MANIFOLD MICRO-SURGERY WITH LINEARLY NEARLY EUCLIDEAN METRICS

Anonymous authors
Paper under double-blind review

ABSTRACT

The Ricci flow is a method of manifold surgery, which can trim manifolds to more regular. However, in most cases, the Rich flow tends to develop singularities and lead to divergence of the solution. In this paper, we propose linearly nearly Euclidean metrics to assist manifold micro-surgery, which means that we prove the dynamical stability and convergence of such metrics under the Ricci-DeTurck flow. From the information geometry and mirror descent points of view, we give the approximation of the steepest descent gradient flow on the linearly nearly Euclidean manifold with dynamical stability. In practice, the regular shrinking or expanding of Ricci solitons with linearly nearly Euclidean metrics will provide a geometric optimization method for the solution on a manifold.

1 INTRODUCTION

In general relativity (Wald, 2010), a complete Riemannian manifold $(\mathcal{M}, g)$ endowed with a linearly nearly flat spacetime metric g_{ij} is considered for linearized gravity to solve the Newtonian limit. The form of this metric is $g_{ij} = \eta_{ij} + \gamma_{ij}$, where η_{ij} represents a flat Minkowski metric whose background is special relativity and γ_{ij} is small from η_{ij}. An adequate definition of "smallness" in this context is that the components of γ_{ij} are much smaller than 1 in some global inertial coordinate system of η_{ij}. Now, let us step out of the physical world and give a similar metric $g_{ij} = \delta_{ij} + \gamma_{ij}$ in Riemannian n-manifold $(\mathcal{M}^n, g)$, i.e. the linearly nearly Euclidean metric, where δ_{ij} represents a flat Euclidean metric and γ_{ij} is small from δ_{ij}.

A natural problem for such a linearly nearly Euclidean metric is: how does the metric evolve over time with respect to the Ricci flow while ensuring the constant topological structure? Let us review some stability analyses of different manifolds along with the Ricci flow.

For the Riemannian n-dimensional manifold $(\mathcal{M}^n, g)$ that is isometric to the Euclidean n-dimensional space $(\mathbb{R}^n, \delta)$, Schnürer et al. (Schnürer et al., 2007) have showed the stability of Euclidean space under the Ricci flow for a small C^0 perturbation. Koch et al. (Koch & Lamm, 2012) have given the stability of the Euclidean space along with the Ricci flow in the L^∞-Norm. Moreover, for the decay of the L^∞-Norm on Euclidean space, Appleton (Appleton, 2018) has given the proof of a different method. Considering the stability of integrable and closed Ricci-flat metrics, Sesum (Sesum, 2006) has proved that the convergence rate is exponential because the spectrum of the Lichnerowicz operator is discrete. Furthermore, Deruelle et al. (Deruelle & Kröncke, 2021) have proved that an asymptotically locally Euclidean Ricci-flat metric is dynamically stable under the Ricci flow, for an $L^2 \cap L^\infty$ perturbation on non-flat and non-compact Ricci-flat manifolds.

In this paper, we consider a complete Riemannian n-dimensional manifold $(\mathcal{M}^n, g)$ endowed with linearly nearly Euclidean metrics $g(t) = \delta + \gamma(t)$. We prove the stability of linearly nearly Euclidean manifolds under the Ricci-DeTurck flow in the L^2-Norm if initial metrics are integrable and linearly stable, i.e. has a manifold structure of finite dimension. We mean that any Ricci-DeTurck flow which starts from near g exists for all time and converges to a linearly nearly Euclidean metric near g. Moreover, we use the Einstein summation convention and denote generic constants by C or C_1.

Furthermore, we define and construct linearly nearly Euclidean manifolds based on information geometry and mirror descent algorithm. Based on a symmetrized convex function, we obtain linearly nearly Euclidean divergences that can be the set of stationary points of such manifolds along with the Ricci-DeTurck flow. Therefore, we can easily approximate the steepest descent gradient flow under

the Ricci-DeTurck flow. When we use the gradient flow to learn a neural network, we observe the evolution of its metric is consistent with the micro-surgery process under the Ricci-DeTurck flow.

2 RICCI FLOW

Let us introduce a partial differential equation, the Ricci flow, without explanation. The concept of the Ricci flow first published by Hamilton (Hamilton et al., 1982) on the manifold $\mathcal{M}$ of a time-dependent Riemannian metric $g(t)$ with the initial metric g_0:

$$\begin{aligned} \frac{\partial}{\partial t} g(t) &= -2\,\mathrm{Ric}(g(t)) \\ g(0) &= g_0 \end{aligned} \tag{1}$$

where Ric denotes the Ricci curvature tensor whose definition can be found in Appendix A.

The purpose of the Ricci flow is to prove Thurston's Geometrization Conjecture and Poincaré Conjecture because the Ricci flow is like a surgical scalpel, trimming irregular manifolds into regular manifolds to facilitate observation and discussion (Sheridan & Rubinstein, 2006).

In general, in order to possess good geometric and topological properties, we expect the metric to become converge and round with the help of the Ricci flow. "become round" means that the solution will not shrink to a point but converge to a constant circle. However, in most cases, we do not even know the convergence of the solution and whether the solution will develop a singularity. Next, we will discuss these issues for brevity.

2.1 SHORT TIME EXISTENCE

To show that there exists a unique solution for a short time, we must check if the system of the Ricci flow is strongly parabolic.

Theorem 1 *When $u : \mathcal{M} \times [0, T) \to \mathcal{E}$ is a time-dependent section of the vector bundle $\mathcal{E}$ where $\mathcal{M}$ is some Riemannian manifold, if the system of the Ricci flow is strongly parabolic at u_0 then there exists a solution on some time interval $[0, T)$, and the solution is unique for as long as it exists.*

Proof. The proofs can be found in (Ladyzhenskaia et al., 1988). □

Definition 1 *The Ricci flow is strongly parabolic if there exists $\delta > 0$ such that for all covectors $\varphi \neq 0$ and all symmetric $h_{ij} = \frac{\partial g_{ij}(t)}{\partial t} \neq 0$, the principal symbol of $-2\,\mathrm{Ric}$ satisfies*

$$[-2\,\mathrm{Ric}](\varphi)(h)_{ij} h^{ij} = g^{pq} \left(\varphi_p \varphi_q h_{ij} + \varphi_i \varphi_j h_{pq} - \varphi_q \varphi_i h_{jp} - \varphi_q \varphi_j h_{ip}\right) h^{ij} > \delta \varphi_k \varphi^k h_{rs} h^{rs}.$$

Since the inequality cannot always be satisfied, the Ricci flow is not strongly parabolic, which makes us unable to prove the existence of the solution based on Theorem 1.

