower bound $\dfrac{d^2\alpha}{d\sigma^2} \geqq -C$, for $C < \infty$ independent of the point $x \in \mathscr{M}$ at which we compute α, and this gives the semiconvexity of $E(\mathscr{V}_\tau)$ (with respect to σ, or equivalently τ). For values of σ where $\sigma \to E(\mathscr{V}_\tau)$ admits a second derivative in the sense of Alexandrov, the identity

$$\frac{d^2}{d\sigma^2} E(\mathscr{V}_\tau) := \lim_{\delta \to 0} \frac{E(\mathscr{V}_{(\sigma+\delta)^2}) + E(\mathscr{V}_{(\sigma-\delta)^2}) - 2E(\mathscr{V}_{\sigma^2})}{\delta^2} = \int_{\mathscr{M}} \frac{\partial^2 \alpha}{\partial \sigma^2}(\sigma)\, d\mathscr{V}_{\tau_1}$$

follows from (3.19) thanks to Fatou's lemma, as in the proof of [11], Lemma 8. This combines with Lemma 3.1 to give (3.16), and a similar approach gives (3.17).

By semiconvexity, the one-sided derivative of $E(\mathscr{V}_\tau)$ at $\tau = \tau_1$ must exist, allowing the possibility that it is $-\infty$. If we take any sequence $\tau_k \downarrow \tau_1$, and set $K = \bigcap_k K_{\tau_k}$, then we may exploit (3.19) once again to give that

$$\left.\frac{d}{d\tau}\right|_{\tau_1^+} E(\mathscr{V}_\tau) = \lim_{k \to \infty} \int_K \frac{\alpha(\tau_k) - \alpha(\tau_1)}{\tau_k - \tau_1}\, d\mathscr{V}_{\tau_1}.$$

If α were a convex function of τ for each x, then the monotone convergence theorem would tell us that

$$\left.\frac{d}{d\tau}\right|_{\tau_1^+} E(\mathscr{V}_\tau) = \int_K \alpha'(\tau_1)\, d\mathscr{V}_{\tau_1}, \tag{3.20}$$

and it is not hard to see that the same conclusion follows from the known semiconvexity of α. By definition of α, keeping in mind that the $\mathscr{L}$-Jacobi fields on which α depends were chosen as in Remark 2.19, we have (at $\tau = \tau_1$)

$$\begin{aligned} \alpha'(\tau_1) &= -\frac{1}{2}\,\mathrm{tr}\left[\frac{d}{d\tau}\langle Y_i, Y_j\rangle_{g(\tau)}\right] \\ &= -\frac{1}{2}\,\mathrm{tr}[2\,\mathrm{Ric}(Y_i, Y_j) + \langle D_\tau Y_i, Y_j\rangle + \langle Y_i, D_\tau Y_j\rangle] \\ &= -R - \sum_i \langle D_\tau Y_i, \hat{Y}_i\rangle \\ &= -R + \frac{1}{2\sqrt{\tau_1}}\Delta\varphi. \end{aligned} \tag{3.21}$$

Returning to (3.20), and exploiting the semiconcavity of φ to be sure that the singular part of the distributional Laplacian $\Delta_{\mathscr{D}'}\varphi$ of φ is weakly negative, we have

$$\begin{aligned} \left.\frac{d}{d\tau}\right|_{\tau_1^+} E(\mathscr{V}_\tau) &= \int_K \left(-R(x,\tau_1) + \frac{1}{2\sqrt{\tau_1}}\Delta\varphi\right) d\mathscr{V}_{\tau_1} \\ &\geqq \int_{\mathscr{M}} \left(-R(x,\tau_1) + \frac{1}{2\sqrt{\tau_1}}\Delta_{\mathscr{D}'}\varphi\right) d\mathscr{V}_{\tau_1}. \end{aligned} \tag{3.22}$$

Because $d\mathscr{V}_{\tau_1} = f_{\tau_1}\, d\mu(\tau_1)$, this gives (3.18) as desired. □

We have given (3.16) for use in future work. Here we only require (3.17), and then only the version of it one obtains by integrating with respect to τ (not σ). Indeed, by semiconcavity of $\sigma E(\mathscr{V}_\tau)$, we can see that (3.17) holds in the distributional sense, and integrates to

$$\left[2\tau^{\frac{3}{2}}\frac{dE(\mathscr{V}_\tau)}{d\tau}\right]_{\tau_1}^{\tau_2} \geqq \int_{\tau_1}^{\tau_2}\left(\tau^{\frac{3}{2}}\int_{\mathscr{M}} H\big(X(\tau)\big)\,d\mathscr{V}_{\tau_1} - \frac{n}{2}\tau^{-\frac{1}{2}}\right)d\tau \tag{3.23}$$

$$= \int_{\mathscr{M}\times\mathscr{M}} \mathscr{K}(x,\tau_1,y,\tau_2)\,d\pi(x,y) - n(\sqrt{\tau_2}-\sqrt{\tau_1}),$$

where $\mathscr{K}$ is defined in (A.8) of Appendix A, and π is the push forward of $\mathscr{V}_{\tau_1}$ by the map $x \to \big(x, F(x)\big)$ as in Remark 2.8.

Meanwhile, by exploiting Lemma 2.11 to write $\nabla\varphi(x) = \nabla_1 Q\big(x,\tau_1;F(x),\tau_2\big)$ for appropriate x, where $\nabla_1 Q$ represents the gradient of Q with respect to its first argument, and with respect to $g(\tau_1)$, we can rewrite (3.18) as

$$\left.\frac{d}{d\tau}\right|_{\tau_1} E(\mathscr{V}_\tau) \geqq -\int_{\mathscr{M}\times\mathscr{M}}\left(R(x,\tau_1) + \left\langle\frac{\nabla_1 Q(x,\tau_1;y,\tau_2)}{2\sqrt{\tau_1}}, \nabla\ln f_{\tau_1}(x)\right\rangle\right)d\pi(x,y), \tag{3.24}$$

where π is again the push forward of $\mathscr{V}_{\tau_1}$ by the map $x \to \big(x,F(x)\big)$. Taking this viewpoint, we have the right notation to give the analogous inequality for the other one-sided derivative:

$$\left.\frac{d}{d\tau}\right|_{\tau_2} E(\mathscr{V}_\tau) \leqq \int_{\mathscr{M}\times\mathscr{M}}\left(-R(y,\tau_2) + \left\langle\frac{\nabla_2 Q(x,\tau_1;y,\tau_2)}{2\sqrt{\tau_2}}, \nabla\ln f_{\tau_2}(y)\right\rangle\right)d\pi(x,y), \tag{3.25}$$

where $\nabla_2 Q$ is the gradient with respect to the y argument, and with respect to $g(\tau_2)$.

