RICCI FLOW, ENTROPY AND OPTIMAL TRANSPORTATION*

ROBERT J. MCCANN† AND PETER M. TOPPING‡

Abstract. Let a smooth family of Riemannian metrics $g(\tau)$ satisfy the backwards Ricci flow equation on a compact oriented n-dimensional manifold M. Suppose two families of normalized n-forms $\omega(\tau) \geq 0$ and $\tilde{\omega}(\tau) \geq 0$ satisfy the forwards (in τ) heat equation on M generated by the connection Laplacian $\Delta_{g(\tau)}$. If these n-forms represent two evolving distributions of particles over M, the minimum root-mean-square distance $W_2(\omega(\tau), \tilde{\omega}(\tau), \tau)$ to transport the particles of $\omega(\tau)$ onto those of $\tilde{\omega}(\tau)$ is shown to be non-increasing as a function of τ, without sign conditions on the curvature of $(M, g(\tau))$. Moreover, this contractivity property is shown to characterise supersolutions to the Ricci flow.

1. Introduction

On a compact oriented n-dimensional manifold M, let $g(\tau)$ be a smooth family of metrics for $\tau \in [\tau_1, \tau_2]$. We are particularly interested in the case that $g(\tau)$ satisfies the backwards Ricci flow equation

$$\frac{\partial g}{\partial \tau} = 2\,\mathrm{Ric}(g) \tag{1}$$

where $\mathrm{Ric}(g)$ is the Ricci tensor of g. Given terminal data $g(\tau_2)$, such a family can always be constructed for τ_1 sufficiently close to τ_2 (see Hamilton [12], DeTurck [10], [29, Ch. 5]). The geodesic distance $d(\mathbf{x}, \mathbf{y}, \tau)$ between two points $\mathbf{x}, \mathbf{y} \in M$, with respect to $g(\tau)$, evolves according to the formula

$$d^2(\mathbf{x}, \mathbf{y}, \tau) = \inf_{\sigma(0)=\mathbf{x}, \sigma(1)=\mathbf{y}} \int_0^1 \left| \frac{d\sigma}{ds} \right|^2_{g(\tau)} ds, \tag{2}$$

where the infimum is taken over smooth curves $\sigma : [0,1] \to M$ joining $\mathbf{x}$ to $\mathbf{y}$. Similarly, given two Borel probability measures ν and $\tilde{\nu}$ on M, the 2-Wasserstein distance $W_2(\nu, \tilde{\nu}, \tau)$ between them evolves according to its definition

$$W_2^2(\nu, \tilde{\nu}, \tau) = \inf_{\pi \in \Gamma(\nu, \tilde{\nu})} \int_{M \times M} d^2(\mathbf{x}, \mathbf{y}, \tau)\, d\pi(\mathbf{x}, \mathbf{y}). \tag{3}$$

Date: November 12, 2008.

*This paper supercedes the earlier paper *Diffusion is a 2-Wasserstein contraction on any manifold evolving by reverse Ricci flow (2006)*.

The infimum is taken over the space $\Gamma(\nu,\tilde{\nu})$ of Borel probability measures π on $M \times M$ which have marginals ν and $\tilde{\nu}$, in the sense that

$$\begin{aligned}(4)\qquad \int_M f(\mathbf{x})d\nu(\mathbf{x}) = \int_{M\times M} f(\mathbf{x})\, d\pi(\mathbf{x},\mathbf{y}); &\qquad \text{and}\\ \int_{M\times M} f(\mathbf{y})\, d\pi(\mathbf{x},\mathbf{y}) \;\; = \int_M f(\mathbf{y})d\tilde{\nu}(\mathbf{y}),&\end{aligned}$$

for each continuous test function $f \in C(M)$.

In this paper, we are particularly interested in the case of measures ν and $\tilde{\nu}$ which are induced by n-forms ω and $\tilde{\omega}$ respectively, in the sense that

$$\nu(A) = \int_A \omega,$$

for every Borel $A \subset M$, and similarly for $\tilde{\omega}$. (We will often corrupt notation by considering the Wasserstein distance between ω and $\tilde{\omega}$ rather than ν and $\tilde{\nu}$.) The advantage of defining W_2 as an infimum over joint probabilities π rather than smooth $2n$-forms on $M \times M$ is that $\Gamma(\nu,\tilde{\nu})$ is a weak-$*$ compact subset of the dual space to $(C(M\times M), \|\cdot\|_\infty)$, so the infimum in (3) is therefore attained by some joint probability π_0. The structure of the minimising π_0 will be recalled in the proofs below; it is not generally smooth.

Following a construction from Perelman's work on Ricci flow [23], [29, Chapter 6], let $\omega(\mathbf{x},\tau)$ evolve under the heat equation

$$(5)\qquad \frac{\partial \omega}{\partial \tau} = \Delta_{g(\tau)}\omega,$$

where Δ_g is the connection Laplacian with respect to g. This evolution preserves the total mass:

$$\frac{d}{d\tau}\int_M \omega = 0,$$

and gives a smooth n-form $\omega(\tau)$ at later times τ. In particular, the measures induced by $\omega(\tau)$ at later times continue to be probability measures, absolutely continuous with respect to the measure induced by any smooth volume form on M. If we write $\omega(\mathbf{x},\tau) = u(\mathbf{x},\tau)dV$, where $dV = dV_{g(\tau)}$ is the volume form associated to $g(\tau)$, then the non-negative function u solves the conjugate heat equation

$$(6)\qquad \frac{\partial u}{\partial \tau} = \Delta_{g(\tau)}u - \left(\frac{1}{2}\,\mathrm{tr}\,\frac{\partial g}{\partial \tau}\right)u,$$

which in the special case of Ricci flow is

$$(7)\qquad \frac{\partial u}{\partial \tau} = \Delta_{g(\tau)}u - Ru,$$

where $R = \operatorname{tr}\operatorname{Ric}$ is the scalar curvature, since the volume form dV evolves according to $\frac{\partial}{\partial\tau}dV = (\frac{1}{2}\operatorname{tr}\frac{\partial g}{\partial\tau})\,dV$ or $\frac{\partial}{\partial\tau}dV = R\,dV$ in the special case of Ricci flow (see [29, (2.5.7)]). By the strong maximum principle, $u > 0$ for $\tau > \tau_1$.

We precede our main theorem by one of its corollaries, which asserts that the diffusion (5) of the form $\omega(\tau)$ couples with the backwards Ricci flow to produce a 2-Wasserstein contraction:

Corollary 1 (Coupled contraction)**.** *On a compact oriented manifold M, suppose a smooth family of metrics $g(\tau)$ satisfies the backwards Ricci flow equation (1) on the same interval $[\tau_1, \tau_2] \subset \mathbf{R}$ that $\omega(\mathbf{x}, \tau) \geq 0$ and $\tilde{\omega}(\mathbf{x}, \tau) \geq 0$ are unit mass solutions to the diffusion equation (5). Then $W_2(\omega(\tau), \tilde{\omega}(\tau), \tau)$ is a non-increasing function of $\tau \in [\tau_1, \tau_2]$, where 2-Wasserstein distance W_2 is defined by (3).*

This result should be compared to 2-Wasserstein contractivity of the ordinary heat flow in a stationary metric, which can be established assuming $\operatorname{Ric} \geq 0$: see e.g. Sturm & von Renesse [27], and the subsequent works of Lott & Villani [16] [17] and Sturm [24] [25] [26], which build on the Riemannian adaptation by Otto & Villani [22] and Cordero-Erausquin, McCann & Schmuckenschläger [8] [9], of Jordan, Kinderlehrer & Otto's gradient flow formulation of the dynamics [14] [21] from Euclidean space and McCann's displacement convexity [19]. In the Euclidean context, W_2-contractivity of the heat evolution was also established by Ambrosio, Gigli & Savaré [1] and Carrillo, McCann & Villani [6]. The connection between entropy, Ricci curvature, and convergence of diffusion dates back at least to Bakry & Emery [2].

We remark that in our Ricci flow setting, no sign condition on the Ricci curvature is required. In a region where this curvature is negative, the evolution of the metric (1) shrinks distances just enough to compensate for any lack of contractivity of the diffusion, whereas in Ricci positive regions, the diffusive contraction turns out to be strong enough to compensate for expansion of distances by (1).

The part of the proof of our main theorem which leads to Corollary 1 will be based on displacement semiconvexity and other estimates for the Boltzmann-Shannon entropy along appropriate Wasserstein geodesics (see Section 3).

To state our main theorem, we need to introduce the notion of a supersolution to the Ricci flow.

Definition 1. *A super Ricci flow (parametrised backwards in time) is a smooth family $g(\tau)$ of metrics, $\tau \in [\tau_1, \tau_2]$, such that at each $\tau \in (\tau_1, \tau_2)$, and at each point on M, we have*

$$-\frac{\partial g}{\partial \tau} + 2\operatorname{Ric}(g(\tau)) \geq 0. \tag{8}$$

Note that τ is reverse-time compared to the time parameter t in the classical Ricci flow literature, and so $-\frac{\partial}{\partial\tau}$ is a derivative forwards in time t.

Our main theorem asserts that the contractivity of diffusions backwards in t (forwards in τ) as in Corollary 1 characterises super Ricci flows, and there is a third equivalent condition involving *forwards in* t (backwards in τ) solutions to heat equations.