It is possible to understand which parts have an impact on its non-parabolic by the linearization of the Ricci curvature tensor.

Lemma 1 *The linearization of $-2\,\mathrm{Ric}$ can be rewritten as*

$$\begin{aligned} &D[-2\,\mathrm{Ric}](h)_{ij} = g^{pq} \nabla_p \nabla_q h_{ij} + \nabla_i V_j + \nabla_j V_i + O(h_{ij}) \\ &\text{where} \quad V_i = g^{pq} \left(\frac{1}{2} \nabla_i h_{pq} - \nabla_q h_{pi}\right). \end{aligned} \tag{2}$$

Proof. The proofs can be found in Appendix B.1. □

The term $O(h_{ij})$ will have no contribution to the principal symbol of $-2\,\mathrm{Ric}$, so ignoring it will not affect our discussion of this problem. By carefully observing the above equation, one finds that the impact on the non-parabolic of the Ricci flow comes from the terms in V, not the term $g^{pq} \nabla_p \nabla_q h_{ij}$. The solution is followed by the DeTurck Trick (DeTurck, 1983) that has a time-dependent reparameterization of the manifold:

$$\begin{aligned} \frac{\partial}{\partial t} \bar{g}(t) &= -2\,\mathrm{Ric}(\bar{g}(t)) - \mathcal{L}_{\frac{\partial \varphi(t)}{\partial t}} \bar{g}(t) \\ \bar{g}(0) &= \bar{g}_0 + d, \end{aligned} \tag{3}$$

where d is a symmetric (0,2)-tensor on $\mathcal{M}$. See Appendix B.2 for details. By choosing $\frac{\partial \varphi(t)}{\partial t}$ to cancel the effort of the terms in V, the reparameterized Ricci flow is strongly parabolic. Thus, one can say that the Ricci-DeTurck flow has a unique solution, the pullback metric, for a short time.

2.2 Curvature Explosion at Singularity

In this subsection, we will present the behavior of the Ricci flow in finite time and show that the evolution of the curvature is close to divergence. The core demonstration is followed with Theorem 4, which requires some other proof as a foreshadowing.

Theorem 2 *Given a smooth Riemannian metric g_0 on a closed manifold $\mathcal{M}$, there exists a maximal time interval $[0, T)$ such that a solution $g(t)$ of the Ricci flow, with $g(0) = g_0$, exists and is smooth on $[0, T)$, and this solution is unique.*

Proof. The proofs can be found in (Sheridan & Rubinstein, 2006). □

Theorem 3 *Let $\mathcal{M}$ be a closed manifold and $g(t)$ a smooth time-dependent metric on $\mathcal{M}$, defined for $t \in [0, T)$. If there exists a constant $C < \infty$ for all $x \in \mathcal{M}$ such that*

$$\int_0^T \left| \frac{\partial}{\partial t} g_x(t) \right|_{g(t)} dt \leq C, \tag{4}$$

then the metrics $g(t)$ converge uniformly as t approaches T to a continuous metric $g(T)$ that is uniformly equivalent to $g(0)$ and satisfies

$$e^{-C} g_x(0) \leq g_x(T) \leq e^{C} g_x(0).$$

Proof. The proofs can be found in Appendix B.3. □

Corollary 1 *Let $(\mathcal{M}, g(t))$ be a solution of the Ricci flow on a closed manifold. If $|\operatorname{Rm}|_{g(t)}$ is bounded on a finite time $[0, T)$, then $g(t)$ converges uniformly as t approaches T to a continuous metric $g(T)$ which is uniformly equivalent to $g(0)$.*

Proof. The bound on $|\operatorname{Rm}|_{g(t)}$ implies one on $|\operatorname{Ric}|_{g(t)}$. Based on Equation (1), we can extend the bound on $|\frac{\partial}{\partial t} g(t)|_{g(t)}$. Therefore, we obtain an integral of a bounded quantity over a finite interval is also bounded, by Theorem 3. □

Theorem 4 *If g_0 is a smooth metric on a compact manifold $\mathcal{M}$, the Ricci flow with $g(0) = g_0$ has a unique solution $g(t)$ on a maximal time interval $t \in [0, T)$. If $T < \infty$, then*

$$\lim_{t \to T} \left(\sup_{x \in \mathcal{M}} |\operatorname{Rm}_x(t)| \right) = \infty. \tag{5}$$

Proof. For a contradiction, we assume that $|\operatorname{Rm}_x(t)|$ is bounded by a constant. It follows from Corollary 1 that the metrics $g(t)$ converge uniformly in the norm induced by $g(t)$ to a smooth metric $g(T)$. Based on Theorem 2, it is possible to find a solution to the Ricci flow on $t \in [0, T)$ because the smooth metric $g(T)$ is uniformly equivalent to initial metric $g(0)$.

Hence, one can extend the solution of the Ricci flow after the time point $t = T$, which is the result for continuous derivatives at $t = T$. This tell us that the time T of existence of the Ricci flow has not been maximal, which contradicts our assumption. In other words, $|\operatorname{Rm}_x(t)|$ is unbounded. □

As approaching the singular time T, the Riemann curvature $|\operatorname{Rm}|_{g(t)}$ becomes no longer convergent and tends to explode.