Combining (3.23), (3.24) and (3.25), we obtain the following corollary which is what we shall require in the proof of Theorem 1.1.

Corollary 3.3. *Under the hypotheses of Lemma* 3.2, *we have*

$$\int_{\mathscr{M}\times\mathscr{M}}\big(\mathscr{K} - 2\tau_1^{\frac{3}{2}}R(x,\tau_1) - \tau_1\langle\nabla_1 Q, \nabla\ln f_{\tau_1}(x)\rangle_{g(\tau_1)} + 2\tau_2^{\frac{3}{2}}R(y,\tau_2) \tag{3.26}$$

$$- \tau_2\langle\nabla_2 Q, \nabla\ln f_{\tau_2}(y)\rangle_{g(\tau_2)}\big)\,d\pi(x,y)$$

$$\leqq n(\sqrt{\tau_2}-\sqrt{\tau_1}),$$

where π is the optimal transference plan from $\mathscr{V}(\tau_1)$ to $\mathscr{V}(\tau_2)$ (for $\mathscr{L}$-optimal transport).

4. Proof of Theorem 1.1

We will follow [11], Section 4, as closely as possible. All we need to show is that

$$\left.\frac{d^+\Theta}{ds}\right|_{s=0} := \limsup_{s\downarrow 0}\frac{\Theta(s)-\Theta(0)}{s} \leqq 0.$$

Defining $h(s) := V\big(v_1(\tau_1(s)), \tau_1(s); v_2(\tau_2(s)), \tau_2(s)\big)$, this is equivalent to proving that

$$\text{(4.1)} \qquad \frac{d^+h}{ds}\bigg|_{s=0} \leqq -\frac{1}{2}h(0) + n(\sqrt{\overline{\tau_2}} - \sqrt{\overline{\tau_1}}).$$

Let us write $dv_i(\tau) = u_i(\tau)\, d\mu(\tau)$, so $u_i(\tau) : \mathscr{M} \to (0, \infty)$ is the probability density of $v_i(\tau)$, for $i = 1, 2$. Since v_1 and v_2 are diffusions, by Fourier's law, if we take families $\psi_i^\tau : \mathscr{M} \to \mathscr{M}$ of diffeomorphisms with $\psi_i^{\overline{\tau_i}}$ the identity, generated by $-\nabla \ln u_i$, then $(\psi_i^\tau)_\# v_i(\overline{\tau_i}) = v_i(\tau)$ for τ near $\overline{\tau_i}$.

If we now let π_0 be the minimiser for the variational problem (1.4) in the case of $V\big(v_1(\bar{\tau}_1), \bar{\tau}_1; v_2(\bar{\tau}_2), \bar{\tau}_2\big)$, and use $\pi_s := (\psi_1^{\tau_1(s)} \times \psi_2^{\tau_2(s)})_\# \pi_0$ as a competitor for the variational problem defining $V\big(v_1(\tau_1(s)), \tau_1(s); v_2(\tau_2(s)), \tau_2(s)\big)$, then we may compute similarly to [11] that

$$\text{(4.2)} \qquad h(s) - h(0) \leqq \int_{\mathscr{M}\times\mathscr{M}} \big(Q\big(\psi_1^{\tau_1(s)}(x), \tau_1(s); \psi_2^{\tau_2(s)}(y), \tau_2(s)\big) - Q(x, \overline{\tau_1}; y, \overline{\tau_2})\big)\, d\pi_0(x, y),$$

and hence (keeping in mind Lemma A.2 and Corollary 2.12) we have

$$\text{(4.3)} \qquad \frac{d^+h}{ds}\bigg|_{s=0} \leqq \int_{\mathscr{M}\times\mathscr{M}} \frac{d}{ds}\bigg|_{s=0} Q\big(\psi_1^{\tau_1(s)}(x), \tau_1(s); \psi_2^{\tau_2(s)}(y), \tau_2(s)\big)\, d\pi_0(x, y)$$
$$= \int_{\mathscr{M}\times\mathscr{M}} \bigg(\langle \nabla_1 Q, -\nabla \ln u_1(\overline{\tau_1})\rangle \overline{\tau_1} + \frac{\partial Q}{\partial \tau_1}\overline{\tau_1} + \langle \nabla_2 Q, -\nabla \ln u_2(\overline{\tau_2})\rangle \overline{\tau_2} + \frac{\partial Q}{\partial \tau_2}\overline{\tau_2}\bigg)\, d\pi_0,$$

where the inner products used are $g(\tau_1)$ and $g(\tau_2)$ respectively. Exploiting Lemma A.6 from Appendix A, this may be written

$$\text{(4.4)} \qquad \frac{d^+h}{ds}\bigg|_{s=0} \leqq \int_{\mathscr{M}\times\mathscr{M}} \big(-\overline{\tau_1}\langle \nabla_1 Q, \nabla \ln u_1(\overline{\tau_1})\rangle - \overline{\tau_2}\langle \nabla_2 Q, \nabla \ln u_2(\overline{\tau_2})\rangle + 2\overline{\tau_2}^{\frac{3}{2}} R(y, \overline{\tau_2}) - 2\overline{\tau_1}^{\frac{3}{2}} R(x, \overline{\tau_1}) + \mathscr{K}\big)\, d\pi_0 - \frac{1}{2}h(0).$$

By Corollary 3.3, we deduce our desired (4.1).

Appendix A. Theory of $\mathscr{L}$-length

In this appendix we briefly survey the theory of Perelman's $\mathscr{L}$-length, and the distance Q it induces. Some of this material can be found in the foundational paper [13], §7, described for $L(y, \tau) := Q(x, 0; y, \tau)$. Details of the remaining parts can either be found in [5], Chapter 7, or [19], or can be arrived at by adapting the analogous theory of Riemannian length in Riemannian geometry.