Theorem 2. *Suppose that M is a compact, oriented manifold equipped with a smooth family of metrics $g(\tau)$ for $\tau \in [\tau_1, \tau_2] \subset \mathbf{R}$. Then the following are equivalent:*

(A) $g(\tau)$ is a super Ricci flow (i.e. satisfies (8)*);*

(B) whenever $\tau_1 < a < b < \tau_2$ and $\omega(\mathbf{x}, \tau) \geq 0$ and $\tilde{\omega}(\mathbf{x}, \tau) \geq 0$ are unit mass solutions to the diffusion equation (5) for $\tau \in (a, b)$, the function $W_2(\omega(\tau), \tilde{\omega}(\tau), \tau)$ is nonincreasing in $\tau \in (a, b)$, where 2*-Wasserstein distance W_2 is defined by (3).*

(C) whenever $\tau_1 < a < b < \tau_2$ and $f : M \times (a, b) \to \mathbf{R}$ is a solution to $-\frac{\partial f}{\partial\tau} = \Delta_{g(\tau)} f$, the function $\sup_M |\nabla f(\cdot, \tau)|$ *is nondecreasing in τ.*

This theorem is related to a result of Sturm and von Renesse [27] showing that fixed Riemannian manifolds with $\mathrm{Ric} \geq 0$ can be characterised in terms of the properties of the solutions of heat equations. In our situation – working with respect to an evolving metric – one must distinguish between forwards in t (backwards in τ) solutions to the heat equation $\frac{\partial f}{\partial t} = \Delta f$, which do not have preserved mass in our situation, and backwards in t (forwards in τ) solutions to the diffusion equation $\frac{\partial u}{\partial\tau} = \Delta u - \left(\frac{1}{2}\operatorname{tr}\frac{\partial g}{\partial\tau}\right) u$ which do have preserved mass.

Remark 3. *Our characterisations indicate how one can define a super Ricci flow in certain weaker contexts than having a smooth family of Riemannian manifolds. For example, one could consider one-parameter families of path metric spaces, each equipped with a reference measure such as the Hausdorff measure of non-trivial dimension induced by its metric. Using the ideas of entropy convexity for Ricci flow in this paper, it is possible to make sense of weak super Ricci flow definitions without constructing any notion of diffusion. This provides a dynamic analogue of the approach of the previously mentioned papers of Lott & Villani* [16] [17] *and Sturm* [24] [25] [26]*, which address the static case. A weak Ricci flow can then be defined to be a weak super Ricci flow which at each time expands distances no faster (to first order in time) than any other super Ricci flow which coincides with the given super Ricci flow at that time.*

Remark 4. *The orientability assumption in Theorem 2 and Corollary 1 is only required to make sense of the inequalities $\omega \geq 0$ (meaning that the form ω is a nonnegative multiple of the volume form $dV_{g(\tau)}$) and $\tilde{\omega} \geq 0$. By reformulating the theorem in terms of the measures induced by ω and $\tilde{\omega}$, one gets a result* *which is also true for non-orientable manifolds (albeit at the expense of clarity of exposition). The same is true for Lemma 8 below.*

Acknowledgement: We would like to thank Gerhard Huisken and Alessio Figalli for useful comments during the course of this work. John Lott has informed us that he has independently been considering the links between optimal transportation and Ricci flow, and has formally derived a result similar to our Corollary 1 [15]. We are also grateful to him for pointing out related developments due to Carfora [5], and for comments on this work. This research was supported in part by Natural Sciences and Engineering Research Council of Canada Grant 217006-03, United States National Science Foundation Grant DMS 0354729, an Engineering and Physical Sciences Research Council (UK) Advanced Research Fellowship, and the Leverhulme Trust. RJM is pleased to acknowledge the hospitality of the Institut Henri Poincaré (Paris) and PT thanks the Max Planck Albert Einstein Institute (Golm) and the Free University (Berlin) where parts of this work were performed.

Added in proof: Since this work appeared in preprint form in 2006, there have been several developments which we briefly survey. In [30] from 2007, a notion of $\mathcal{L}$-optimal transportation was introduced and a contractivity result was proved that generalises the 2-Wasserstein contractivity on Ricci flows in this paper. That viewpoint also allows one to recover essentially all of the monotonic quantities that Perelman introduced in [23] to study finite-time singularities of Ricci flow – see [30] and the subsequent paper of Lott [18] from 2008 where Perelman's *reduced volume* was shown also to arise this way. That latter paper [18] also includes a new rigorous proof of the 2-Wasserstein contractivity on Ricci flows that makes up part of this paper. Finally it turns out to be fruitful to extend the results of this paper to 1-Wasserstein contractivity. The extension to this situation was made by Tom Ilmanen, and in [4] a new space-time Ricci soliton construction was made which was inspired by an attempt to reconcile this 1-Wasserstein contractivity with the $\mathcal{L}$-Wasserstein contractivity results from [30].

2. Properties of the distance function

We consider now the distance function $d : M \times M \times [\tau_1, \tau_2] \to [0, \infty)$ associated to an arbitrary smooth family of Riemannian metrics $g(\tau)$ on M for $\tau \in [\tau_1, \tau_2]$.

Since we are working with a smooth flow on a compact manifold M, over a compact time interval, we may assume that the diameter is bounded (that is, $d(\cdot, \cdot, \cdot) \leq C$) and $|\frac{\partial g}{\partial \tau}|$ is bounded. This latter fact implies control on the rate that distances can expand or shrink (c.f. [29, Lemma 5.3.2]) and we can deduce that the distance function $\tau \mapsto d(\mathbf{x}, \mathbf{y}, \tau)$ is a Lipschitz function on $[\tau_1, \tau_2]$ with Lipschitz constant independent of $\mathbf{x}$ and $\mathbf{y}$.

If $K < \infty$ is an upper bound for the Lipschitz constant of $d^2(\mathbf{x}, \mathbf{y}, \cdot) : [\tau_1, \tau_2] \to \mathbf{R}$, for all $\mathbf{x}, \mathbf{y} \in M$, then we may work directly from the definition of 2-Wasserstein distance (3) to see that for fixed unit mass non-negative n-forms ω and $\tilde{\omega}$, the function $\tau \mapsto W_2^2(\omega, \tilde{\omega}, \tau)$ is Lipschitz, with Lipschitz constant no more than K. With the hindsight of the proof of Theorem 2, it will be clear that with $\omega(\tau)$ and $\tilde{\omega}(\tau)$ evolving smoothly, as in the theorem, the function $\tau \to W_2^2(\omega(\tau), \tilde{\omega}(\tau), \tau)$ is also Lipschitz away from $\tau = \tau_1$, but for now let us observe the more elementary fact that it is continuous on the whole of $[\tau_1, \tau_2]$.

Meanwhile, at each instant τ, the squared distance function $d^2(\mathbf{x}, \mathbf{y}, \tau)$ defined by (2) is smooth on an open subset of $M \times M$ whose complement (the *cut locus*) is denoted by $\mathrm{Cut}(M, g(\tau)) \subset M \times M$. A minimising geodesic links each pair of points $\mathbf{x}, \mathbf{y} \in M$ by completeness. For $(\mathbf{x}, \mathbf{y}) \notin \mathrm{Cut}(M, g(\tau))$, this geodesic is unique, and we represent it as a constant speed smooth map $s \in [0, 1] \longrightarrow \sigma(\mathbf{x}, \mathbf{y}, s, \tau) \in M$ with $\sigma(\mathbf{x}, \mathbf{y}, 0, \tau) = \mathbf{x}$ and $\sigma(\mathbf{x}, \mathbf{y}, 1, \tau) = \mathbf{y}$, which attains the infimum (2). When appropriate, we will abbreviate $\sigma(\mathbf{x}, \mathbf{y}, s, \tau)$ by $\sigma(s)$ and write $d\sigma/ds$ instead of $\partial\sigma/\partial s$.

We also need to consider the space-time cut locus

$$\mathrm{Cut}_{\mathrm{ST}} := \{(\mathbf{x}, \mathbf{y}, \tau) \in M \times M \times [\tau_1, \tau_2] \mid (\mathbf{x}, \mathbf{y}) \in \mathrm{Cut}(M, g(\tau))\}.$$

The following elementary properties of $\mathrm{Cut}_{\mathrm{ST}}$ will be proved in the appendix.

Lemma 5. *Suppose M is a compact manifold with a smooth family $g(\tau)$ of Riemannian metrics for $\tau \in (\tau_1, \tau_2)$. Then $\mathrm{Cut}_{\mathrm{ST}}$ is closed in $M \times M \times (\tau_1, \tau_2)$.*

Moreover, on the complement of $\mathrm{Cut}_{\mathrm{ST}}$, writing the unique constant speed minimising geodesic from $\mathbf{x}$ to $\mathbf{y}$, with respect to $g(\tau)$, as $s \in [0, 1] \longrightarrow \sigma(\mathbf{x}, \mathbf{y}, s, \tau) \in M$, the point $\sigma(\mathbf{x}, \mathbf{y}, s, \tau)$ is smoothly dependent on $\mathbf{x}$, $\mathbf{y}$, s and τ, and in particular, the squared distance function $d^2(\mathbf{x}, \mathbf{y}, \tau)$ is smoothly dependent on $\mathbf{x}$, $\mathbf{y}$ and τ.