3 Evolution of Linearly Nearly Euclidean Metrics

Next, this paper will focus on linearly nearly Euclidean metrics, proving that them can have a good performance in terms of stability, i.e., the convergence of a Ricci-DeTurck flow $\bar{g}(t)$ to a linearly

nearly Euclidean metric $\bar{g}(\infty)$. Before that, we have to construct a family $\bar{g}_0$ of linearly nearly Euclidean reference metrics such that $\frac{\partial}{\partial t}\bar{g}_0(t) = O((\bar{g}(t) - \bar{g}_0(t))^2)$. Let

$$\mathcal{F} = \left\{ \bar{g}(t) \in \mathcal{M}^n \;\middle|\; 2\,\mathrm{Ric}(\bar{g}(t)) + \mathcal{L}_{\frac{\partial \varphi(t)}{\partial t}} \bar{g}(t) = 0 \right\}$$

be the set of stationary points under the Ricci-DeTurck flow. We are able to establish a manifold

$$\tilde{\mathcal{F}} = \mathcal{F} \cap \mathcal{U} \tag{6}$$

where $\mathcal{U}$ is an L^2-neighbourhood of integral $\bar{g}_0$.

3.1 Analysis on Linearly Nearly Euclidean Metrics

Let us give the definition of linearly nearly Euclidean metrics without further explanation:

Definition 2 *A complete Riemannian n-manifold $(\mathcal{M}^n, g_0)$ is said to be linearly nearly Euclidean with one end of order $\tau > 0$ if there exists a compact set $K \subset \mathcal{M}$, a radius r, a point x in $\mathcal{M}$ and a diffeomorphism satisfying $\phi : \mathcal{M}\backslash K \to (\mathbb{R}^n \backslash B(x,r))/SO(n)$, where $B(x,r)$ is the ball and $SO(n)$ is a finite group acting freely on $\mathbb{R}^n\backslash\{0\}$, then*

$$\left|\partial^k(\phi_*\gamma_0)\right|_\delta = O(r^{-\tau-k}) \quad \forall k \geq 0 \tag{7}$$

holds on $(\mathbb{R}^n\backslash B(x,r))/SO(n)$. g_0 can be linearly decomposed into a form containing the Euclidean metric δ:

$$g_0(t) = \delta + \gamma_0(t). \tag{8}$$

In this paper, we consider the linear stability and integrability of the initial metric g_0. Fortunately, similar to the proof process of (Koiso, 1983; Besse, 2007), we can proceed that g_0 is integral and linearly stable.

Definition 3 *A complete linearly nearly Euclidean n-manifold $(\mathcal{M}^n, g_0)$ is said to be linearly stable if the L^2 spectrum of the Lichnerowicz operator $L_{g_0} := \Delta_{g_0} + 2\,\mathrm{Rm}(g_0)*$ is in $(-\infty, 0]$ where Δ_{g_0} is the Laplacian, when L_{g_0} acting on d_{ij} satisfies*

$$\begin{aligned} L_{g_0}(d) &= \Delta_{g_0} d + 2\,\mathrm{Rm}(g_0) * d \\ &= \Delta_{g_0} d + 2\,\mathrm{Rm}(g_0)_{iklj} d_{mn} g_0^{km} g_0^{ln}. \end{aligned}$$

Definition 4 *A n-manifold $(\mathcal{M}^n, g_0)$ is said to be integrable if a neighbourhood of g_0 has a smooth structure.*

3.2 Short Time Convergence in the L^2-Norm

For convenience, we rewrite the Ricci-DeTurck flow (3) in terms of the difference $d(t) := \bar{g}(t) - \bar{g}_0$, such that

$$\begin{aligned} \frac{\partial}{\partial t} d(t) = \frac{\partial}{\partial t}\bar{g}(t) &= -2\,\mathrm{Ric}(\bar{g}(t)) + 2\,\mathrm{Ric}(\bar{g}_0) + \mathcal{L}_{\frac{\partial \varphi'(t)}{\partial t}} \bar{g}_0 - \mathcal{L}_{\frac{\partial \varphi(t)}{\partial t}} \bar{g}(t) \\ &= \Delta d(t) + \mathrm{Rm} * d(t) + F_{\bar{g}^{-1}} * \nabla^{\bar{g}_0} d(t) * \nabla^{\bar{g}_0} d(t) + \nabla^{\bar{g}_0}\left(G_{\Gamma(\bar{g}_0)} * d(t) * \nabla^{\bar{g}_0} d(t)\right), \end{aligned} \tag{9}$$

where the tensors F and G depend on $\bar{g}^{-1}$ and $\Gamma(\bar{g}_0)$. Note that $\bar{g}_0$ is a linearly nearly Euclidean metric which satisfies the above formula, where $d_0(t) = \bar{g}_0(t) - \bar{g}_0$, so that $d(t) - d_0(t) = \bar{g}(t) - \bar{g}_0(t)$ holds. Note that $\|\cdot\|_{L^2}$ or $\|\cdot\|_{L^\infty}$ denotes the L^2-Norm or L^∞-Norm with respect to the metric $\bar{g}_0$.

Lemma 2 *Let $(\mathcal{M}^n, \bar{g}_0)$ be a complete linearly nearly Euclidean n-manifold. If $\bar{g}(0)$ is a metric satisfying $\|\bar{g}(0) - \bar{g}_0\|_{L^\infty} < \epsilon$ where $\epsilon > 0$, then there exist a constant $C < \infty$ and a unique Ricci–DeTurck flow $\bar{g}(t)$ that satisfies*

$$\|\bar{g}(t) - \bar{g}_0\|_{L^\infty} < C\|\bar{g}(0) - \bar{g}_0\|_{L^\infty} < C \cdot \epsilon.$$

If a Ricci-DeTurck flow in $\mathcal{B}_{L^\infty}(\bar{g}_0, \epsilon)$ for $t \geq 1$, there exist constants such that

$$\left\|\nabla^k\left(\bar{g}(t) - \bar{g}_0\right)\right\|_{L^\infty} < C(k)\epsilon, \quad \forall k \in \mathbb{N}.$$

Proof. The similar statement for the case of negative Einstein metrics is given in (Bamler, 2010). The proofs can be translated easily to the case of linearly nearly Euclidean metrics by referring the details (Bamler, 2011). □

