Bridging the Gap between Optimal Transport and Ricci Flow

Jiaxi Zhao

2019/10/21

1 Ricci Curvature Bound and the Geodesic Convexity of Entropy

Ricci curvature lower bound is shown to be connected with a great many important properties of sample space. In this section, we provide a complete list of the result mainly from the work of Sturm & Renesse [1].

Theorem 1. *For any smooth connected compact Riemannian manifold M and any $K \in \mathbb{R}$, the following properties are equivalent:*
(1)$\mathrm{Ric} \geq K$.
(2)The normalized Riemannian uniform distribution on spheres

$$\sigma_{r,x}(A) := \frac{H^{n-1}\left(A \cap \partial B_r(x)\right)}{H^{n-1}\left(\partial B_r(x)\right)}, \quad A \in \mathscr{B}(M),$$

satisfies the asymptotic estimate

$$d_1^W\left(\sigma_{r,x}, \sigma_{r,y}\right) \leq \left(1 - \frac{K}{2n} r^2\right) + o\left(r^2\right) \cdot d(x,y).$$

where the error term is uniform w.r.t. $x, y \in M$.
(3)The normalized Riemannian uniform distribution on balls

$$m_{r,x}(A) := \frac{m\left(A \cap \partial B_r(x)\right)}{m\left(\partial B_r(x)\right)}, \quad A \in \mathscr{B}(M),$$

satisfies the asymptotic estimate

$$d_1^W\left(m_{r,x}, m_{r,y}\right) \leq \left(1 - \frac{K}{2(n+2)} r^2 + o\left(r^2\right)\right) \cdot d(x,y).$$

where the error term is uniform w.r.t. $x, y \in M$.

1.1 Generalization of the Ricci Tensor

Various generalizations of the Ricci curvature tensor has been proposed, including the one from Bakry-Emerycite BE

$$\mathrm{Ric}_f = \mathrm{Ric} + \mathrm{Hess}\, f, \tag{1}$$

and the one further generalized by Sturmcite Sturm

$$\mathrm{Ric}_f^N = \mathrm{Ric} + \mathrm{Hess}\, f - \frac{1}{N} df \otimes df. \tag{2}$$

Notice that those generalization of the Bakry-Emery Ricci tensor is in the regime of smooth metric space. And theorems concerned with the compactness of the metric space and the bound of the diameter are also proved as below(We state the theorem in French)

Theorem 2. *Si M est complete, avec* $\mathrm{Ric}_f^N \geq (n-1)\,H\,(N \leq +\infty)$.
(1)Quand $N < +\infty$, M est compact et

$$\mathrm{diam}\,(M) \leq \sqrt{\frac{n+N-1}{n-1}\frac{\pi}{\sqrt{H}}}.$$

Another intéressant point du bound de la curvature Ricci est que il y a une intégral form du bound de Ricci tensor sur le diameter du manifold. Nous on écrivons au bas.

Theorem 3. *Si M est complete, $\exists p \in M$, s.t. toute de façon de geodesic(aucun de geodesic) que partit de p*

$$\int_0^{+\infty} \mathrm{Ric}\,(\dot\gamma(s), \dot\gamma(s))\,ds = +\infty.$$

M est compact.

2 Ricci Flow and the Gradient Flow of Relative Entropy

First recall the definition of relative entropy $H(\mu_t|\nu) = \int d\mu_t \log \frac{d\mu_t}{d\nu}$. Via elementary geometric calculation in Wasserstein geometry, it can be shown that entropy dissipation along the gradient w.r.t. the relative entropy $H\,(\cdot|\nu)$ is given by:

$$-\frac{d}{dt}H\,(\mu_t|\nu) = \int \left|\nabla \log \frac{d\mu_t}{d\nu}\right|^2 d\mu_t. \tag{3}$$

Here we can first consider the flat case then move to the manifold casewhere the RHS is exactly the well-known Fisher information functional.