Remark 6. *By translating time, let us assume that $0 \in (\tau_1, \tau_2)$. By virtue of the lemma, given $(\mathbf{x}, \mathbf{y}) \notin \mathrm{Cut}(M, g(0))$, and two smooth maps $X, Y : [\tau_1, \tau_2] \to M$ with $X(0) = \mathbf{x}$ and $Y(0) = \mathbf{y}$, we may precisely compute the evolution of $d^2(X(\tau), Y(\tau), \tau)$. One gets terms owing to the evolution of $X(\tau)$ and $Y(\tau)$, and of the metric $g(\tau)$:*

$$(9) \qquad \begin{aligned} \left.\frac{d}{d\tau}\right|_{\tau=0} \frac{d^2(X(\tau), Y(\tau), \tau)}{2} &= -\left\langle \left.\frac{dX}{d\tau}\right|_{\tau=0}, \left.\frac{d\sigma}{ds}\right|_{s=0^+} \right\rangle + \left\langle \left.\frac{dY}{d\tau}\right|_{\tau=0}, \left.\frac{d\sigma}{ds}\right|_{s=1^-} \right\rangle \\ &\quad + \int_0^1 \frac{1}{2} \frac{\partial g}{\partial \tau} \left(\frac{d\sigma}{ds}, \frac{d\sigma}{ds} \right) \bigg|_{\sigma(\mathbf{x}, \mathbf{y}, s, 0)} ds, \end{aligned}$$

where the shorthand $\sigma(s)$ refers to the specific geodesic $\sigma(\mathbf{x}, \mathbf{y}, s, 0)$.

3. Derivatives of the classical entropy along Wasserstein geodesics

In this section we consider Wasserstein geodesics on a fixed Riemannian manifold. We begin by recalling briefly the strategy for showing W_2-contractivity of the diffusion equation on a fixed, Ricci non-negative manifold. The central idea of the contractivity estimate [1] [6] [27] goes back to Jordan, Kinderlehrer & Otto's realization [14] that the heat equation represents steepest descent of the Boltzmann-Shannon entropy

$$E(u) = \int_M (\log u)\, u\, dV \tag{10}$$

with respect to 2-Wasserstein distance. In Euclidean space, (displacement) convexity of the entropy [19] along W_2-geodesics [3] allowed Otto [21] to quantify rates of convergence to the heat kernel. This displacement convexity extends to Ricci non-negative manifolds [8] as conjectured by Otto & Villani [22], and actually characterizes Ricci non-negativity as observed by Sturm & von Renesse [27]. On a manifold whose Ricci curvature takes both signs, the second derivative of the entropy (10) is estimated from below by a lower bound for the Ricci curvature [9] [27] — a fact used by Lott & Villani [16] [17] and Sturm [24] [25] [26] to develop a theory of Ricci bounds on measured length spaces. From the entropy, we shall require a more precise manifestation of displacement convexity (part of Lemma 8 below) to balance the possible metric expansion arising from (1). We derive this manifestation (14) following a Jacobi-field calculation of Cordero-Erausquin, McCann & Schmuckenschläger [9], in the spirit of classical comparison geometry, instead of their original proof [8]. This calculation explicitly links the behaviour of the entropy $E(u)$ along W_2-geodesics, to an appropriate average of the Ricci curvature along ordinary geodesics.

Definition 7 (Push-forward)**.** *Given manifolds M and $\hat{M}$, any Borel map $F : M \longrightarrow \hat{M}$ and probability measure ν on M induce a Borel probability measure on $\hat{M}$, called the push-forward of ν through F, denoted $F_\#\nu$ and defined by $(F_\#\nu)[V] = \nu[F^{-1}(V)]$ for all Borel $V \subset \hat{M}$. For Borel test functions $v : \hat{M} \longrightarrow \mathbf{R} \cup \{\pm\infty\}$, it follows that*

$$\int_{\hat{M}} v\, d(F_\#\mu) = \int_M (v \circ F)\, d\mu. \tag{11}$$

Since the lemma below applies to a manifold with a fixed metric g, rather than a flowing metric $g(\tau)$, we adapt our notation $\sigma(\mathbf{x}, \mathbf{y}, s, \tau)$, from Section 2, and our notation $W_2(\nu, \tilde{\nu}, \tau)$, by dropping the time argument τ.

Lemma 8 (Derivatives of the entropy along Wasserstein geodesics)**.** *Suppose M is a compact oriented manifold with a smooth metric g. Let $\omega > 0$ and $\tilde{\omega} > 0$*

be smooth n-forms on M with unit total mass, inducing probability measures ν and $\tilde{\nu}$. Let $\pi_0 \in \Gamma(\nu, \tilde{\nu})$ denote the minimising measure on $M \times M$ from the definition of $W_2(\nu, \tilde{\nu})$.

Then there exists a family of probability measures $\nu(s)$, for $s \in [0,1]$, with $\nu(0) = \nu$ and $\nu(1) = \tilde{\nu}$, such that

$$\frac{W_2(\nu, \nu(s))}{s} = W_2(\nu, \tilde{\nu}) = \frac{W_2(\nu(s), \tilde{\nu})}{1-s} \tag{12}$$

for each $s \in (0,1)$. For each $s \in [0,1]$, there exists a non-negative function $u(s) \in L^1(M)$ such that $\nu(s)$ is the measure induced by $u(s)\,dV_g$. The entropy $E(u(s))$ is semiconvex for $s \in [0,1]$, and a.e. $\tilde{s} \in (0,1)$ satisfies

$$\frac{d^2}{ds^2}\bigg|_{\tilde{s}} E(u(s)) := \lim_{\delta \to 0} \frac{E(u(\tilde{s}+\delta)) + E(u(\tilde{s}-\delta)) - 2E(u(\tilde{s}))}{\delta^2} \tag{13}$$

$$\geq \int_{M\times M} \operatorname{Ric}\left(\frac{d\sigma}{ds}, \frac{d\sigma}{ds}\right)\bigg|_{\sigma(\mathbf{x},\mathbf{y},\tilde{s})} d\pi_0(\mathbf{x},\mathbf{y}). \tag{14}$$

Moreover,

$$\frac{d}{ds}\bigg|_{s=0^+} E(u(s)) := \lim_{s \searrow 0} \frac{E(u(s)) - E(u(0))}{s}$$

$$\geq \int_{M\times M} \langle \frac{d\sigma}{ds}\bigg|_{\sigma(\mathbf{x},\mathbf{y},0^+)}, \nabla \log u(0)\bigg|_{\mathbf{x}} \rangle d\pi_0(\mathbf{x},\mathbf{y}). \tag{15}$$

By exchanging ω and $\tilde{\omega}$ in (15) (equivalently, by transforming s to $1-s$) we also have

$$-\frac{d}{ds}\bigg|_{s=1^-} E(u(s)) \geq \int_{M\times M} \langle -\frac{d\sigma}{ds}\bigg|_{\sigma(\mathbf{x},\mathbf{y},1^-)}, \nabla \log u(1)\bigg|_{\mathbf{y}} \rangle d\pi_0(\mathbf{x},\mathbf{y}),$$

and through (14), (15) and the semiconvexity of $E(u(s))$, the lemma yields what we will require in the proof of Theorem 2:

Corollary 9. *Suppose M is a compact oriented manifold with a smooth metric g. Let $\omega = u\,dV > 0$ and $\tilde{\omega} = \tilde{u}\,dV > 0$ be smooth n-forms on M with unit total mass. Let π_0 denote the minimising measure on $M \times M$ from the definition of $W_2(\omega, \tilde{\omega})$. Then*

$$\begin{aligned} &\int_{M\times M} \left(\langle \frac{d\sigma}{ds}\bigg|_{\sigma(\mathbf{x},\mathbf{y},1^-)}, \nabla \log \tilde{u}\bigg|_{\mathbf{y}} \rangle - \langle \frac{d\sigma}{ds}\bigg|_{\sigma(\mathbf{x},\mathbf{y},0^+)}, \nabla \log u\bigg|_{\mathbf{x}} \rangle \right) d\pi_0(\mathbf{x},\mathbf{y}) \\ &\quad \geq \int_0^1 \left(\int_{M\times M} \operatorname{Ric}\left(\frac{d\sigma}{ds}, \frac{d\sigma}{ds}\right)\bigg|_{\sigma(\mathbf{x},\mathbf{y},s)} d\pi_0(\mathbf{x},\mathbf{y}) \right) ds. \end{aligned} \tag{16}$$

Proof. (Lemma 8.) Before beginning the proof, we highlight a few implicit assertions within the statement of the lemma. First, we have defined π_0 as *the* minimiser; uniqueness here follows from [20] because ω and $\tilde{\omega}$ are smooth, and thus ν and $\tilde{\nu}$ do not charge sets of zero volume. Second, the semiconvexity of $E(u(s))$ and the smoothness of $u(0)$ and $u(1)$ tacitly imply that $u(s) \in L \log L$ for each $s \in [0,1]$ – that is, $E(u(s))$ is finite. Third, implicit in the integrals (14) and (15) is the existence of a geodesic $s \in [0,1] \longrightarrow \sigma(\mathbf{x},\mathbf{y},s)$ for π_0-almost all $(\mathbf{x},\mathbf{y}) \in M \times M$; this relies on a result of Cordero-Erausquin, McCann & Schmuckenschläger [8] which asserts that $\pi_0[\mathrm{Cut}(M,g)] = 0$. Fourth, the limits in (13) and (15) exist owing to the semiconvexity of $E(u(s))$.