Throughout this appendix, we will be considering a smooth (reverse) Ricci flow $g(\tau)$ on a closed manifold $\mathscr{M}$, defined on an open time interval $(\hat{\tau}_1, \hat{\tau}_2)$ where $0 < \hat{\tau}_1 < \tau_1 < \tau_2 < \hat{\tau}_2$. We use the notation

$$\Upsilon := \{(x, \tau_a; y, \tau_b) \mid x, y \in \mathscr{M} \text{ and } \hat{\tau}_1 < \tau_a < \tau_b < \hat{\tau}_2\}.$$

We will continue to abbreviate $Q(x, \tau_1; y, \tau_2)$ by $Q(x, y)$.

We have already recalled, in the introduction and Section 2, the definitions of $\mathscr{L}$-length, $\mathscr{L}$-geodesics and the $\mathscr{L}$-exponential map, and we shall require these in this appendix. (One can also define an $\mathscr{L}$-exponential map *backwards* in time.) The notion of $\mathscr{L}$-geodesic also induces a notion of $\mathscr{L}$-Jacobi field, and we refer the reader to [5], Chapter 7, for details.

The first point to note is that the distance function $Q(x, \tau_1; y, \tau_2)$ defined in the introduction is locally Lipschitz on Υ, where we use the metric on Υ arising as the sum of $g(\tau_1)$, $d\tau_1^2$, $g(\tau_2)$ and $d\tau_2^2$ corresponding to the four respective arguments of Q. This follows by a variation on the argument for showing that Perelman's L distance is locally Lipschitz (see e.g. [5], Lemma 7.30, or [19]). In particular, for fixed τ_1 and τ_2, the function $Q(\cdot, \tau_1; y, \tau_2)$ is Lipschitz with Lipschitz constant independent of $y \in \mathscr{M}$.

Just as for Riemannian distance, Q need not be smooth; Q will fail to be smooth on some subset $\mathscr{L}Cut \subset \Upsilon$ which we will define and study now. It is convenient to define

$$\text{(A.1)} \qquad \Omega(x, \tau_1; \tau_2) := \{Z \in T_x\mathscr{M} \mid \gamma : [\tau_1, \tau_2] \to \mathscr{M} \text{ defined by } \gamma(\tau) = \mathscr{L}_{\tau_1, \tau} \exp_x(Z) \text{ is a minimising } \mathscr{L}\text{-geodesic}\}.$$

This set clearly shrinks as τ_2 is increased, and exhausts $T_x\mathscr{M}$ in the limit $\tau_2 \downarrow \tau_1$. (Beware, however, that it need not be star-shaped as would its classical Riemannian analogue.) Define $\Omega^*(x, \tau_1; \hat{\tau}_2)$ to be the intersection of all the sets $\Omega(x, \tau_1; \tau_2)$, over $\tau_2 \in (\tau_1, \hat{\tau}_2)$. For $Z \in T_x\mathscr{M} \backslash \Omega^*(x, \tau_1; \hat{\tau}_2)$, define $\bar{\tau}(x, \tau_1; Z) := \sup\{\tau \in (\tau_1, \hat{\tau}_2) \mid Z \in \Omega(x, \tau_1; \tau)\} \in (\tau_1, \hat{\tau}_2)$. We can then define the possibly empty set

$$\text{(A.2)} \qquad \mathscr{L}Cut := \{(x, \tau_1; \mathscr{L}_{\tau_1, \bar{\tau}(x, \tau_1; Z)} \exp_x(Z), \bar{\tau}(x, \tau_1; Z)) \mid x \in \mathscr{M}, \tau_1 \in (\hat{\tau}_1, \hat{\tau}_2), Z \in T_x\mathscr{M} \backslash \Omega^*(x, \tau_1; \hat{\tau}_2)\}.$$

It will also be convenient to define the slice $\mathscr{L}Cut_{\tau_1, \tau_2}$ to be the subset of $\mathscr{L}Cut$ consisting of those points of the form $(x, \tau_1; y, \tau_2)$ for some $x, y \in \mathscr{M}$.

Remark A.1. Much of the classical theory of cut loci in Riemannian geometry carries over to $\mathscr{L}Cut$. In particular, $\mathscr{L}Cut$ can be characterised as the union of two sets: the first consisting of points $(x, \tau_1; y, \tau_2)$ such that there exists more than one minimising $\mathscr{L}$-geodesic $\gamma : [\tau_1, \tau_2] \to \mathscr{M}$ with $\gamma(\tau_1) = x$ and $\gamma(\tau_2) = y$, and the second consisting of points $(x, \tau_1; y, \tau_2)$ such that y is conjugate to x (with respect to $\mathscr{L}$-Jacobi fields) along a minimising $\mathscr{L}$-geodesic $\gamma : [\tau_1, \tau_2] \to \mathscr{M}$ with $\gamma(\tau_1) = x$ and $\gamma(\tau_2) = y$.

This characterisation is used to prove the first three parts of the following lemma, using the techniques of the proof of [11], Lemma 5. Implicit here is the existence of minimising $\mathscr{L}$-geodesics between given end-points (see, for example, [5], Lemma 7.27).

Lemma A.2. *We have that*:

(i) *The set $\mathscr{L}Cut$ is closed in Υ.*

(ii) *The function Q is smooth on $\Upsilon \backslash \mathscr{L}Cut$.*

(iii) *The minimising $\mathscr{L}$-geodesic corresponding to each point in $\Upsilon \backslash \mathscr{L}Cut$ is smoothly dependent on that point in the sense that if we associate to each point $(x, \tau_1; y, \tau_2) \in \Upsilon \backslash \mathscr{L}Cut$ the vector $Z \in \Omega(x, \tau_1; \tau_2) \subset T_x\mathscr{M}$ for which $\mathscr{L}_{\tau_1,\tau_2} \exp_x(Z) = y$, then Z depends smoothly on $(x, \tau_1; y, \tau_2)$.*

(iv) *On $\Upsilon \backslash \mathscr{L}Cut$ we have*

$$\text{(A.3)} \qquad \frac{\partial Q}{\partial \tau_1}(x, \tau_1; y, \tau_2) = \sqrt{\tau_1}\left(\frac{|Z|^2}{\tau_1} - R(x, \tau_1)\right); \quad \nabla_1 Q(x, \tau_1; y, \tau_2) = -2Z,$$

where $\nabla_1 Q$ denotes the gradient with respect to the first argument, x, using the metric $g(\tau_1)$. Analogous formulae hold for the derivatives with respect to y and τ_2.