On the other hand, if we fix the first argument, that's consider functional $H(\mu|\cdot)$, the entropy dissipation w.r.t. gradient flow associated with this entropy is provided:

$$-\frac{d}{dt}H\,(\mu|\nu_t) = \int \left|\nabla \frac{d\mu}{d\nu_t}\right|^2 d\nu_t. \tag{4}$$

notice these two dissipation rates are not the same, but up to a factor $\frac{d\mu}{d\nu}$ in the integrand.

On the Ricci side, similar equation can also be written down as following. We define the entropy functional H as below:

$$H(u_t, g_t) = \int u_t \log u_t dV_t. \tag{5}$$

notice that the entropy functional in the Ricci side relates to both the metric g_t(influence dV_t) and function u_t. And the flow equation describes this dynamics is given by:

$$\begin{cases} \dfrac{\partial g_t}{\partial t} = -\,2\,\mathrm{Ric}(g_t), \\ \Box^* u_t = 0, \qquad \Box^* = \dfrac{\partial}{\partial t} + \Delta_t - R_t. \end{cases} \tag{6}$$

notice here that both Laplacian on the manifold Δ_t and curvature tensor R_t has subscript t to indicate its dependence on the time-evolving metric g_t. And the entropy dissipation along this dynamics is given by:

$$\begin{aligned} -\frac{d}{dt}H\,(u_t, g_t) &= -\int \left((1+\log u_t)\,\frac{\partial u_t}{\partial t} - R_t u_t \log u_t\right) dV_t \\ &= \int \left((1+\log u_t)\,(-\Delta_t u_t + R_t u_t) - R_t u_t \log u_t\right) dV_t \\ &= \int \left((1+\log u_t)\,(-\Delta_t u_t) + R_t u_t\right) dV_t \\ &= \int \left(\frac{|\nabla_t u_t|^2}{u_t} + R_t u_t\right) dV_t = \mathscr{F}(u_t, g_t). \end{aligned} \tag{7}$$

Actually, the dynamics 6 is a decoupled version of another flow, namely:

$$\begin{cases} \dfrac{\partial g_t}{\partial t} = -2\left(\mathrm{Ric}(g_t) + \mathrm{Hess}_t f_t\right), \\ \Box^* u_t = 0, \qquad \Box^* = \dfrac{\partial}{\partial t} + \Delta_t - R_t. \end{cases} \tag{8}$$

There's still some wrong here with the evolution of u

Next we compute the gradient of the Fisher information functional we defined previously:

$$\begin{aligned} \frac{d}{dt}\mathscr{F}(u_t, g_t) &= \int \left(\frac{|\nabla_t u_t|^2}{u_t} + R_t u_t \right) dV_t \\ &= \int \left(\frac{|\nabla_t u_t|^2}{u_t} + R_t u_t \right) \frac{1}{2}\operatorname{tr} h dV_t + \int \left(\frac{|\nabla_t u_t|^2}{u_t} + R_t k_t + u_t \left(\langle -\mathrm{Ric}_t, h_t \rangle + \delta^2 h + \Delta h \right) \right) dV_t \\ &= \end{aligned} \tag{9}$$

where we assume $\frac{\partial g_t}{\partial t} = h_t$, $\frac{\partial}{\partial t} u_t = k_t$.

2.1 A Trivial Example

We first show that the Ricci flow dynamics 6 on $\mathbb{R}^n$ space we introduced in previous section is exactly the gradient flow of entropy on the Wasserstein density space $\mathscr{P}(\mathbb{R}^n)$. Simplifying the Ricci flow dynamics 6 as:

$$\begin{cases} \dfrac{\partial g_t}{\partial t} = 0 \\ \dfrac{\partial}{\partial t} u_t + \Delta_t u_t = 0 \end{cases}$$

under which we find that the metric tensor g_t remains the same while the probability density function u_t evolves under heat equation, since g_t remains the same, the Laplacian Δ_t does the same. As is known in the Wasserstein community, such evolution, the Fokker-Planck equation of the standard Brownian motion, is exactly the gradient flow w.r.t. the entropy functional on $\mathbb{R}^n$.