Let us begin by recalling the basic facts about the minimiser of (3) established in [20]. The minimising joint measure $\pi_0 \in \Gamma(\nu,\tilde{\nu})$ is unique (as mentioned above) and can be expressed $\pi_0 = (id \times F)_{\#}\nu$ as the push-forward of ν through a Borel map $\mathbf{x} \longrightarrow (\mathbf{x}, F(\mathbf{x}))$. The map $F : M \longrightarrow M$ can be written $F(\mathbf{x}) = \exp_{\mathbf{x}} \nabla\theta$, for some potential $\theta : M \longrightarrow \mathbf{R}$ whose negation is $d^2/2$-concave, meaning $-\theta = ((-\theta)_{d^2/2})_{d^2/2}$, where the operation

$$\phi_{d^2/2}(\mathbf{y}) := \min_{\mathbf{x}\in M} d^2(\mathbf{x},\mathbf{y})/2 - \phi(\mathbf{x}) \tag{17}$$

defines a variant of the Legendre-Fenchel transform adapted to functions $\phi : M \longrightarrow \mathbf{R}$ on a Riemannian manifold. In particular, θ is semiconvex and admits a second order Taylor expansion on a set $\mathrm{dom}\, D^2\theta \subset M$ of full volume [9]. Define, for $\mathbf{x} \in \mathrm{dom}\, D^2\theta$, the displacement $F_s(\mathbf{x}) := \exp_{\mathbf{x}} s\nabla\theta$ which interpolates geodesically between id and F. It is by now well-known that $\nu(s)$ satisfies (12) if and only if $\nu(s) = (F_s)_{\#}\nu$ [16]. Fixing $\tilde{s} \in (0,1)$, a σ-compact set $K \subset \mathrm{dom}\, D^2\theta$ of full measure exists [9, Proposition 5] on which the Monge-Ampère equation

$$u(0,\mathbf{x}) = u(s, F_s(\mathbf{x})) \det A_{\mathbf{x}}(s) > 0 \tag{18}$$

holds for all $\mathbf{x} \in K$ and $s \in [0,1]$ such that $s - \tilde{s} \in \mathbf{Q}$ is rational. Here $s \longrightarrow A_{\mathbf{x}}(s)$ is the unique $n \times n$ matrix of Jacobi fields along the geodesic $F_s(\mathbf{x})$ verifying $A_{\mathbf{x}}(0) = I$ and $A'_{\mathbf{x}}(0) = D^2\theta(\mathbf{x})$ (working with respect to a parallel orthonormal frame along the geodesic). Furthermore, $F_s(K)$ is a Borel set of full mass for u_s and we have $(\mathbf{x}, F_1(\mathbf{x})) \notin \mathrm{Cut}(M,g)$ and $F_s(\mathbf{x}) = \sigma(\mathbf{x}, F(\mathbf{x}), s)$ for $\mathbf{x} \in K$ [8]. By compactness, our manifold admits a Ricci curvature bound $\mathrm{Ric} \geq \lambda g$ for some $\lambda \in \mathbf{R}$. Theorem 10 of Cordero-Erausquin, McCann & Schmuckenschläger [9] asserts convexity of $E(u(s)) + \lambda W_2^2(\omega,\tilde{\omega})s^2/2$ on $s \in [0,1]$. In other words, $E(u(s))$ is semiconvex and has a second order Taylor expansion a.e. in $[0,1]$. If $\tilde{s}$ is such a point, the limit (13) exists and can be computed along a rational sequence $\mathbf{Q} \ni \delta \to 0$. From the fact that

$\nu(s) = (F_s)_\# \nu$ we find

$$\begin{aligned} &E(u(\tilde{s}+\delta)) + E(u(\tilde{s}-\delta)) - 2E(u(\tilde{s})) \\ &= \int_K \omega(\mathbf{x}) \log \frac{u(\tilde{s}+\delta, F_{\tilde{s}+\delta}(\mathbf{x}))u(\tilde{s}-\delta, F_{\tilde{s}-\delta}(\mathbf{x}))}{u(\tilde{s},\mathbf{x})^2}. \end{aligned}$$

Using (18) when δ is rational, and knowing the limit exists, we find that

$$\left.\frac{d^2}{ds^2}\right|_{\tilde{s}} E(u(s)) = \lim_{\delta \to 0} \int_K \omega(\mathbf{x}) \frac{\varphi(\tilde{s}+\delta,\mathbf{x}) + \varphi(\tilde{s}-\delta,\mathbf{x}) - 2\varphi(\tilde{s},\mathbf{x})}{\delta^2} \tag{19}$$

where $\varphi(s,\mathbf{x}) := -\log\det A_{\mathbf{x}}(s)$ (a smooth function of s for each $\mathbf{x} \in K$). By working directly with the definition of Jacobi fields, one can estimate, for $\mathbf{x} \in K$,

$$\frac{\partial^2 \varphi}{\partial s^2}(s,\mathbf{x}) \geq \frac{1}{n}\left(\frac{\partial \varphi}{\partial s}(s,\mathbf{x})\right)^2 + \left.\mathrm{Ric}\left(\frac{d\sigma}{ds}, \frac{d\sigma}{ds}\right)\right|_{\sigma(\mathbf{x},F(\mathbf{x}),s)} \tag{20}$$

as in Lemma 6 of [9] (c.f. [13, §17] or [11, (4.18)], say). One deduces first from this a lower bound for $\frac{\partial^2\varphi}{\partial s^2}(s,\mathbf{x})$, uniformly in $s \in [0,1]$ and $\mathbf{x} \in K$. (We are using the boundedness of the diameter of (M,g) here to control $\frac{d\sigma}{ds}$.) This then gives us a uniform lower bound on the ratio $\frac{1}{\delta^2}(\varphi(\tilde{s}+\delta,\mathbf{x})+\varphi(\tilde{s}-\delta,\mathbf{x})-2\varphi(\tilde{s},\mathbf{x}))$ when $\delta > 0$ is small enough for all terms to be well-defined. Consequently, we may address (19) with Fatou's lemma to deduce

$$\begin{aligned} \left.\frac{d^2}{ds^2}\right|_{\tilde{s}} E(u(s)) &\geq \int_K \omega(\mathbf{x}) \frac{\partial^2\varphi}{\partial s^2}(\tilde{s},\mathbf{x}) \\ &\geq \int_K \omega(\mathbf{x}) \left.\mathrm{Ric}\left(\frac{d\sigma}{ds}, \frac{d\sigma}{ds}\right)\right|_{\sigma(\mathbf{x},F(\mathbf{x}),\tilde{s})} \end{aligned} \tag{21}$$

using (20). Since $\pi_0 = (id \times F)_\# \nu$, and K carries the full mass of the measure ν induced by ω, (11) yields the desired estimate (14).

For the final estimate (15), we follow a similar argument with $\tilde{s} = 0$ and compute as above, that

$$\begin{aligned} \frac{E(u(s)) - E(u(0))}{s} &= \int_K \frac{u(s)\log u(s) - u(0)\log u(0)}{s} dV \\ &= \frac{1}{s}\int_K \log\left(\frac{u(s,F_s(\mathbf{x}))}{u(0,\mathbf{x})}\right)\omega(\mathbf{x}) \\ &= \frac{1}{s}\int_K \varphi(s,\mathbf{x})\omega(\mathbf{x}), \end{aligned} \tag{22}$$

where the final equality is holding for any rational s by (18). Before taking the limit $s \searrow 0$ in $\mathbf{Q}$, we need:

Claim:

$$\liminf_{s\searrow 0}\int_K \frac{\varphi(s,\cdot)}{s}d\nu \geq -\int_K \Delta\theta d\nu.$$

Indeed, for $\mathbf{x}\in K$, we have $\varphi(0,\mathbf{x})=0$ and $\frac{\partial\varphi}{\partial s}(0,\mathbf{x})=-\Delta\theta(\mathbf{x})$, so by Taylor's expansion, we have

$$\frac{\varphi(s,\mathbf{x})}{s} = -\Delta\theta(\mathbf{x}) + \frac{1}{s}\int_0^s (s-b)\frac{\partial^2\varphi}{\partial s^2}(b,\mathbf{x})db,$$

and using again the fact that $\frac{\partial^2\varphi}{\partial s^2}(b,\mathbf{x})$ is bounded below, uniformly in $\mathbf{x}$, we find that

$$\frac{\varphi(s,\mathbf{x})}{s} \geq -\Delta\theta(\mathbf{x}) - sC,$$

for some $C<\infty$. Fatou's lemma may then be invoked to conclude the proof of the claim:

$$\begin{aligned} 0 &= \int_K \liminf_{s\searrow 0}\left(\frac{\varphi(s,\mathbf{x})}{s}+\Delta\theta(\mathbf{x})\right)d\nu(\mathbf{x}) \\ &\leq \liminf_{s\searrow 0}\int_K\left(\frac{\varphi(s,\mathbf{x})}{s}+\Delta\theta(\mathbf{x})\right)d\nu(\mathbf{x}). \end{aligned}$$