Lemma 3 *Let $(\mathcal{M}^n, \bar{g}_0)$ be a linearly nearly Euclidean n-manifold. For a Ricci–DeTurck flow $\bar{g}(t)$ on a maximal time interval $t \in [0, T)$, if it satisfies $\|\bar{g}(0) - \bar{g}_0\|_{L^\infty} < \epsilon$ where $\epsilon > 0$, then there exists a constant $C < \infty$ for $t \in (0, T)$ such that*

$$\|\bar{g}(t) - \bar{g}_0\|_{L^2} < C. \tag{10}$$

Proof. Based on Lemma 2, we can consider about $\|\bar{g}(t) - \bar{g}_0\|_{L^2}$. Let κ be a function such that $\kappa = 1$ on $B(x, r)$, $\kappa = 0$ on $\mathcal{M}^n \backslash B(x, 2r)$ and $|\nabla \kappa| \leq 2/r$ where $x \in \mathcal{M}^n$ and a radius r.

Followed by Equation (9), we obtain

$$\begin{aligned}
\frac{\partial}{\partial t} \int_{\mathcal{M}} |d(t)|^2 \kappa^2 \mathrm{d}\mu \leq & 2 \int_{\mathcal{M}} \left\langle \Delta d(t), \kappa^2 d(t) \right\rangle \mathrm{d}\mu + C \| \operatorname{Rm} \|_{L^\infty} \int_{\mathcal{M}} |d(t)|^2 \kappa^2 \mathrm{d}\mu \\
& + C \|d(t)\|_{L^\infty} \int_{\mathcal{M}} |\nabla d(t)|^2 \kappa^2 \mathrm{d}\mu + \int_{\mathcal{M}} \langle \nabla (G_\Gamma * d * \nabla d), d \rangle \kappa^2 \mathrm{d}\mu \\
\leq & -2 \int_{\mathcal{M}} |\nabla d(t)|^2 \kappa^2 \mathrm{d}\mu + C \int_{\mathcal{M}} |\nabla d(t)| |d(t)| |\nabla \kappa| \kappa \mathrm{d}\mu \\
& + C(\bar{g}_0) \int_{\mathcal{M}} |d(t)|^2 \kappa^2 \mathrm{d}\mu + C \|d(t)\|_{L^\infty} \int_{\mathcal{M}} |\nabla d(t)|^2 \kappa^2 \mathrm{d}\mu \\
\leq & (-2 + C \cdot \epsilon + C_1) \int_{\mathcal{M}} |\nabla d(t)|^2 \kappa^2 \mathrm{d}\mu + C(\bar{g}_0) \int_{\mathcal{M}} |d(t)|^2 \kappa^2 \mathrm{d}\mu \\
& + \frac{1}{C_1} \int_{\mathcal{M}} |d(t)|^2 |\nabla \kappa|^2 \mathrm{d}\mu \\
\leq & \left(C(\bar{g}_0) + \frac{2}{C_1 r^2} \right) \int_{B(x,2r)} |d(t)|^2 \mathrm{d}\mu.
\end{aligned}$$

Note that we can always find a suitable C_1 to make the above formula true. By integration in time t, we can further obtain

$$\int_{\mathcal{M}} |d(t)|^2 \kappa^2 \mathrm{d}\mu \leq \int_{\mathcal{M}} |d(0)|^2 \kappa^2 \mathrm{d}\mu + \left(C(\bar{g}_0) + \frac{2}{C_1 r^2} \right) \int_0^t \int_{B(x,2r)} |d(s)|^2 \mathrm{d}\mu \mathrm{d}s < \infty.$$

Consequently, we can find a finite ball that satisfies this estimate. □

Corollary 2 *Based on Lemma 3, we further have*

$$\sup \int_{\mathcal{M}} |d(t)|^2 \kappa^2 \mathrm{d}\mu < \infty. \tag{11}$$

Proof. We obtain

$$\begin{aligned}
\sup \int_{\mathcal{M}} |d(t)|^2 \kappa^2 \mathrm{d}\mu \leq & \sup \int_{\mathcal{M}} |d(0)|^2 \kappa^2 \mathrm{d}\mu \\
& + N \left(C(\bar{g}_0) + \frac{2}{C_1 r^2} \right) \int_0^t \sup \int_{\mathcal{M}} |d(s)|^2 \kappa^2 \mathrm{d}\mu \mathrm{d}s,
\end{aligned}$$

where each ball of radius $2r$ on $\mathcal{M}$ can be covered by N balls of radius r because $(\mathcal{M}^n, \bar{g}_0)$ is linearly nearly Euclidean. By the Gronwall inequality, we have

$$\sup \int_{\mathcal{M}} |d(t)|^2 \kappa^2 \mathrm{d}\mu \leq \exp \left(N \left(C(\bar{g}_0) + \frac{2}{C_1 r^2} \right) t \right) \sup \int_{\mathcal{M}} |d(0)|^2 \kappa^2 \mathrm{d}\mu.$$

For the L^2-Norm, the Ricci-DeTurck flow in linearly nearly Euclidean manifolds has a solution for a short time. □

3.3 LONG TIME STABILITY IN THE L^2-NORM

Before starting the discussion about long time stability of linearly nearly Euclidean metrics, we need some prior knowledge:

Lemma 4 *Let $\bar{g}(t)$ be a Ricci–DeTurck flow on a maximal time interval $t \in (0, T)$ in an L^2 neighbourhood of $\bar{g}_0$. We have the following estimate such that:*

$$\left\| \frac{\partial}{\partial t} d_0(t) \right\|_{L^2} \leq C \left\| \nabla^{\bar{g}_0(t)} \left(d(t) - d_0 \right) \right\|_{L^2}^2 .$$

Proof. Let $\{e_1(t), e_2(t), \ldots, e_n(t)\}$ be a family of L^2-orthonormal bases of the kernel $\ker_{L^2}$ such that $\frac{\partial}{\partial t} e_i(t)$ depends linearly on $\frac{\partial}{\partial t} d_0(t)$. For an isomorphism orthogonal projection $\Pi : T_{\bar{g}_0}\tilde{\mathcal{F}} \to \ker_{L^2}$, by the Hardy inequality (Minerbe, 2009), one has similar proofs by referring the details (Deruelle & Kröncke, 2021). □

Theorem 5 *Let $(\mathcal{M}^n, \bar{g}_0)$ be a linearly nearly Euclidean n-manifold which is linearly stable and integrable. Furthermore, there exists a constant $\alpha_{\bar{g}_0}$ such that*

$$\left(\Delta d(t) + \mathrm{Rm}(\bar{g}_0) * d(t), d(t) \right)_{L^2} \leq -\alpha_{\bar{g}_0} \left\| \nabla^{\bar{g}_0} h \right\|_{L^2}^2$$

for all $\bar{g}(t) \in \tilde{\mathcal{F}}$ which is as in (6).