Remark 1. *A gap between the flow in Wasserstein space and flow in Ricci side is that we consider different object in two contexts. In Wasserstein space, we fix the base manifold's geometry g and study only the dynamics concerned with the probability evolution μ_t, while in Ricci side we pack the ground metric g and probability u together as a whole thing and consider its evolution. Actually, specifying both the geometry and a probability measure μ on it is a crucial prespective in modern geometry.*

Thus, a natural question arise: *Is there an analogue in Ricci side for the gradient flow w.r.t. the relative entropy on the Wasserstein density manifold?*

We can argue here by an heuristic illustration. The probability evolution w.r.t. the gradient flow of relative entropy $H(\cdot|\nu)$ on the Wasserstein density space is given by:

$$\frac{\partial \mu_t}{\partial t} = \nabla \cdot (\mu_t \nabla \log \nu) - \Delta \mu_t = \mu_t \Delta \log \nu + \nabla \mu_t \cdot \nabla \nu - \Delta \mu_t \tag{10}$$

Compared with the evolution of Ricci dynamics 6 we find such correspondence:

$$\Delta \log \nu \leftrightarrow R_t, \quad \nabla \mu_t \cdot \nabla \nu \leftrightarrow ? \tag{11}$$

where the ? means that we cannot find a counterpart in Ricci side, indicating the gap we have to bridge.

2.2 Ricci flow and Heat flow

Another interesting point is that Ricci flow is also connected with the heat flow, both in the sense that curvature tensor can be viewed as the Hessian of metric and Ricci curvature as Laplacian and in the sense that they both have some kind of "smoothify" effect, at least intuitively.

2.3 Bochner Formula

Bochner formula is one of the most famous result in the field of geometry analysis. At each point p, we choose a normal coordinate $\{E_i, i = 1, ..., n\}$, i.e. $g_{ij} = \delta_{ij}, \nabla_{E_i} E_j = 0$. Then, we have

$$
\begin{aligned}
\Delta |\nabla f|^2 &= \operatorname{tr}\left(\operatorname{Hess} |\nabla f|^2\right) \\
&= \sum_{i=1}^{n} \left(\operatorname{Hess} |\nabla f|^2 (E_i, E_i)\right) \\
&= \sum_{i=1}^{n} g\left(\nabla_{E_i} \nabla |\nabla f|^2, E_i\right) \\
&= \sum_{i=1}^{n} E_i \left(g\left(\nabla |\nabla f|^2, E_i\right)\right) \\
&= \sum_{i=1}^{n} E_i \left(E_i \left(|\nabla f|^2\right)\right) \\
&= 2\sum_{i=1}^{n} E_i \left(g\left(\nabla_{E_i} \nabla f, \nabla f\right)\right) \\
&= 2\sum_{i=1}^{n} E_i \left(\operatorname{Hess} f\left(E_i, \nabla f\right)\right) \\
&= 2\sum_{i=1}^{n} E_i \left(\operatorname{Hess} f\left(\nabla f, E_i\right)\right) \\
&= 2\sum_{i=1}^{n} E_i \left(g\left(\nabla_{\nabla f} \nabla f, E_i\right)\right).
\end{aligned}
$$

Further, we have

$$
\begin{aligned}
&\sum_{i=1}^{n} E_i \left(g\left(\nabla_{\nabla f} \nabla f, E_i\right)\right) \\
= &\sum_{i=1}^{n} \left(g\left(\nabla_{E_i}\left(\nabla_{\nabla f} \nabla f\right), E_i\right) + g\left(\left(\nabla_{\nabla f} \nabla f\right), \nabla_{E_i} E_i\right)\right) \\
= &\sum_{i=1}^{n} g\left(\nabla_{E_i}\left(\nabla_{\nabla f} \nabla f\right), E_i\right),
\end{aligned}
$$

where we use the assumption that the frame $\{E_i, i = 1, ..., n\}$ is a normal frame. Continuing the calculation, we have