Using the claim, we may now take a limit in (22) to give

$$\begin{aligned} \frac{d}{ds}\bigg|_{s=0^+} E(u(s)) &:= \lim_{s\searrow 0}\frac{E(u(s))-E(u(0))}{s} \\ &\geq -\int_K \Delta\theta d\nu \\ &\geq -\int_M (\Delta_{\mathcal{D}'}\theta)u(0)dV \qquad (23) \\ &= \int_M \langle\nabla\theta,\nabla u(0)\rangle\frac{d\nu}{u(0)} \end{aligned}$$

since $0<u(0)\in C^\infty(M)$. in the last inequality we used semiconvexity of θ to know that the distributional Laplacian $\Delta_{\mathcal{D}'}\theta$ was a signed measure with non-negative singular part, and thus pass from its absolutely continuous part $\Delta\theta$ on K to the full distributional Laplacian on M. Appealing to the facts that $\pi_0=(id\times F)_{\#}\nu$, and $\frac{d\sigma}{ds}|_{\sigma(\mathbf{x},F(\mathbf{x}),0^+)}=\nabla\theta(\mathbf{x})$ on the set K of full measure, we recover the desired conclusion (15). ■

Remark 10. *With a little more work, one can in fact show that equality holds in the first inequality of* (23). *Thus, the difference between the left-hand side and the right-hand side in* (15) *can be written precisely in terms of the integral of the singular part of the distributional Laplacian* $\Delta_{\mathcal{D}'}\theta$. *This clarifies the speculation in the last few lines of* [9].

4. Proof of Theorem 2; (A) $\implies$ (B)

We now return to study the coupled system described in the introduction, and prove that (A) implies (B) in Theorem 2, and hence prove Corollary 1 by restricting super Ricci flows to Ricci flows.

Proof. Recall from Section 2 that $h(\tau) := W_2^2(\omega(\tau), \tilde{\omega}(\tau), \tau)/2$ is a continuous function of τ on $[\tau_1, \tau_2]$. By translating time, we may assume that $0 \in (\tau_1, \tau_2)$, and prove that

$$\left.\frac{d^+ h}{d\tau}\right|_{\tau=0} := \limsup_{\tau \searrow 0} \frac{h(\tau) - h(0)}{\tau} \leq 0.$$

We define, as in the introduction, the function $u : M \times (\tau_1, \tau_2) \to (0, \infty)$ by $\omega(\mathbf{x}, \tau) = u(\mathbf{x}, \tau) dV_{g(\tau)}$. In contrast to the construction in Section 3, the maximum principle and parabolicity now guarantee that u is smooth and positive. Let $\psi_\tau : M \to M$ be the family of diffeomorphisms generated by the time-dependent vector field $-\nabla \log u$, with ψ_0 the identity map. Then analogously to the situation in [23] and [29, Chapter 6] for Ricci flow, one can calculate the pull-back

$$\psi_\tau^* \omega(\tau) = \omega(0)$$

for each $\tau \in (\tau_1, \tau_2)$; the diffeomorphism property makes this equivalent to a push-forward: abusing the distinction between n-forms and measures, we have $(\psi_\tau)_\# \omega(0) = \omega(\tau)$. One can also make the identical construction for $\tilde{\omega}$, yielding diffeomorphisms $\tilde{\psi}_\tau$, generated by the vector field $-\nabla \log \tilde{u}$.

Let π_0 be the (unique) optimal transport plan taking $\omega(0)$ to $\tilde{\omega}(0)$. A reasonable competitor for the optimal transport plan at time τ is $(\psi_\tau \times \tilde{\psi}_\tau)_\# \pi_0$. In particular, we note that the marginals of this measure are (the measures induced by) $\omega(\tau)$ and $\tilde{\omega}(\tau)$, making it a valid transport plan. We then know, by (3) and the definition of push-forward measures, that

$$\begin{aligned} h(\tau) &\leq \frac{1}{2} \int_{M \times M} d^2(\mathbf{x}, \mathbf{y}, \tau)\, d((\psi_\tau \times \tilde{\psi}_\tau)_\# \pi_0)(\mathbf{x}, \mathbf{y}) \\ &= \frac{1}{2} \int_{M \times M} d^2(\psi_\tau(\mathbf{x}), \tilde{\psi}_\tau(\mathbf{y}), \tau)\, d\pi_0(\mathbf{x}, \mathbf{y}). \end{aligned} \tag{24}$$

By definition of π_0, we now have

$$\frac{h(\tau) - h(0)}{\tau} \leq \int_{M \times M} \frac{d^2(\psi_\tau(\mathbf{x}), \tilde{\psi}_\tau(\mathbf{y}), \tau) - d^2(\mathbf{x}, \mathbf{y}, 0)}{2\tau}\, d\pi_0(\mathbf{x}, \mathbf{y}), \tag{25}$$

for $\tau > 0$. We wish to take the limit $\tau \searrow 0$. On any time interval $I \subset\subset (\tau_1, \tau_2]$, we have u bounded away from 0, and so $d^2(\psi_\tau(\mathbf{x}), \tilde{\psi}_\tau(\mathbf{y}), \tau)$ is a Lipschitz function of $\tau \in I$, with Lipschitz constant independent of $\mathbf{x}$ and $\mathbf{y}$. Therefore, we may appeal to the Dominated Convergence Theorem and then Remark 6 (which gives a formula for the τ derivative of $d^2(\psi_\tau(\mathbf{x}), \tilde{\psi}_\tau(\mathbf{y}), \tau)$ valid for $(\mathbf{x}, \mathbf{y})$ outside the vanishing π_0-measure [8] set $\mathrm{Cut}(M, g(0)))$ to establish that

$$
\begin{aligned}
\left.\frac{d^+h}{d\tau}\right|_{\tau=0} &\leq \lim_{\tau\to 0} \int_{M\times M} \frac{d^2(\psi_\tau(\mathbf{x}), \tilde{\psi}_\tau(\mathbf{y}), \tau) - d^2(\mathbf{x}, \mathbf{y}, 0)}{2\tau}\, d\pi_0(\mathbf{x}, \mathbf{y}) \\
&= \int_{M\times M} \left.\frac{d}{d\tau}\right|_{\tau=0} \frac{d^2(\psi_\tau(\mathbf{x}), \tilde{\psi}_\tau(\mathbf{y}), \tau)}{2}\, d\pi_0(\mathbf{x}, \mathbf{y}) \\
&= \int_{M\times M} \Bigg(\left.\langle \nabla \log u, \frac{d\sigma}{ds}\rangle\right|_{\sigma(\mathbf{x},\mathbf{y},0^+,0)} - \left.\langle \nabla \log \tilde{u}, \frac{d\sigma}{ds}\rangle\right|_{\sigma(\mathbf{x},\mathbf{y},1^-,0)} \\
&\qquad + \int_0^1 \frac{1}{2}\frac{\partial g}{\partial\tau}\left(\frac{d\sigma}{ds}, \frac{d\sigma}{ds}\right)\Bigg|_{\sigma(\mathbf{x},\mathbf{y},s,0)} ds \Bigg)\, d\pi_0(\mathbf{x}, \mathbf{y}).
\end{aligned} \tag{26}
$$

By considering the W_2-geodesic between $\omega(0)$ and $\tilde{\omega}(0)$ – and in particular, by invoking Corollary 9, we conclude that

$$
\left.\frac{d^+h}{d\tau}\right|_{\tau=0} \leq \int_{M\times M} \int_0^1 \frac{1}{2}\left(\frac{\partial g}{\partial\tau} - 2\mathrm{Ric}\right)\left(\frac{d\sigma}{ds}, \frac{d\sigma}{ds}\right)\Bigg|_{\sigma(\mathbf{x},\mathbf{y},s,0)} ds\, d\pi_0(\mathbf{x}, \mathbf{y}) \leq 0.
$$

■

5. Proof of Theorem 2; (B) $\Longrightarrow$ (C)

Define

$$
\Sigma = \{(\mathbf{x}, \tau, \hat{\mathbf{x}}, \hat{\tau}) \in M \times [\tau_1, \tau_2] \times M \times [\tau_1, \tau_2] \ : \ \tau < \hat{\tau}\},
$$

and a smooth function $u : \Sigma \to (0, \infty)$ as follows. Given $(\hat{\mathbf{x}}, \hat{\tau}) \in M \times (\tau_1, \tau_2]$, we define $u(\mathbf{x}, \tau, \hat{\mathbf{x}}, \hat{\tau})$ for $\tau < \hat{\tau}$ by asking that

$$
\begin{aligned}
\Box u := -\frac{\partial u}{\partial\tau} - \Delta_{\mathbf{x}} u = 0 \qquad & \text{for } (\mathbf{x}, \tau) \in M \times (\tau_1, \hat{\tau}) \\
u(\cdot, \hat{\tau}, \hat{\mathbf{x}}, \hat{\tau}) = \delta_{\hat{\mathbf{x}}} \qquad & \text{on } M
\end{aligned} \tag{27}
$$

where $\Delta_{\mathbf{x}}$ is the Laplacian with respect to the $\mathbf{x}$ entry of $u(\mathbf{x}, \tau, \hat{\mathbf{x}}, \hat{\tau})$ using the metric $g(\tau)$, and the initial condition is the usual shorthand for

$$
\lim_{\tau\nearrow\hat{\tau}} \int_M u(\mathbf{x}, \tau, \hat{\mathbf{x}}, \hat{\tau})\varphi(\mathbf{x}, \tau) d\mathbf{x} = \varphi(\hat{\mathbf{x}}, \hat{\tau}) \text{ for all } \varphi \in C^\infty(M \times (\tau_1, \hat{\tau}]), \tag{28}
$$

where $d\mathbf{x}$ is the Riemannian volume measure with respect to $g(\tau)$ for the parameter $\mathbf{x}$.