Proof. The similar proofs can be found in (Devyver, 2014) with some minor modifications. Due to the linear stability requirement of linearly nearly Euclidean manifolds in Definition 3, $-L_{\bar{g}_0}$ is non-negative. Then there exists a positive constant $\alpha_{\bar{g}_0}$ such that

$$\alpha_{\bar{g}_0} \left(-\Delta d(t), d(t) \right)_{L^2} \leq \left(-\Delta d(t) - \mathrm{Rm}(\bar{g}_0) * d(t), d(t) \right)_{L^2} .$$

By Taylor expansion, one repeatedly uses elliptic regularity and Sobolev embedding (Pacini, 2010) to obtain the estimate. □

Corollary 3 *Let $(\mathcal{M}^n, \bar{g}_0)$ be a linearly nearly Euclidean n-manifold which is integrable. For a Ricci–DeTurck flow $\bar{g}(t)$ on a maximal time interval $t \in [0, T]$, if it satisfies $\|\bar{g}(t) - \bar{g}_0\|_{L^\infty} < \epsilon$ where $\epsilon > 0$, then there exists a constant $C < \infty$ for $t \in [0, T]$ such that the evolution inequality satisfies*

$$\|d(t) - d_0\|_{L^2}^2 \geq C \int_0^T \left\| \nabla^{\bar{g}_0(t)} \left(d(t) - d_0 \right) \right\|_{L^2}^2 \mathrm{d}t.$$

Proof. Based on Equation (9), we know

$$\begin{aligned} \frac{\partial}{\partial t}(d(t) - d_0) =& \Delta(d(t) - d_0) + \mathrm{Rm} * (d(t) - d_0) \\ &+ F_{\bar{g}^{-1}} * \nabla^{\bar{g}_0}(d(t) - d_0) * \nabla^{\bar{g}_0}(d(t) - d_0) \\ &+ \nabla^{\bar{g}_0} \left(G_{\Gamma(\bar{g}_0)} * (d(t) - d_0) * \nabla^{\bar{g}_0}(d(t) - d_0) \right) . \end{aligned}$$

Followed by Lemma 4 and Theorem 5, we further obtain

$$\begin{aligned} \frac{\partial}{\partial t} \|d(t) - d_0\|_{L^2}^2 =& 2 \left(\Delta(d(t) - d_0) + \mathrm{Rm} * (d(t) - d_0), d(t) - d_0 \right)_{L^2} \\ &+ \left(F_{\bar{g}^{-1}} * \nabla^{\bar{g}_0}(d(t) - d_0) * \nabla^{\bar{g}_0}(d(t) - d_0), d(t) - d_0 \right)_{L^2} \\ &+ \left(\nabla^{\bar{g}_0} \left(G_{\Gamma(\bar{g}_0)} * (d(t) - d_0) * \nabla^{\bar{g}_0}(d(t) - d_0) \right), d(t) - d_0 \right)_{L^2} \\ &+ \left(d(t) - d_0, \frac{\partial}{\partial t} d_0(t) \right)_{L^2} + \int_{\mathcal{M}} (d(t) - d_0) * (d(t) - d_0) * \frac{\partial}{\partial t} d_0(t) \mathrm{d}\mu \\ \leq& -2\alpha_{\bar{g}_0} \left\| \nabla^{\bar{g}_0} \left(d(t) - d_0 \right) \right\|_{L^2}^2 \\ &+ C \left\| (d(t) - d_0) \right\|_{L^\infty} \left\| \nabla^{\bar{g}_0} \left(d(t) - d_0 \right) \right\|_{L^2}^2 \\ &+ \left\| \frac{\partial}{\partial t} d_0(t) \right\|_{L^2} \|d(t) - d_0\|_{L^2} \\ \leq& \left(-2\alpha_{\bar{g}_0} + C \cdot \epsilon \right) \left\| \nabla^{\bar{g}_0} \left(d(t) - d_0 \right) \right\|_{L^2}^2 . \end{aligned}$$

Let ϵ be small enough that $-2\alpha_{\bar{g}_0} + C \cdot \epsilon < 0$ holds, we can find

$$\frac{\partial}{\partial t}\|d(t) - d_0\|_{L^2}^2 \leq -C \left\|\nabla^{\bar{g}_0}\left(d(t) - d_0\right)\right\|_{L^2}^2$$

holds. □

Theorem 6 *Let $(\mathcal{M}^n, \bar{g}_0)$ be a linearly nearly Euclidean n-manifold which is linearly stable and integrable. For every $\epsilon_1 > 0$, there exists a $\epsilon_2 > 0$ satisfying: For any metric $\bar{g}(t) \in \mathcal{B}_{L^2}(\bar{g}_0, \epsilon_2)$, there is a complete Ricci–DeTurck flow $(\mathcal{M}^n, \bar{g}(t))$ starting from $\bar{g}(t)$ converging to a linearly nearly Euclidean metric $\bar{g}(\infty) \in \mathcal{B}_{L^2}(\bar{g}_0, \epsilon_1)$. Note that $\mathcal{B}_{L^2}(\bar{g}_0, \epsilon)$ is the ϵ-ball with respect to the L^2-Norm induced by $\bar{g}_0$ and centred at $\bar{g}_0$.*