The equations of part (iv) are similar to Perelman's formulae for the derivatives of L, from [13], §7. When we write $|Z|$ here, we mean its length with respect to $g(\tau_1)$.

To extend the lemma above, we need further notation. Suppose

$$(x, \tau_1; y, \tau_2) \in \Upsilon \backslash \mathscr{L}Cut,$$

and let $\gamma : [\tau_1, \tau_2] \to \mathscr{M}$ be the minimising $\mathscr{L}$-geodesic from x to y. We write $X(\tau) = \gamma'(\tau)$, so $\sqrt{\tau_1}X(\tau_1)$ coincides with the Z of the previous lemma. In place of (A.3) we have

$$\text{(A.4)} \qquad \frac{\partial Q}{\partial \tau_1}(x, \tau_1; y, \tau_2) = \sqrt{\tau_1}\big(|X(\tau_1)|^2 - R(x, \tau_1)\big);$$

$$\nabla_1 Q(x, \tau_1; y, \tau_2) = -2\sqrt{\tau_1}X(\tau_1),$$

and the corresponding formulae for the other derivatives of Q are then

$$\text{(A.5)} \quad \frac{\partial Q}{\partial \tau_2}(x, \tau_1; y, \tau_2) = \sqrt{\tau_2}\big(R(y, \tau_2) - |X(\tau_2)|^2\big); \quad \nabla_2 Q(x, \tau_1; y, \tau_2) = 2\sqrt{\tau_2}X(\tau_2).$$

We now have enough control on Q off $\mathscr{L}Cut$, but we need at least some control on Q across its whole domain. Up to now, we know simply that it is locally Lipschitz. The first observation is that although $Q(\cdot, y)$ need not be everywhere differentiable on $\mathscr{M}$, it does admit a supergradient everywhere:

Lemma A.3. *For all $x, y \in \mathscr{M}$, if we pick $Z \in \Omega(x, \tau_1; \tau_2) \subset T_x\mathscr{M}$ such that $y = \mathscr{L}_{\tau_1,\tau_2} \exp_x(Z)$, then $-2Z$ is a supergradient of $Q(\cdot, y)$ at x.*

One can easily construct an upper barrier for $Q(\cdot, y)$ at x which implies this lemma, either using the so-called Calabi trick, or by considering the $\mathscr{L}$-lengths of a smooth variation of a minimising $\mathscr{L}$-geodesic from x to y. See [5], Lemma 7.32, for this latter approach to prove the corresponding result for L instead of Q.

We now turn to look at second derivative properties of Q. We need several times in the paper that $Q(\cdot, y)$ is semiconcave. (This means that near each point, one can add a smooth function to give a (geodesically) concave function. This notion is independent of the choice of metric [1].) However, we also need $Q(\cdot, y)$ to be uniformly semiconcave (see [6]) which is stronger and is the content of the following lemma.

Lemma A.4 (cf. [6], Corollary 3.13). *There exists $C \leqq \infty$ such that for all $x, y \in \mathscr{M}$ and $v \in T_x\mathscr{M}$,*

$$\text{(A.6)} \qquad \limsup_{r \to 0} \frac{Q\big(\exp_x^{g(\tau_1)}(rv), y\big) + Q\big(\exp_x^{g(\tau_1)}(-rv), y\big) - 2Q(x, y)}{r^2} \leqq C.$$

We stress that C here is independent of x and y (and v). Note that according to [6], Lemma 3.11, such a uniform estimate implies semiconcavity. This uniform estimate can be proved by exploiting the second variation calculations of [13], §7. See also the discussion of Hessian bounds in [19].

Although Lemma A.4 gives an upper bound for the left-hand side of (A.6), there is no lower bound in general.

Lemma A.5 (cf. [6], Proposition 2.5). *For each point $(x, \tau_1; y, \tau_2) \in \mathscr{L}Cut$, the function $Q(\cdot, \tau_1; y, \tau_2)$ is not smooth at x, and its Hessian is unbounded below in the sense that*

$$\text{(A.7)} \qquad \liminf_{u \to 0} \frac{Q\big(\exp_x^{g(\tau_1)}(u), y\big) + Q\big(\exp_x^{g(\tau_1)}(-u), y\big) - 2Q(x, y)}{|u|^2} = -\infty.$$

To prove this, one can follow the proof of the corresponding Riemannian result from [6]. One should deal with each point of $\mathscr{L}Cut$ differently depending on its place in the characterisation of $\mathscr{L}Cut$ mentioned in Remark A.1. For example, if there are two distinct minimising $\mathscr{L}$-geodesics from x to y along which x and y are not conjugate, then each can be used to construct upper barriers for $Q(\cdot, y)$ with different gradients at x, and the result is clear in that case. When x and y are conjugate along a minimising $\mathscr{L}$-geodesic, then one considers the $\mathscr{L}$-index form (refer to [5], (7.134)) analogously to the proof of [6], Proposition 2.5.

Finally, the proof of the main theorem in Section 4 will require a formula involving the derivatives of Q which we derive now. Suppose $(x, \tau_1; y, \tau_2) \in \Upsilon \backslash \mathscr{L}Cut$, let $\gamma : [\tau_1, \tau_2] \to \mathscr{M}$ be the minimising $\mathscr{L}$-geodesic from x to y, and write $X(\tau) = \gamma'(\tau)$ as before. Following [13], we define

$$\text{(A.8)} \qquad \mathscr{K} = \mathscr{K}(x, \tau_1, y, \tau_2) := \int_{\tau_1}^{\tau_2} \tau^{\frac{3}{2}} H\big(X(\tau)\big)\, d\tau,$$

where $H(\cdot)$ is the Hamilton Harnack quantity defined in Lemma 3.1. Perelman was only considering the case $\tau_1 = 0$ at this point in his work, but the direct analogue of [13], (7.4) is

$$\tau_2^{\frac{3}{2}}\big(R(y,\tau_2) + |X(\tau_2)|^2\big) - \tau_1^{\frac{3}{2}}\big(R(x,\tau_1) + |X(\tau_1)|^2\big) = -\mathscr{K}(x,\tau_1,y,\tau_2) + \frac{1}{2}Q(x,\tau_1;y,\tau_2). \tag{A.9}$$

Combining with (A.4) and (A.5), we find (corresponding to [13], (7.5)) the following lemma.