The equation in (27) is the usual forward-in-t heat equation (with $\hat{\mathbf{x}}$ and $\hat{\tau}$ fixed) since $\frac{\partial}{\partial\tau} = -\frac{\partial}{\partial t}$. In fact, for fixed $\mathbf{x}$ and τ, the function u will then satisfy the "conjugate heat equation" in $\hat{\mathbf{x}}$ and $\hat{\tau}$:

Lemma 11. *The function u defined as above satisfies, for fixed $\mathbf{x}$ and τ,*

$$\begin{aligned}\Box^* u := \frac{\partial u}{\partial \hat\tau} - \Delta_{\hat{\mathbf{x}}} u + \left(\frac{1}{2}\operatorname{tr}\frac{\partial g}{\partial \hat\tau}\right) u = 0 \text{ for } (\hat{\mathbf{x}}, \hat\tau) \in M \times (\tau, \tau_2) \\ u(\mathbf{x}, \tau, \cdot, \tau) = \delta_{\mathbf{x}} \text{ on } M\end{aligned} \tag{29}$$

where $\Delta_{\hat{\mathbf{x}}}$ is the Laplacian with respect to the $\hat{\mathbf{x}}$ entry of $u(\mathbf{x}, \tau, \hat{\mathbf{x}}, \hat\tau)$ using the metric $g(\hat\tau)$, the notation $\frac{\partial g}{\partial \hat\tau}$ represents the τ-derivative of g at $\hat\tau$ (evaluated at $\hat{\mathbf{x}}$), and the initial condition is the usual shorthand for

$$\lim_{\hat\tau \searrow \tau} \int_M u(\mathbf{x}, \tau, \hat{\mathbf{x}}, \hat\tau)\varphi(\hat{\mathbf{x}}, \hat\tau) d\hat{\mathbf{x}} = \varphi(\mathbf{x}, \tau) \text{ for all } \varphi \in C^\infty(M \times [\tau, \tau_2)), \tag{30}$$

where $d\hat{\mathbf{x}}$ is the Riemannian volume measure with respect to $g(\hat\tau)$ for the parameter $\hat{\mathbf{x}}$.

For fixed $\mathbf{x}$ and τ, one should interpret $u(\mathbf{x}, \tau, \hat{\mathbf{x}}, \hat\tau)d\hat{\mathbf{x}}$ as the probability measure of a Brownian path starting at $(\mathbf{x}, \tau)$ and diffusing forwards in τ until $\hat\tau$. In contrast, for fixed $\hat{\mathbf{x}}$ and $\hat\tau$, the function $u(\cdot, \tau, \hat{\mathbf{x}}, \hat\tau)$ is a likelihood function, not a probability density.

Proof. Having defined $u : \Sigma \to (0, \infty)$ by (27), define $v : \Sigma \to (0, \infty)$ to be the solution of (29). It remains to prove that $u \equiv v$.

Fix $\tau, \hat\tau$ with $\tau_1 < \tau < \hat\tau < \tau_2$ and $\mathbf{x}, \hat{\mathbf{x}} \in M$. We wish to show that $u(\mathbf{x}, \tau, \hat{\mathbf{x}}, \hat\tau) = v(\mathbf{x}, \tau, \hat{\mathbf{x}}, \hat\tau)$. Writing $U(\mathbf{z}, \eta) := u(\mathbf{z}, \eta, \hat{\mathbf{x}}, \hat\tau)$ and $V(\mathbf{z}, \eta) := v(\mathbf{x}, \tau, \mathbf{z}, \eta)$, this would be $U(\mathbf{x}, \tau) = V(\hat{\mathbf{x}}, \hat\tau)$.

For a, b with $\tau < a < b < \hat\tau$, integration by parts (see [29, §6.3]) tells us that

$$\begin{aligned}\left[\int_M V(\mathbf{z}, \eta)U(\mathbf{z}, \eta) d\mathbf{z}\right]_{\eta=a}^{\eta=b} &= -\int_a^b \int_M V(\mathbf{z}, \eta)(\Box U)(\mathbf{z}, \eta) d\mathbf{z}\, d\eta \\ &\quad + \int_a^b \int_M U(\mathbf{z}, \eta)(\Box^* V)(\mathbf{z}, \eta) d\mathbf{z}\, d\eta \\ &= 0,\end{aligned} \tag{31}$$

where $d\mathbf{z}$ is the Riemannian volume element associated to $g(\eta)$ for the parameter $\mathbf{z}$. By (28),

$$\lim_{b \nearrow \hat\tau} \int_M V(\mathbf{z}, b)U(\mathbf{z}, b) d\mathbf{z} = V(\hat{\mathbf{x}}, \hat\tau). \tag{32}$$

Similarly, by (30) (which holds for v, not u, by assumption in this proof) we have

$$\lim_{a \searrow \tau} \int_M V(\mathbf{z}, a)U(\mathbf{z}, a) d\mathbf{z} = U(\mathbf{x}, \tau). \tag{33}$$

Combining (31), (32) and (33), we conclude that $U(\mathbf{x},\tau) = V(\hat{\mathbf{x}},\hat{\tau})$ as desired.
∎

Armed with the function u and its properties, we are in a position to prove the implication (B) $\implies$ (C) of Theorem 2. The proof is inspired by the work of Sturm and von Renesse [27].

Proof. Suppose that $\tau_1 < a < b < \tau_2$. By Lemma 11, we know that for fixed $\mathbf{x} \in M$ and $a \in (\tau_1, \tau_2)$, the function $\mathbf{y} \mapsto u(\mathbf{x}, a, \mathbf{y}, \tau)$ is the probability density, with respect to $g(\tau)$, of Brownian diffusion in the direction of τ, starting at $(\mathbf{x}, a)$ (for $\tau \in (a, \tau_2)$). Since we are assuming (B), we then know that $W_2(u(\mathbf{x}, a, \mathbf{y}, \tau)d\mathbf{y}, u(\tilde{\mathbf{x}}, a, \tilde{\mathbf{y}}, \tau)d\tilde{\mathbf{y}}, \tau)$ is a nonincreasing function of τ. Moreover, by construction,

$$W_2(u(\mathbf{x}, a, \mathbf{y}, \tau)d\mathbf{y}, u(\tilde{\mathbf{x}}, a, \tilde{\mathbf{y}}, \tau)d\tilde{\mathbf{y}}, \tau) \to d(\mathbf{x}, \tilde{\mathbf{x}}, a)$$

as $\tau \searrow a$. Consequently, for $\tau \in (a, \tau_2)$,

$$W_2(u(\mathbf{x}, a, \mathbf{y}, \tau)d\mathbf{y}, u(\tilde{\mathbf{x}}, a, \tilde{\mathbf{y}}, \tau)d\tilde{\mathbf{y}}, \tau) \leq d(\mathbf{x}, \tilde{\mathbf{x}}, a). \tag{34}$$

We are trying to show (C) which we recast into the following equivalent condition:

(C′): If $\tau_1 < \tilde{a} < a < b < \tilde{b} < \tau_2$, and $f : M \times (\tilde{a}, \tilde{b}) \to \mathbf{R}$ solves $-\frac{\partial f}{\partial \tau} = \Delta_{g(\tau)} f$, then

$$Lip(f, a) \leq Lip(f, b),$$

where

$$Lip(f, \tau) := \sup_{\mathbf{x}, \tilde{\mathbf{x}} \in M; \mathbf{x} \neq \tilde{\mathbf{x}}} \frac{|f(\mathbf{x}, \tau) - f(\tilde{\mathbf{x}}, \tau)|}{d(\mathbf{x}, \tilde{\mathbf{x}}, \tau)}$$

is the Lipschitz constant of $f(\cdot, \tau)$.