Proof. By Lemma 2, one can find so small $\epsilon_2 > 0$ such that $d(t) \in \mathcal{B}_{L^2}(0, \epsilon_2)$ holds. By Lemma 4 and Corollary 3, we have

$$\begin{aligned}
\|d_0(T)\|_{L^2} &\leq C \int_1^T \left\|\frac{\partial}{\partial t} d_0(t)\right\|_{L^2} \mathrm{d}t \\
&\leq C \int_1^T \left\|\nabla^{\bar{g}_0}\left(d(t) - d_0(t)\right)\right\|_{L^2}^2 \mathrm{d}t \\
&\leq C \left\|d(1) - d_0(1)\right\|_{L^2}^2 \leq C\|d(1)\|_{L^2}^2 \leq C \cdot (\epsilon_2)^2 .
\end{aligned}$$

Furthermore, we obtain

$$\|d(T) - d_0(T)\|_{L^2} \leq \|d(1) - d_0(1)\|_{L^2} \leq C \cdot \epsilon_2.$$

By the triangle inequality, we get

$$\|d(T)\|_{L^2} \leq C \cdot (\epsilon_2)^2 + C \cdot \epsilon_2.$$

Followed by Corollary 2 and Lemma 4, such T should be pushed further outward, because

$$\lim_{t \to +\infty} \sup \left\|\frac{\partial}{\partial t} d_0(t)\right\|_{L^2} \leq \lim_{t \to +\infty} \sup \left\|\nabla^{\bar{g}_0}\left(d(t) - d_0(t)\right)\right\|_{L^2}^2 = 0.$$

Thus, as t approaches to $+\infty$, $\bar{g}(t)$ converges to $\bar{g}(\infty) = \bar{g}_0 + d_0(\infty)$. By the Euclidean Sobolev inequality (Minerbe, 2009), $d(t) - d_0(t)$ converges to 0 as t goes to $+\infty$,

$$\lim_{t \to +\infty} \|d(t) - d_0(t)\|_{L^2} \leq \lim_{t \to +\infty} C \left\|\nabla^{\bar{g}_0}\left(d(t) - d_0(t)\right)\right\|_{L^2} = 0.$$

We now conclude a result for linearly nearly Euclidean manifolds under the Ricci-DeTurck flow, which ensures a infinite time existence. □

4 Gradient Flow under the Ricci-DeTurck flow

Above, we have clarified the convergence of linearly nearly Euclidean manifolds under the Ricci-DeTurck flow. Furthermore, we will consider the solution of gradient flow. Empirically, we introduce information geometry (Amari & Nagaoka, 2000; Amari, 2016) and mirror descent algorithm (Bubeck et al., 2015) to approximate the gradient flow.

4.1 A Linearly Nearly Euclidean Metric

From the perspective of information geometry and mirror descent algorithm, the metric $\bar{g}$ can be deduced by the divergence that needs to satisfy certain criteria (Basseville, 2013). We now consider two nearby points P and Q in a manifold $\mathcal{M}$, where these two points are expressed in coordinates as $\boldsymbol{\xi}_P$ and $\boldsymbol{\xi}_Q$, where $\boldsymbol{\xi}$ is a column vector. Moreover, the divergence is defined as half the square of an infinitesimal distance between two sufficiently close points in Definition 5.

Definition 5 *$D[P : Q]$ is called a divergence when it satisfies the following criteria:*

(1) $D[P:Q] \geq 0$, (2) $D[P:Q] = 0$ when and only when $P = Q$, (3) When P and Q are sufficiently close, by denoting their coordinates by $\boldsymbol{\xi}_P$ and $\boldsymbol{\xi}_Q = \boldsymbol{\xi}_P + d\boldsymbol{\xi}$, the Taylor expansion of D is written as

$$D[\boldsymbol{\xi}_P : \boldsymbol{\xi}_P + d\boldsymbol{\xi}] = \frac{1}{2}\sum_{i,j} \bar{g}_{ij}(\boldsymbol{\xi}_P) d\xi_i d\xi_j + O(|d\boldsymbol{\xi}|^3),$$

and metric $\bar{g}_{ij}$ is positive-definite, depending on $\boldsymbol{\xi}_P$.

Furthermore, we construct such divergences by assist of Definition 6. We introduce a symmetrized convex function:

$$\phi(\boldsymbol{\xi}) = \sum_i \frac{1}{\tau^2} \log \frac{1}{2}\left(\exp(\tau\xi_i) + \exp(-\tau\xi_i)\right) = \sum_i \frac{1}{\tau^2} \log\left(\cosh(\tau\xi_i)\right) \tag{12}$$

where τ is a constant parameter. Note that we omit the construction process of the convex function, which is inspired by the Definition 7.

Definition 6 *The Bregman divergence (Bregman, 1967) $D_B[\boldsymbol{\xi} : \boldsymbol{\xi}']$ is defined as the difference between a convex function $\phi(\boldsymbol{\xi})$ and its tangent hyperplane $z = \phi(\boldsymbol{\xi}') + (\boldsymbol{\xi} - \boldsymbol{\xi}')\nabla\phi(\boldsymbol{\xi}')$, depending on the Taylor expansion at the point $\boldsymbol{\xi}'$:*

$$D_B[\boldsymbol{\xi} : \boldsymbol{\xi}'] = \phi(\boldsymbol{\xi}) - \phi(\boldsymbol{\xi}') - (\boldsymbol{\xi} - \boldsymbol{\xi}')\nabla\phi(\boldsymbol{\xi}').$$

Definition 7 *Let $\mathcal{M}$ be a Riemannian manifold, for the tangent vector $\boldsymbol{v} \in T_x\mathcal{M}$ in a point $x \in \mathcal{M}$ where $T_x\mathcal{M}$ is the tangent space, there is a unique geodesic $\gamma_{\boldsymbol{v}}(t)$ locally that satisfies $\gamma_{\boldsymbol{v}}(0) = x$ and $\gamma'_{\boldsymbol{v}}(0) = \boldsymbol{v}$. The exponential map $\exp_x : T_x\mathcal{M} \mapsto \mathcal{M}$ corresponding to $\gamma_{\boldsymbol{v}}(t)$ is defined as $\exp_x(\boldsymbol{v}) = \gamma_{\boldsymbol{v}}(1)$.*