$$
\begin{aligned}
&\sum_{i=1}^{n} g\left(\nabla_{E_i} \nabla_{\nabla f} \nabla f, E_i\right) \\
= &\sum_{i=1}^{n} \left(g\left(\nabla_{E_i} \nabla_{\nabla f} \nabla f - \nabla_{\nabla f} \nabla_{E_i} \nabla f - \nabla_{[E_i, \nabla f]} \nabla f, E_i\right) + g\left(\nabla_{\nabla f} \nabla_{E_i} \nabla f + \nabla_{[E_i, \nabla f]} \nabla f, E_i\right)\right) \\
= &\sum_{i=1}^{n} g\left(\mathrm{R}(E_i, \nabla f)\nabla f, E_i\right) + \sum_{i=1}^{n} g\left(\nabla_{\nabla f} \nabla_{E_i} \nabla f, E_i\right) + \sum_{i=1}^{n} g\left(\nabla_{[E_i, \nabla f]} \nabla f, E_i\right).
\end{aligned}
$$

For other two terms, separately we have

$$\begin{aligned}
&\sum_{i=1}^{n} g\left(\nabla_{[E_i,\nabla f]}\nabla f, E_i\right) \\
=&\sum_{i=1}^{n} g\left(\nabla_{\nabla_{E_i}\nabla f-\nabla_{\nabla f}E_i}\nabla f, E_i\right) \\
=&\sum_{i=1}^{n} g\left(\nabla_{\nabla_{E_i}\nabla f}\nabla f, E_i\right) \\
=&\sum_{i=1}^{n} \mathrm{Hess}\, f\left(\nabla_{E_i}\nabla f, E_i\right) \\
=&\sum_{i=1}^{n} \mathrm{Hess}\, f\left(E_i, \nabla_{E_i}\nabla f\right) \\
=&\sum_{i=1}^{n} g\left(\nabla_{E_i}\nabla f, \nabla_{E_i}\nabla f\right) = \left|\mathrm{Hess}\, f\right|^2,
\end{aligned}$$

and

$$\sum_{i=1}^{n} g\left(\nabla_{\nabla f}\nabla_{E_i}\nabla f, E_i\right)$$

2.4 Connection between the log-Sobolev inequality and Perelmen's W-functional

2.4.1 Case of $\mathbb{R}^n$

In Topping's note, he connects

2.4.2 Case of $\mathbb{S}^n$

Here, we consider the case of high dimensional sphere. In the context of measure concentration cite concentration slide, people have found that uniform measure on the high dimensional sphere has a lot to do with the Gaussian measure on the high dimensional Euclidean space.

Consider the Perelmen W-functional

$$W(g, f, \tau) = \int \left(\tau\left(|\nabla f|^2 + R\right) - n - f\right) u dV.$$

Notice that an interesting point is that, when we are in Euclidean space, the scalar curvature R vanishes and the dimensional constant n is canceled by Gaussian kernel. While in spherical case, the dimensional constant n is directly canceled by the scalar curvature for sphere.

3 Wasserstein Distance under Ricci Flow

Studying the quantities along the Ricci flow is of special interest, for example, Yamabe problem is concerned with ... Also, in the previous section, studying Fisher information functional along the Ricci flow provides various connection between both the geometric and probabilistic quantities. In this section, we focus on the series of results which is concerned with the evolution of the Wasserstein distance under both the Ricci flow and the heat flow.

We first state the result belongs to MacCan and Topping:

Theorem 4.

4 Brunn-Minkowski Inequality

The Brunn-Minkowski inequality can be thought as the basis of the geometry analysis.

It's important to understand the convex body with the function such as support function defined on it.

4.1 Development of Brunn-Minkowski Inequality

Quasimassintegral: perimeter; Dual Brunn-Minkowski inequality and L^p Brunn-Minkowski inequality

Appearance of Monge-Ampère equation in the analytic method of Brunn-Minkowski inequality. It comes from the variational formulation of the Minkowski problems. The Monge-Ampère equation for the dual formulation have another term which can not be fitted into the framework of optimal transport.

It will be interesting to study the geometric aspect of the dual Minkowski problem.

References

[1] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy and ricci curvature. *Communications on Pure and Applied Mathematics*, 58(7):923–940, 2005.