Lemma A.6. *Under Ricci flow, Q satisfies*

$$\tau_2\frac{\partial Q}{\partial \tau_2} + \tau_1\frac{\partial Q}{\partial \tau_1} = 2\tau_2^{\frac{3}{2}}R(y,\tau_2) - 2\tau_1^{\frac{3}{2}}R(x,\tau_1) + \mathscr{K} - \frac{1}{2}Q. \tag{A.10}$$

Appendix B. Wasserstein and $\mathscr{L}$-Wasserstein distance, and their infinitesimal versions

Throughout this appendix, we will be considering a (reverse) Ricci flow $g(\tau)$ defined on a closed manifold $\mathscr{M}$, with τ ranging over an open interval containing $[\tau_1, \tau_2]$, for $0 < \tau_1 < \tau_2$. We will be concerned with the limit in which τ_1 and τ_2 approach each other.

We start by noting the most elementary relationship between V and W_2.

Lemma B.1. *Suppose ν_1 and ν_2 are Borel probability measures on $\mathscr{M}$. Define $\underline{R} := \inf R$, $\bar{R} := \sup R$ and $R_1 := \sup|\mathrm{Ric}|$, where the infimum and supremums are taken over the whole of space-time. Then*

$$\frac{2}{3}\underline{R}(\tau_2^{\frac{3}{2}} - \tau_1^{\frac{3}{2}}) + \frac{e^{-2R_1(\tau_2-\tau_1)}}{2(\sqrt{\tau_2}-\sqrt{\tau_1})}W_2^2(\nu_1,\nu_2,\tau_1) \leqq V(\nu_1,\tau_1;\nu_2,\tau_2) \tag{B.1}$$

and

$$V(\nu_1,\tau_1;\nu_2,\tau_2) \leqq \frac{2}{3}\bar{R}(\tau_2^{\frac{3}{2}} - \tau_1^{\frac{3}{2}}) + \frac{e^{2R_1(\tau_2-\tau_1)}}{2(\sqrt{\tau_2}-\sqrt{\tau_1})}W_2^2(\nu_1,\nu_2,\tau_1). \tag{B.2}$$

To prove Lemma B.1, one needs merely to integrate the inequalities of the following proposition with respect to the optimal transference plans π associated to the $\mathscr{L}$-Wasserstein distance and Wasserstein distance respectively.

Proposition B.2. *If $\gamma : [\tau_1, \tau_2] \to \mathscr{M}$ is a minimising $\mathscr{L}$-geodesic such that*

$$\gamma(\tau_1) = x \in \mathscr{M} \quad \textit{and} \quad \gamma(\tau_2) = y \in \mathscr{M},$$

and such that $\underline{R} \leqq R\big(\gamma(\tau),\tau\big) \leqq \bar{R}$ for $\tau \in [\tau_1,\tau_2]$, then

$$\frac{2}{3}\underline{R}(\tau_2^{\frac{3}{2}} - \tau_1^{\frac{3}{2}}) + \frac{e^{-2R_1(\tau_2-\tau_1)}}{2(\sqrt{\tau_2}-\sqrt{\tau_1})}d^2(x,y,\tau_1) \leqq Q(x,\tau_1;y,\tau_2) \tag{B.3}$$

and

$$Q(x,\tau_1;y,\tau_2) \leqq \frac{2}{3}\bar{R}(\tau_2^{\frac{3}{2}} - \tau_1^{\frac{3}{2}}) + \frac{e^{2R_1(\tau_2-\tau_1)}}{2(\sqrt{\tau_2}-\sqrt{\tau_1})} d^2(x,y,\tau_1). \tag{B.4}$$

In turn, the estimates of Proposition B.2 follow by elementary consideration of (respectively) the $\mathscr{L}$-lengths of a minimising $\mathscr{L}$-geodesic $\gamma : [\tau_1, \tau_2] \to \mathscr{M}$ from x to y, and a minimising Riemannian geodesic from x to y on $(\mathscr{M}, g(\tau_1))$ (parametrised with respect to $\sigma := \sqrt{\tau}$). For similar considerations, see [5], Lemma 7.13. One should keep in mind that the bound $|\mathrm{Ric}| \leqq R_1$ on a Ricci flow constrains the length of vectors—and hence the distance between two points—to grow/shrink at most exponentially [16], Lemma 5.3.2. This control on the evolution of distances also implies that

$$W_2(v_1, v_2, \tau_\alpha) \leqq e^{R_1|\tau_\beta - \tau_\alpha|} W_2(v_1, v_2, \tau_\beta).$$

This combines with Lemma B.1 to give the following corollary.

Corollary B.3. *If $v_1(\tau)$ and $v_2(\tau)$ are continuous families of Borel probability measures* (*with respect to $W_2(\cdot,\cdot,\tau_0)$*) *for τ in a neighbourhood of $\tau_0 \in (\tau_1, \tau_2)$, and $\tau_1(s)$, $\tau_2(s)$ are continuous functions of a real variable s such that $\tau_1(s) \to \tau_0$ and $\tau_2(s) \to \tau_0$ as $s \to 0$, then*

$$2\big(\sqrt{\tau_2}(s) - \sqrt{\tau_1}(s)\big) V\big(v_1(\tau_1(s)), \tau_1(s); v_2(\tau_2(s)), \tau_2(s)\big) \to W_2^2\big(v_1(\tau_0), v_2(\tau_0), \tau_0\big)$$

as $s \to 0$.