Theorem 7 *For a convex function ϕ defined by Equation (12), the linearly nearly Euclidean divergence between two points $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ is*

$$D_{LNE}[\boldsymbol{\xi}' : \boldsymbol{\xi}] = \sum_i \left[\frac{1}{\tau^2}\log\frac{\cosh(\tau\xi'_i)}{\cosh(\tau\xi_i)} - \frac{1}{\tau}(\xi'_i - \xi_i)\tanh(\tau\xi_i)\right] \tag{13}$$

where the Riemannian metric is

$$\begin{aligned}\bar{g}_{ij}(\boldsymbol{\xi}(t)) &= \delta_{ij} - \left[\tanh(\tau\boldsymbol{\xi})\tanh(\tau\boldsymbol{\xi})^\top\right]_{ij} \\ &= \begin{bmatrix} 1 - \tanh(\tau\xi_1(t))\tanh(\tau\xi_1(t)) & \cdots & -\tanh(\tau\xi_1(t))\tanh(\tau\xi_n(t)) \\ \vdots & \ddots & \vdots \\ -\tanh(\tau\xi_n(t))\tanh(\tau\xi_1(t)) & \cdots & 1 - \tanh(\tau\xi_n(t))\tanh(\tau\xi_n(t)) \end{bmatrix}.\end{aligned} \tag{14}$$

Proof. The proofs can be found in Appendix C.1. □

By Theorem 7, the form of the metrics $\bar{g}(t)$ is consistent with the definition of linearly nearly Euclidean metrics. We only need to adjust parameter τ to satisfy Definition 2. Moreover, we can prove that the linearly nearly Euclidean divergence satisfies the criteria of divergence followed by Definition 5.

4.2 Weak Approximation of the Special Case

With the form of linearly nearly Euclidean metrics, we can perform micro-surgery under the Ricci-DeTurck flow on the target manifold. Moreover, we dynamically consider the gradient flow followed with the optimal descent direction on this manifold.

Lemma 5 *The steepest descent gradient flow measured by the linearly nearly Euclidean divergence is defined as*

$$\tilde{\partial}_{\boldsymbol{\xi}} = \bar{g}^{-1}(t)\partial_{\boldsymbol{\xi}} = \left[\delta_{ij} - \tanh(\tau\boldsymbol{\xi}(t))\tanh(\tau\boldsymbol{\xi}(t))^\top\right]^{-1}\partial_{\boldsymbol{\xi}}. \tag{15}$$

Proof. The proofs can be found in Appendix C.2. □

In particular, we put forward higher requirements for this metric on the basis of Definition 2, i.e., it needs to be a strictly diagonally-dominant matrix. Therefore, this gradient flow is a weak approximation under the manifold micro-surgery.

Corollary 4 *The weak approximation of the gradient flow measured by the linearly nearly Euclidean divergence is defined as*

$$\tilde{\partial}_{\boldsymbol{\xi}} \approx \left[\delta_{ij} + \tanh(\tau\boldsymbol{\xi}(t))\tanh(\tau\boldsymbol{\xi}(t))^{\top}\right]\partial_{\boldsymbol{\xi}} \tag{16}$$

if the metric satisfies strictly diagonally-dominant.

Proof. The proofs can be found in Appendix C.3. □

Remark: For a neural network that is specified by connection weights, the set of all such weighs forms a manifold. When we use the gradient flow to learn a neural network ($\boldsymbol{\xi}(t)$ is composed of weights), we observe the evolution of its metric is consistent with the micro-surgery process under the Ricci-DeTurck flow (see Appendix C.4). Consequently, the training of a neural manifold is also a surgery, i.e., the manifold is gradually regular, whose process is stable and eventually converges.

4.3 Strong Approximation with Neural Networks

Bypassing the requirement of weak approximation in Corollary 4, our goal is to approximate the gradient flow, $\bar{g}^{-1}(t)\partial_{\boldsymbol{\xi}}$ in Lemma 5, from the assist of multi-layer perceptron (MLP) neural network because a neural network with a single hidden layer and a finite number of neurons can be used to approximate a continuous function on compact subsets (Jejjala et al., 2020), which is stated by the universal approximation theorem (Cybenko, 1989; Hornik, 1991).

As an $n \times n$ symmetric matrix, $\bar{g}(t)$ can be decomposed in terms of the combination of entries P and A, where P is the entries made up of the elements of the lower triangular matrix that contains $n(n-1)/2$ real parameters and A is the entries made up of the elements of the diagonal matrix that contains n real parameters. During the training in Figure 1, $\tilde{g}(t)$ can be used as strong approximation of $\bar{g}^{-1}(t)$ in the gradient flow.

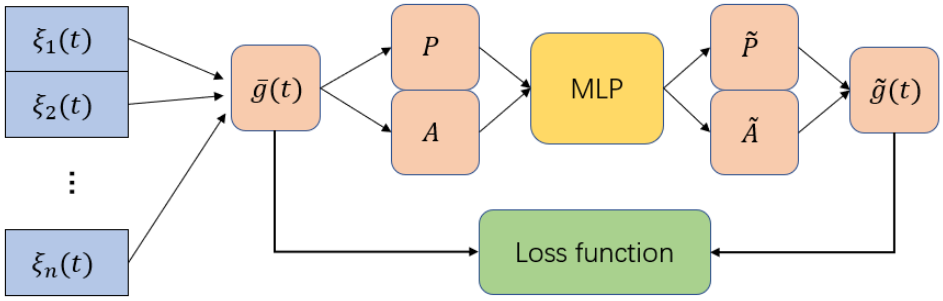


Figure 1: Flow chart of strong approximation. The new entries $\tilde{P}$ and $\tilde{A}$ produced by neural network get combined into a new metric $\tilde{g}(t)$ that is used to minimize the loss function by combining with the metric $\bar{g}(t)$, where the loss function is defined by Equation (17).

$$\mathbb{L} = \|\boldsymbol{I} - \bar{g}(t)\tilde{g}(t)\|^2. \tag{17}$$

5 Conclusion

In this paper, we have analysed the evolution of linearly nearly Euclidean metrics under the Ricci-DeTurck flow, including proof of convergence in short and infinite time. Furthermore, we construct a linearly nearly Euclidean metric with the assist of the information geometry and use it as a springboard to reach the approximation of gradient flow. This kind of stable micro-surgery under the Ricci-DeTurck flow can be used to optimize n-dimensional manifolds in terms of geometry, e.g., assisting the training of complex and redundant neural networks (Martens & Grosse, 2015). We hope that this paper will open exciting future directions for behavioral analysis of deep learning with geometric methods.