To prove this, we write $f(\cdot, a)$ in terms of $u : \Sigma \to (0, \infty)$ and $f(\cdot, b)$:

$$f(\mathbf{x}, a) = \int_M u(\mathbf{x}, a, \mathbf{y}, b) f(\mathbf{y}, b) d\mathbf{y}.$$

Now for $\mathbf{x}, \tilde{\mathbf{x}} \in M$, let $\pi(\mathbf{y}, \tilde{\mathbf{y}})$ be any transport plan between the measures $u(\mathbf{x}, a, \mathbf{y}, b)d\mathbf{y}$ and $u(\tilde{\mathbf{x}}, a, \tilde{\mathbf{y}}, b)d\tilde{\mathbf{y}}$. Then

$$f(\mathbf{x}, a) = \int_{M \times M} f(\mathbf{y}, b) d\pi(\mathbf{y}, \tilde{\mathbf{y}}),$$

and similarly,

$$f(\tilde{\mathbf{x}}, a) = \int_{M \times M} f(\tilde{\mathbf{y}}, b) d\pi(\mathbf{y}, \tilde{\mathbf{y}}).$$

Subtracting, we may estimate

$$
\begin{aligned}
|f(\mathbf{x},a)-f(\tilde{\mathbf{x}},a)| &\leq \int_{M\times M} |f(\mathbf{y},b)-f(\tilde{\mathbf{y}},b)|d\pi(\mathbf{y},\tilde{\mathbf{y}}) \\
&\leq Lip(f,b)\int_{M\times M} d(\mathbf{y},\tilde{\mathbf{y}},b)d\pi(\mathbf{y},\tilde{\mathbf{y}}) \\
&\leq Lip(f,b)\left(\int_{M\times M} d^2(\mathbf{y},\tilde{\mathbf{y}},b)d\pi(\mathbf{y},\tilde{\mathbf{y}})\right)^{\frac{1}{2}},
\end{aligned} \tag{35}
$$

where we have used the Cauchy-Schwarz inequality. If we now choose π to be the *optimal* transport plan, we find that

$$
\begin{aligned}
|f(\mathbf{x},a)-f(\tilde{\mathbf{x}},a)| &\leq Lip(f,b)W_2(u(\mathbf{x},a,\mathbf{y},b)d\mathbf{y},u(\tilde{\mathbf{x}},a,\tilde{\mathbf{y}},b)d\tilde{\mathbf{y}},b) \\
&\leq Lip(f,b)d(\mathbf{x},\tilde{\mathbf{x}},a),
\end{aligned} \tag{36}
$$

by (34). Dividing by $d(\mathbf{x},\tilde{\mathbf{x}},a)$ and taking the supremum over $\mathbf{x},\tilde{\mathbf{x}}\in M$ $(\mathbf{x}\neq\tilde{\mathbf{x}})$ we conclude that

$$Lip(f,a)\leq Lip(f,b)$$

as desired. ■

6. Proof of Theorem 2; (C) $\implies$ (A)

We now complete a circle of implications which establishes Theorem 2, mirroring [27].

Proof. Suppose on the contrary, that $g(\tau)$ is not a super Ricci flow, despite (C) holding. Then there exists a time $\tau_0\in(\tau_1,\tau_2)$, a point $\mathbf{x}\in M$ and a vector $X\in T_{\mathbf{x}}M$ of unit length when measured using $g(\tau_0)$, such that

$$\left(-\frac{\partial g}{\partial\tau}(\tau_0)+2\mathrm{Ric}(g(\tau_0))\right)(X,X)<0. \tag{37}$$

Let us work on the fixed Riemannian manifold $(M,g(\tau_0))$ for a moment. Choose $R>0$ less than the injectivity radius of $(M,g(\tau_0))$. Let $\{x^i\}$ be normal coordinates centred at $\mathbf{x}$, such that $\frac{\partial}{\partial x^1}=X$, defined in the ball $B(\mathbf{x},R)$. Let $\Psi: B(\mathbf{x},R/2)\to\mathbf{R}$ be the signed distance function from the level set $\{x^1=0\}$ such that $X(\Psi)=1$ (rather than -1). Then Ψ is a Lipschitz function with the property that $|\nabla\Psi|\leq 1$ almost everywhere in $B(\mathbf{x},R/2)$, with equality in some neighbourhood of $\mathbf{x}$. By virtue of being a signed distance function from a hypersurface, the Hessian of Ψ at $\mathbf{x}$ can be calculated to be

$$\mathrm{Hess}(\Psi)(Y,Z)=-d\Psi(I\!I(Y^T,Z^T))$$

where Y^T, Z^T are the projections onto the hypersurface of arbitrary vectors $Y, Z \in T_{\mathbf{x}}M$, and $I\!I(\cdot,\cdot)$ represents the second fundamental form of the hypersurface. By construction, we have $I\!I(\cdot,\cdot) = 0$ at $\mathbf{x}$, so

$$\mathrm{Hess}(\Psi) = 0$$

at $\mathbf{x}$. Now define $\varphi : M \to [0, R/2]$ to be the Lipschitz cut-off function $\varphi = [R/2 - d(\cdot, \mathbf{x})]_+$, which is supported in $B(\mathbf{x}, R/2)$. Define a Lipschitz function $f_0 : M \to \mathbf{R}$ to be the function Ψ truncated from above by φ, and from below by $-\varphi$. In other words, set

$$f_0(\mathbf{y}) = \max\{\min\{\Psi(\mathbf{y}), \varphi(\mathbf{y})\}, -\varphi(\mathbf{y})\}.$$

This globally defined function is smooth in a neighbourhood of $\mathbf{x}$, has Lipschitz constant equal to 1, and retains from Ψ the properties that

$$\mathrm{Hess}(f_0) = 0 \text{ at } \mathbf{x}; \qquad \nabla f_0(\mathbf{x}) = X; \qquad |\nabla f_0| = 1 \text{ near } \mathbf{x}. \tag{38}$$

We now drop our focus on the fixed Riemannian manifold $(M, g(\tau_0))$ and consider again space-time. Let $f : M \times (\tau_1, \tau_0] \to \mathbf{R}$ be the continuous solution forwards in time (backwards in τ) of the ordinary heat equation:

$$\begin{aligned} \Box u := -\frac{\partial f}{\partial \tau} - \Delta f = 0 \qquad & \text{on } M \times (\tau_1, \tau_0) \\ f(\cdot, \tau_0) = f_0 \qquad & \text{on } M. \end{aligned} \tag{39}$$

The function f is smooth for $\tau < \tau_0$, and even all the way to $\tau = \tau_0$ in a neighbourhood of $\mathbf{x}$. It also satisfies

$$\limsup_{\tau \nearrow \tau_0} \sup_M |\nabla f(\cdot, \tau)| \leq Lip(f, \tau_0) = 1,$$

and

$$\lim_{\tau \nearrow \tau_0} |\nabla f(\mathbf{x}, \tau)| = 1.$$

Since we are assuming (C), we can deduce that

$$\sup_M |\nabla f(\cdot, \tau)| \leq 1 \qquad \text{for all } \tau \in (\tau_1, \tau_0). \tag{40}$$

In contrast, we can compute at $(\mathbf{x}, \tau_0)$,

$$\begin{aligned} \frac{\partial |\nabla f|^2}{\partial \tau} &= \frac{\partial |df|^2}{\partial \tau} \\ &= -\frac{\partial g}{\partial \tau}(\nabla f, \nabla f) + 2\langle d\frac{\partial f}{\partial \tau}, df\rangle \\ &= -\frac{\partial g}{\partial \tau}(\nabla f, \nabla f) - 2\langle d\Delta f, df\rangle \\ &= -\frac{\partial g}{\partial \tau}(\nabla f, \nabla f) + 2\mathrm{Ric}(\nabla f, \nabla f) - \Delta |df|^2 + 2|\mathrm{Hess}(f)|^2, \end{aligned} \tag{41}$$

by the Bochner formula, and so because $f(\cdot,\tau_0)=f_0$, and by (38), we deduce that

$$\frac{\partial|\nabla f|^2}{\partial\tau}(\mathbf{x},\tau_0)=\left(-\frac{\partial g}{\partial\tau}(\tau_0)+2\mathrm{Ric}(g(\tau_0))\right)(X,X).$$

However, by (37), we then find that

$$\frac{\partial|\nabla f|^2}{\partial\tau}(\mathbf{x},\tau_0)<0,$$

and hence that for some $\tau\in(\tau_1,\tau_0)$, $|\nabla f|^2(\mathbf{x},\tau)>1$, contradicting (40). ■

7. Appendix: Proof of Lemma 5

Proof. Implicit in the proof will be the standard characterisation of $\mathrm{Cut}(M,g)$ as the complement of the set of points $(\mathbf{x},\mathbf{y})\in M\times M$ such that there exists a *unique* shortest constant speed geodesic $s\in[0,1]\longrightarrow\sigma(\mathbf{x},\mathbf{y},s)$ from $\mathbf{x}$ to $\mathbf{y}$, *and* $\mathbf{x}$ and $\mathbf{y}$ are not conjugate along $\sigma(\mathbf{x},\mathbf{y},\cdot)$.

Suppose $(\mathbf{x}_0,\xi_0)\in TM$, $\tau_0\in(\tau_1,\tau_2)$ and $\xi_0\notin\mathrm{TConj}(\mathbf{x}_0,\tau_0)$, where for $\mathbf{x}\in M$, $\tau\in(\tau_1,\tau_2)$,

$$\mathrm{TConj}(\mathbf{x},\tau):=\{\xi\in T_{\mathbf{x}}M\ :\ \exp_{\mathbf{x},g(\tau)}\text{ is critical at }\xi\}.$$

By applying the Inverse Function Theorem to the smooth map $\varphi:TM\times(\tau_1,\tau_2)\to M\times M\times(\tau_1,\tau_2)$ given by $\varphi(\mathbf{x},\xi,\tau)=(\mathbf{x},\exp_{\mathbf{x},g(\tau)}\xi,\tau)$, we see that there exist neighbourhoods $V\subset TM\times(\tau_1,\tau_2)$ of $(\mathbf{x}_0,\xi_0,\tau_0)$ and $U\subset M\times M\times(\tau_1,\tau_2)$ of $(\mathbf{x}_0,\exp_{\mathbf{x}_0,g(\tau_0)}\xi_0,\tau_0)$, such that the restriction $\varphi:V\to U$ is a smooth diffeomorphism.