One of the applications of Proposition B.2 is to relate the $\mathscr{L}$-Wasserstein distance to the infinitesimal $\mathscr{L}$-Wasserstein distance, or $\mathscr{L}$-Wasserstein speed, as defined in Section 1.2, by proving Lemma 1.3. The quickest approach to this is to exploit the following well-known analogue of that lemma for the Wasserstein distance W_2, which follows on from ideas implicit in the work of Benamou-Brenier [2] and the heuristics of Otto-Villani [12], §3. We follow most closely Otto's argument described in [17], §7.6. The Riemannian metric is fixed in the following lemma, so we drop the third parameter τ for W_2 and d.

Lemma B.4. *Suppose that $(\mathscr{M}, g)$ is a closed Riemannian manifold, and $v(t)$ is a smooth family of positive probability measures on $\mathscr{M}$* (*i.e. its density with respect to Riemannian volume measure is smooth and positive*) *for t in a neighbourhood of* 0. *Then*

$$W_2^2\big(v(0), v(t)\big) = t^2 \inf_X \int_{\mathscr{M}} |X|^2 \, dv(0) + o(t^2), \tag{B.5}$$

where the infimum here and later in this section is taken over all advection fields for $v(t)$ at $t = 0$, as defined in Section 1.

Proof. For *each t*, let $X(t)$ be the minimising advection field for $v(\cdot)$ at time t, and write $X = X(0)$. Write $dv(t) = u(t)\, d\mu$ where μ is Riemannian volume measure. Then $X = \nabla v$ where v solves $-\mathrm{div}(u\nabla v) = \frac{\partial u}{\partial t}$ (cf. Section 1).

We first prove that the left-hand side of (B.5) is less than the right-hand side. Let $\psi_t : \mathscr{M} \to \mathscr{M}$ be the family of diffeomorphisms generated by $X(t)$, with ψ_0 the identity. We write π_0 for the push forward of $\nu(0)$ under the diagonal map $x \to (x, x)$, which we view as the optimal transference plan between $\nu(0)$ and $\nu(t)$ at $t = 0$. Taking this viewpoint, we may take as a candidate transference plan at nearby t, the measure π_t obtained by pushing forward $\nu(0)$ under the map $x \to \big(x, \psi_t(x)\big)$. We may then estimate

$$\text{(B.6)} \qquad W_2^2\big(\nu(0), \nu(t)\big) \leqq \int_{\mathscr{M}\times\mathscr{M}} d^2(x, y)\, d\pi_t(x, y) = \int_{\mathscr{M}\times\mathscr{M}} d^2\big(x, \psi_t(x)\big)\, d\nu(0)(x)$$

by definition of push-forward measures. By definition of X, we have

$$d\big(x, \psi_t(x)\big) = t|X(0)|(x) + o(t),$$

uniformly in x, so

$$\text{(B.7)} \qquad W_2^2\big(\nu(0), \nu(t)\big) \leqq t^2 \int_{\mathscr{M}} |X|^2\, d\nu(0) + o(t^2),$$

as desired. To prove the opposite inequality, we compute at $t = 0$ that

$$\text{(B.8)} \qquad \int_{\mathscr{M}} |X|^2\, d\nu(0) = \int_{\mathscr{M}} |\nabla v|^2 u\, d\mu = -\int_{\mathscr{M}} v \operatorname{div}(u\nabla v)\, d\mu.$$

But $-\operatorname{div}\big(u(0)\nabla v\big) = \dfrac{\partial u}{\partial t}(0) = \dfrac{u(t) - u(0)}{t} + o(1)$ uniformly over $\mathscr{M}$, as $t \to 0$, so

$$\text{(B.9)} \qquad \int_{\mathscr{M}} |X|^2\, d\nu(0) + o(1) = \int_{\mathscr{M}} v \frac{\big(d\nu(t) - d\nu(0)\big)}{t} = \frac{1}{t} \int_{\mathscr{M}\times\mathscr{M}} [v(y) - v(x)]\, d\hat{\pi}_t(x, y),$$

where $\hat{\pi}_t$ is the *optimal* transference plan from $\nu(0)$ to $\nu(t)$. By smoothness of v, and compactness of $\mathscr{M}$, we know that $|v(x) - v(y)| \leqq |\nabla v|(x) d(x, y) + C d^2(x, y)$, and so keeping in mind the Cauchy-Schwarz inequality,

$$\begin{aligned}
\text{(B.10)} \qquad |t| \int_{\mathscr{M}} |X|^2\, d\nu(0) + o(t) &\leqq \int_{\mathscr{M}\times\mathscr{M}} [|\nabla v|(x) d(x, y) + C d^2(x, y)]\, d\hat{\pi}_t(x, y) \\
&\leqq \Bigg(\int_{\mathscr{M}\times\mathscr{M}} |\nabla v|^2(x)\, d\hat{\pi}_t(x, y)\Bigg)^{\frac{1}{2}} \Bigg(\int_{\mathscr{M}\times\mathscr{M}} d^2(x, y)\, d\hat{\pi}_t(x, y)\Bigg)^{\frac{1}{2}} \\
&\quad + C \int_{\mathscr{M}\times\mathscr{M}} d^2(x, y)\, d\hat{\pi}_t(x, y) \\
&= \Bigg(\int_{\mathscr{M}} |\nabla v|^2\, d\nu(0)\Bigg)^{\frac{1}{2}} W_2\big(\nu(0), \nu(t)\big) + C W_2^2\big(\nu(0), \nu(t)\big).
\end{aligned}$$

The conclusion (B.7) of the first part of the lemma tells us that the very last term of (B.10) is $O(t^2)$. Therefore, by the first equality of (B.8),

$$\text{(B.11)} \qquad |t| \int_{\mathscr{M}} |X|^2\, d\nu(0) \leqq \Bigg(\int_{\mathscr{M}} |X|^2\, d\nu(0)\Bigg)^{\frac{1}{2}} W_2\big(\nu(0), \nu(t)\big) + o(t)$$

or equivalently

$$t^2 \int_{\mathscr{M}} |X|^2 \, dv(0) \leqq W_2^2\big(v(0), v(t)\big) + o(t^2). \qquad \square \tag{B.12}$$

Lemma B.4 puts us in a position to prove Lemma 1.3 from Section 1.