A first consequence of this is that

$$(\mathbf{x},\xi,\tau)\in V\implies\xi\notin\mathrm{TConj}(\mathbf{x},\tau).\tag{42}$$

Now consider an arbitrary point $(\mathbf{x}_0,\mathbf{y}_0,\tau_0)\notin\mathrm{Cut}_{\mathrm{ST}}$. Let $\xi_0\in T_{\mathbf{x}_0}M$ be the unique shortest vector (shortest with respect to $g(\tau_0)$) such that $\mathbf{y}_0=\exp_{\mathbf{x}_0,g(\tau_0)}\xi_0$. By the characterisation of the cut locus recalled at the start of the proof, $\xi_0\notin\mathrm{TConj}(\mathbf{x}_0,\tau_0)$, and so we may find the neighbourhoods U and V as above, and deduce (42).

Claim: For a possibly smaller neighbourhood V of $(\mathbf{x}_0,\xi_0,\tau_0)$, given any $(\mathbf{x},\xi,\tau)\in V$, the geodesic $s\in[0,1]\longrightarrow\gamma(\mathbf{x},\xi,s,\tau):=\exp_{\mathbf{x},g(\tau)}(s\xi)$ is the unique minimising geodesic (with respect to $g(\tau)$) linking $\mathbf{x}$ to $\exp_{\mathbf{x},g(\tau)}(\xi)$.

Before proving the claim, we remark that combining with (42), it would imply that the open neighbourhood $\varphi(V)$ is disjoint from $\mathrm{Cut}_{\mathrm{ST}}$, from which we would deduce the closedness of $\mathrm{Cut}_{\mathrm{ST}}$. It would also enable us to write, for $(\mathbf{x},\mathbf{y},\tau)\in\varphi(V)$, the geodesic $s\in[0,1]\longrightarrow\sigma(\mathbf{x},\mathbf{y},s,\tau)\in M$ as $\sigma(\mathbf{x},\mathbf{y},s,\tau)=\exp_{\mathbf{x},g(\tau)}(s\xi)$, where $\mathbf{y}=\exp_{\mathbf{x},g(\tau)}\xi$, and $(\mathbf{x},\xi,\tau)\in V$. In particular, by the smoothness of φ^{-1}, we would deduce the smooth dependence of $\sigma(\mathbf{x},\mathbf{y},s,\tau)$ and the squared distance function $d^2(\mathbf{x},\mathbf{y},\tau)$ on their parameters, whilst in $\varphi(V)$, and hence throughout the complement of $\mathrm{Cut}_{\mathrm{ST}}$.

It remains to prove the claim. If false, there exist sequences $\{(\mathbf{x}_i,\xi_i)\}\subset TM$, $\{\tau_i\}\subset(\tau_1,\tau_2)$ such that $(\mathbf{x}_i,\xi_i)\to(\mathbf{x}_0,\xi_0)$ and $\tau_i\to\tau_0$, but with $s\in[0,1]\longrightarrow\gamma(\mathbf{x}_i,\xi_i,s,\tau_i)$ *not* a unique minimising geodesic with respect to $g(\tau_i)$ joining $\mathbf{x}_i$ to $\mathbf{y}_i:=\exp_{\mathbf{x}_i,g(\tau_i)}\xi_i$ for each i. Note that $\mathbf{y}_i\to\mathbf{y}_0$ as $i\to\infty$. By omitting a finite number of terms, we may assume that $(\mathbf{x}_i,\xi_i,\tau_i)\in V$, and thus $(\mathbf{x}_i,\mathbf{y}_i,\tau_i)\in U:=\varphi(V)$, for all i.

Let us choose vectors $\hat{\xi}_i\in T_{\mathbf{x}_i}M$ such that $s\longrightarrow\gamma(\mathbf{x}_i,\hat{\xi}_i,s,\tau_i)$ is a minimising geodesic from $\mathbf{x}_i$ to $\mathbf{y}_i$, with $\hat{\xi}_i\neq\xi_i$. After passing to a subsequence, we may assume that $(\mathbf{x}_i,\hat{\xi}_i)\to(\mathbf{x}_0,\hat{\xi}_0)$ as $i\to\infty$, for some $\hat{\xi}_0\in T_{\mathbf{x}_0}M$. Since $\varphi(\mathbf{x}_i,\hat{\xi}_i,\tau_i)=\varphi(\mathbf{x}_i,\xi_i,\tau_i)=(\mathbf{x}_i,\mathbf{y}_i,\tau_i)$, and the restriction $\varphi:V\to U$ is a diffeomorphism, we must have $(\mathbf{x}_i,\hat{\xi}_i,\tau_i)\notin V$ for all i, and in particular, we must have $\hat{\xi}_0\neq\xi_0$.

Consequently, $s\longrightarrow\gamma(\mathbf{x}_0,\hat{\xi}_0,s,\tau_0)$ and $s\longrightarrow\gamma(\mathbf{x}_0,\xi_0,s,\tau_0)$ must be distinct minimising geodesics between $\mathbf{x}_0$ and $\mathbf{y}_0$, with repect to $g(\tau_0)$, contradicting the assumption that $(\mathbf{x}_0,\mathbf{y}_0,\tau_0)\notin\mathrm{Cut}_{\mathrm{ST}}$.

■

References

[1] L.A. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* 'Lecture Notes in Mathematics ETH Zürich'. Birkhäuser Verlag, Basel, 2005.

[2] D. Bakry and M. Émery. *Diffusions Hypercontractives. Séminaire de Probabilités XIX, 1983/84.* 'Lecture Notes in Math.' **1123**. Springer, Berlin, 1985.

[3] J.D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.* **84** (2000) 375–393.

[4] E. Cabezas-Rivas and P.M. Topping. The canonical shrinking soliton associated to a Ricci flow. Preprint (2008). `http://www.warwick.ac.uk/~maseq`

[5] M. Carfora, Fokker-Planck dynamics and entropies for normalized Ricci flow. *Adv. Theor. Math. Phys.* **11** (2007), 635–681.

[6] J.A. Carrillo, R.J. McCann, and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Rational Mech. Anal.* **179** (2006) 217–263.

[7] I. Chavel. *Riemannian Geometry: a Modern Introduction.* 'Cambridge tracts in math.' **108** Cambridge University Press 1995.

[8] D. Cordero-Erausquin, R.J. McCann and M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.* **146** (2001) 219–257.

[9] D. Cordero-Erausquin, R.J. McCann and M. Schmuckenschläger. Prekopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport. *Ann. Fac. Sci. Toulouse Math. (6)* **15** (2006) 613-635.

[10] D. DeTurck. Deforming metrics in the direction of their Ricci tensors. In *Collected papers on Ricci flow.* Edited by H. D. Cao, B. Chow, S. C. Chu and S. T. Yau. 'Series in Geometry and Topology' **37**. International Press, 2003.

[11] S. Gallot, D. Hulin and J. Lafontaine, *Riemannian Geometry. (Second edition).* Springer-Verlag, 1993.

[12] R.S. Hamilton, Three-manifolds with positive Ricci curvature. *J. Differential Geom.* **17** (1982) 255–306.

[13] R.S. Hamilton, The formation of singularities in the Ricci flow. In *Surveys in Differential Geometry (Cambridge MA, 1993), Vol. II.* 7–136, International Press, Cambridge MA, 1995.

[14] R. Jordan, D. Kinderlehrer and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** (1998) 1–17.

[15] J. Lott. Talk at UCSB (2005).

[16] J. Lott and C. Villani. Ricci curvature in metric measure spaces via optimal transport. To appear in *Annals Math. (2).* `http://arXiv.org/math.DG/0412127`

[17] J. Lott and C. Villani. Weak curvature conditions and functional inequalities. *J. Funct. Anal.* **245** (2007) 311–333.

[18] J. Lott Optimal transport and Perelman's reduced volume. Preprint (2008).

[19] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.* **128** (1997) 153–179.

[20] R.J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.* **11** (2001) 589–608.

[21] F. Otto. The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations* **26** (2001) 101–174.

[22] 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.

[23] G. Perelman The entropy formula for the Ricci flow and its geometric applications. `http://arXiv.org/math.DG/0211159`v1 (2002).

[24] K.-T. Sturm. Generalized Ricci curvature bounds and convergence of metric measure spaces. *C.R. Math. Acad. Sci. Paris* **340** (2005) 235–238.

[25] K.-T. Sturm. A curvature-dimension condition for metric measure spaces. *C.R. Math. Acad. Sci. Paris* **342** (2006) 197–200.

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

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

[28] H. Tanaka. An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **27** (1973) 47–52.

[29] P.M. Topping, *Lectures on the Ricci flow.* London Mathematical Society Lecture Notes Series **325** Cambridge University Presss (2006)
`http://www.warwick.ac.uk/~maseq/RFnotes.html`

[30] P.M. Topping, $\mathcal{L}$-optimal transportation for Ricci flow. Preprint (2007). *J. reine angew. Math., to appear.* `http://www.warwick.ac.uk/~maseq`

†Department of Mathematics, University of Toronto, Ontario Canada M5S 2E4 `mccann@math.toronto.edu`

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