Proof of Lemma 1.3. We prove the lemma in two steps, first proving that the left-hand side of (1.7) is less than the right-hand side, just as in the proof of Lemma B.4. Analogously to that proof, we consider the optimal advection field $X(\tau)$ for $v(\cdot)$ at time τ, and the diffeomorphisms $\psi_\tau : \mathscr{M} \to \mathscr{M}$ it generates (with ψ_{τ_1} the identity) and compute that

$$\begin{aligned} V\big(v(\tau_1), \tau_1; v(\tau_2), \tau_2\big) &\leqq \int_{\mathscr{M}} Q\big(x, \tau_1; \psi_{\tau_2}(x), \tau_2\big) \, dv(\tau_1)(x) \\ &\leqq \int_{\mathscr{M}} \mathscr{L}(\gamma^{x,\tau_1,\tau_2}) \, dv(\tau_1)(x), \end{aligned} \tag{B.13}$$

where $\gamma^{x,\tau_1,\tau_2} : [\tau_1, \tau_2] \to \mathscr{M}$ is the integral curve of $X(\tau)$ starting at $x \in \mathscr{M}$. That is, $\gamma^{x,\tau_1,\tau_2}(\tau) = \psi_\tau(x)$. Working directly from the definition (1.2) of $\mathscr{L}$, we then find that

$$\begin{aligned} V\big(v(\tau_1), \tau_1; v(\tau_2), \tau_2\big) \leqq (\tau_2 - \tau_1)\left(\sqrt{\tau_1} \int_{\mathscr{M}} \big(R(\cdot, \tau_1) + |X|^2_{g(\tau_1)}\big) \, dv(\tau_1)\right) \\ + o(\tau_2 - \tau_1), \end{aligned} \tag{B.14}$$

as $\tau_2 \downarrow \tau_1$. To get the reverse direction, controlling V from below, we must work a little harder. Fix $\xi > 0$, and choose $r = r(\xi) > 0$ sufficiently small so that $R(y, \tau_2) \geqq R(x, \tau_1) - \xi$ whenever $d(x, y, \tau_1) < r$ and $\tau_2 \in (\tau_1, \tau_1 + r^2)$. From now on in this proof, we only consider values of τ_2 in this range. Define the set

$$\begin{aligned} \Sigma(\tau_2) = \{(x, y) \in \mathscr{M} \times \mathscr{M} \mid d(x, y, \tau_1) < r \text{ and } \exists \text{ minimising } \mathscr{L}\text{-geodesic} \\ \text{from } (x, \tau_1) \text{ to } (y, \tau_2) \text{ remaining in } B_{g(\tau_1)}(x, r)\}, \end{aligned} \tag{B.15}$$

(also depending on our fixed τ_1 and r). For $(x, y) \in \Sigma(\tau_2)$, by Proposition B.2,

$$Q(x, \tau_1; y, \tau_2) \geqq \frac{2}{3}\big(\tau_2^{\frac{3}{2}} - \tau_1^{\frac{3}{2}}\big)\big(R(x, \tau_1) - \xi\big) + \frac{e^{-2R_1(\tau_2 - \tau_1)}}{2(\sqrt{\tau_2} - \sqrt{\tau_1})} d^2(x, y, \tau_1), \tag{B.16}$$

which we may integrate with respect to the optimal transference plan π_{τ_2} (optimal between $v(\tau_1)$ and $v(\tau_2)$ for $\mathscr{L}$-optimal transportation (1.4)) to give

$$\begin{aligned} \int_{\Sigma(\tau_2)} Q(x, \tau_1; y, \tau_2) \, d\pi_{\tau_2}(x, y) \geqq \sqrt{\tau_1}(\tau_2 - \tau_1) \int_{\Sigma(\tau_2)} \big(R(x, \tau_1) - \xi\big) \, d\pi_{\tau_2}(x, y) \\ + \frac{e^{-2R_1(\tau_2 - \tau_1)}}{2(\sqrt{\tau_2} - \sqrt{\tau_1})} \int_{\Sigma(\tau_2)} d^2(x, y, \tau_1) \, d\pi_{\tau_2}(x, y) \\ + o(\tau_2 - \tau_1). \end{aligned} \tag{B.17}$$

We will need the full strength of this estimate, but also the weaker consequence that the left-hand side cannot be too negative:

$$\int_{\Sigma(\tau_2)} Q(x,\tau_1;y,\tau_2)\,d\pi_{\tau_2}(x,y) \geqq O(\tau_2-\tau_1). \tag{B.18}$$

Meanwhile, we need to consider the integral of Q over the *complement* of $\Sigma(\tau_2)$. For $(x,y)\in(\mathscr{M}\times\mathscr{M})\backslash\Sigma(\tau_2)$, a slight variation on the argument for the first inequality (B.3) of Proposition B.2 tells us (by definition of $\Sigma(\tau_2)$) that for τ_2 sufficiently close to τ_1 (depending on r and the curvature of the Ricci flow),

$$Q(x,\tau_1;y,\tau_2) \geqq \frac{r^2}{2(\tau_2-\tau_1)}. \tag{B.19}$$

Integrating this, we find that

$$\pi_{\tau_2}\big((\mathscr{M}\times\mathscr{M})\backslash\Sigma(\tau_2)\big) \leqq \frac{2(\tau_2-\tau_1)}{r^2}\int_{(\mathscr{M}\times\mathscr{M})\backslash\Sigma(\tau_2)} Q(x,\tau_1;y,\tau_2)\,d\pi_{\tau_2}(x,y). \tag{B.20}$$

On the other hand, the integral on the right-hand side can be viewed as the difference of the integral over $\mathscr{M}\times\mathscr{M}$ and the integral over $\Sigma(\tau_2)$, and both of these terms are controlled from above by $O(\tau_2-\tau_1)$ thanks to (B.14) and (B.18) respectively. Therefore

$$\pi_{\tau_2}\big((\mathscr{M}\times\mathscr{M})\backslash\Sigma(\tau_2)\big) \leqq O\big((\tau_2-\tau_1)^2\big). \tag{B.21}$$

The benefit of such an estimate is that when we integrate Q over $(\mathscr{M}\times\mathscr{M})\backslash\Sigma(\tau_2)$, terms involving scalar curvature become negligible and can be removed and added at will. In particular, integrating (B.3) of Proposition B.2, we can see that

$$\begin{aligned}
&\int_{(\mathscr{M}\times\mathscr{M})\backslash\Sigma(\tau_2)} Q(x,\tau_1;y,\tau_2)\,d\pi_{\tau_2}(x,y) \\
&\quad\geqq O\big((\tau_2-\tau_1)^2\big) + \frac{e^{-2R_1(\tau_2-\tau_1)}}{2(\sqrt{\tau_2}-\sqrt{\tau_1})}\int_{(\mathscr{M}\times\mathscr{M})\backslash\Sigma(\tau_2)} d^2(x,y,\tau_1)\,d\pi_{\tau_2}(x,y) \\
&\quad\geqq \sqrt{\tau_1}(\tau_2-\tau_1)\int_{(\mathscr{M}\times\mathscr{M})\backslash\Sigma(\tau_2)} \big(R(x,\tau_1)-\xi\big)\,d\pi_{\tau_2}(x,y) \\
&\qquad + \frac{e^{-2R_1(\tau_2-\tau_1)}}{2(\sqrt{\tau_2}-\sqrt{\tau_1})}\int_{(\mathscr{M}\times\mathscr{M})\backslash\Sigma(\tau_2)} d^2(x,y,\tau_1)\,d\pi_{\tau_2}(x,y) + o(\tau_2-\tau_1).
\end{aligned} \tag{B.22}$$

We can now add this to the analogous inequality (B.17) for $\Sigma(\tau_2)$ to give

$$\begin{aligned}
V\big(v(\tau_1),\tau_1;v(\tau_2),\tau_2\big) &\geqq \sqrt{\tau_1}(\tau_2-\tau_1)\int_{\mathscr{M}\times\mathscr{M}} \big(R(x,\tau_1)-\xi\big)\,d\pi_{\tau_2}(x,y) \\
&\quad + \frac{e^{-2R_1(\tau_2-\tau_1)}}{2(\sqrt{\tau_2}-\sqrt{\tau_1})}W_2^2\big(v(\tau_1),v(\tau_2),\tau_1\big) + o(\tau_2-\tau_1),
\end{aligned} \tag{B.23}$$

and by Lemma B.4 (and the definition of push-forward measures) this reduces to

$$\text{(B.24)} \qquad V\big(\nu(\tau_1), \tau_1; \nu(\tau_2), \tau_2\big) \geqq \sqrt{\tau_1}(\tau_2 - \tau_1)\left(\int_{\mathscr{M}} \big(R(\cdot, \tau_1) - \xi\big)\, d\nu(\tau_1) + \int_{\mathscr{M}} |X|^2\, d\nu(\tau_1)\right) + o(\tau_2 - \tau_1).$$

Because $\xi > 0$ was arbitrary, this improves to

$$\text{(B.25)} \qquad V\big(\nu(\tau_1), \tau_1; \nu(\tau_2), \tau_2\big) \geqq (\tau_2 - \tau_1)\left(\sqrt{\tau_1}\int_{\mathscr{M}} \big(R(\cdot, \tau_1) + |X|^2\big)\, d\nu(\tau_1)\right) + o(\tau_2 - \tau_1),$$

which combines with (B.14) to conclude the proof. □

References

[1] *V. Bangert*, Analytische Eigenschaften konvexer Funktionen auf Riemannschen Mannigfaltigkeiten, J. reine angew. Math. **307** (1979), 309–324.
[2] *J. D. Benamou* and *Y. Brenier*, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math. **84** (2000), 375–393.
[3] *P. Bernard* and *B. Buffoni*, Optimal mass transportation and Mather theory, J. Eur. Math. Soc. **9** (2007), 85–121.
[4] *I. Chavel*, Riemannian geometry: a modern introduction, Cambridge tracts math. **108** (1995).
[5] *B. Chow*, *S.-C. Chu*, *D. Glickenstein*, *C. Guenther*, *J. Isenberg*, *T. Ivey*, *D. Knopf*, *P. Lu*, *F. Luo* and *L. Ni*, Ricci flow: Techniques and Applications: Part I: Geometric aspects, Math. Surv. Monogr. **135** (2007).
[6] *D. Cordero-Erausquin*, *R. J. McCann* and *M. Schmuckenschläger*, A Riemannian interpolation inequality à la Borel, Brascamp and Lieb, Invent. Math. **146** (2001), 219–257.
[7] *D. Cordero-Erausquin*, *R. J. McCann* and *M. Schmuckenschläger*, Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport, Ann. Fac. Sci. Toulouse Math. (6) **15** (2006), 613–635.
[8] *R. S. Hamilton*, Three-manifolds with positive Ricci curvature, J. Diff. Geom. **17** (1982), 255–306.
[9] *R. S. Hamilton*, The Harnack estimate for the Ricci flow, J. Diff. Geom. **37** (1993), 225–243.
[10] *R. J. McCann*, Polar factorization of maps on Riemannian manifolds, Geom. Funct. Anal. **11** (2001), 589–608.
[11] *R. J. McCann* and *P. M. Topping*, Ricci flow, entropy and optimal transportation, Amer. J. Math., to appear.
[12] *F. Otto* and *C. Villani*, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal. **173** (2000), 361–400.
[13] *G. Perelman*, The entropy formula for the Ricci flow and its geometric applications, http://arXiv.org/math.DG/0211159v1, 2002.
[14] *G. Perelman*, Ricci flow with surgery on three-manifolds, http://arXiv.org/math.DG/0303109v1, 2003.
[15] *K.-T. Sturm* and *M.-K. von Renesse*, Transport inequalities, gradient estimates, entropy and Ricci curvature, Comm. Pure Appl. Math. **58** (2005), 923–940.
[16] *P. M. Topping*, Lectures on the Ricci flow, L.M.S. Lect. notes ser. **325** (2006).
[17] *C. Villani*, Topics in optimal transportation, Grad. Stud. Math. **58** (2003).
[18] *C. Villani*, Optimal transport, old and new (Saint-Flour 2005), Version of October 2007.
[19] *R. Ye*, On the *l*-function and the reduced volume of Perelman, preprint 2004.

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

Eingegangen 22. November 2007