REFERENCES

S-i Amari and H Nagaoka. Methods of information geometry, volume 191 of translations of mathematical monographs, s. kobayashi and m. takesaki, editors. *American Mathematical Society, Providence, RI, USA*, pp. 2–19, 2000.

Shun-ichi Amari. *Information geometry and its applications*, volume 194. Springer, 2016.

Alexander Appleton. Scalar curvature rigidity and ricci deturck flow on perturbations of euclidean space. *Calculus of Variations and Partial Differential Equations*, 57(5):1–23, 2018.

Richard H Bamler. Stability of hyperbolic manifolds with cusps under ricci flow. *arXiv preprint arXiv:1004.2058*, 2010.

Richard Heiner Bamler. *Stability of Einstein metrics of negative curvature*. Princeton University, 2011.

Michéle Basseville. Divergence measures for statistical data processing—an annotated bibliography. *Signal Processing*, 93(4):621–633, 2013.

Arthur L Besse. *Einstein manifolds*. Springer Science & Business Media, 2007.

Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

Alix Deruelle and Klaus Kröncke. Stability of ale ricci-flat manifolds under ricci flow. *The Journal of Geometric Analysis*, 31(3):2829–2870, 2021.

Dennis M DeTurck. Deforming metrics in the direction of their ricci tensors. *Journal of Differential Geometry*, 18(1):157–162, 1983.

Baptiste Devyver. A gaussian estimate for the heat kernel on differential forms and application to the riesz transform. *Mathematische Annalen*, 358(1):25–68, 2014.

Richard S Hamilton et al. Three-manifolds with positive ricci curvature. *J. Differential geom*, 17 (2):255–306, 1982.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4 (2):251–257, 1991.

Vishnu Jejjala, Damian Kaloni Mayorga Pena, and Challenger Mishra. Neural network approximations for calabi-yau metrics. *arXiv preprint arXiv:2012.15821*, 2020.

Herbert Koch and Tobias Lamm. Geometric flows with rough initial data. *Asian Journal of Mathematics*, 16(2):209–235, 2012.

Norihito Koiso. Einstein metrics and complex structures. *Inventiones mathematicae*, 73(1):71–106, 1983.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Olga Aleksandrovna Ladyzhenskaia, Vsevolod Alekseevich Solonnikov, and Nina N Ural'tseva. *Linear and quasi-linear equations of parabolic type*, volume 23. American Mathematical Soc., 1988.

James Martens and Roger Grosse. Optimizing neural networks with kronecker-factored approximate curvature. In *International conference on machine learning*, pp. 2408–2417, 2015.

Vincent Minerbe. Weighted sobolev inequalities and ricci flat manifolds. *Geometric and Functional Analysis*, 18(5):1696–1749, 2009.

Tommaso Pacini. Desingularizing isolated conical singularities: uniform estimates via weighted sobolev spaces. *arXiv preprint arXiv:1005.3511*, 2010.

Oliver C Schnürer, Felix Schulze, and Miles Simon. Stability of euclidean space under ricci flow. *arXiv preprint arXiv:0706.0421*, 2007.

Natasa Sesum. Linear and dynamical stability of ricci-flat metrics. *Duke Mathematical Journal*, 133 (1):1–26, 2006.

Nick Sheridan and Hyam Rubinstein. Hamilton's ricci flow. *Honour thesis*, 2006.

Robert M Wald. *General relativity*. University of Chicago press, 2010.

A DIFFERENTIAL GEOMETRY

1. Riemann curvature tensor (Rm) is a (1,3)-tensor defined for a 1-form ω:

$$R^l_{ijk}\omega_l = \nabla_i\nabla_j\omega_k - \nabla_j\nabla_i\omega_k$$

where the covariant derivative of F satisfies

$$\nabla_p F^{j_1\dots j_l}_{i_1\dots i_k} = \partial_p F^{j_1\dots j_l}_{i_1\dots i_k} + \sum_{s=1}^{l} F^{j_1\dots q\dots j_l}_{i_1\dots i_k}\Gamma^{j_s}_{pq} - \sum_{s=1}^{k} F^{j_1\dots j_l}_{i_1\dots q\dots i_k}\Gamma^{q}_{pi_s}.$$

In particular, coordinate form of the Riemann curvature tensor is:

$$R^l_{ijk} = \partial_i\Gamma^l_{jk} - \partial_j\Gamma^l_{ik} + \Gamma^p_{jk}\Gamma^l_{ip} - \Gamma^p_{ik}\Gamma^l_{jp}.$$

2. Christoffel symbol in terms of an ordinary derivative operator is:

$$\Gamma^k_{ij} = \frac{1}{2}g^{kl}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}).$$

3. Ricci curvature tensor (Ric) is a (0,2)-tensor:

$$R_{ij} = R^p_{pij}.$$

4. Scalar curvature is the trace of the Ricci curvature tensor:

$$R = g^{ij}R_{ij}.$$

5. Lie derivative of F in the direction $\frac{d\varphi(t)}{dt}$:

$$\mathcal{L}_{\frac{d\varphi(t)}{dt}}F = \left(\frac{d}{dt}\varphi^*(t)F\right)_{t=0}$$

where $\varphi(t) : \mathcal{M} \to \mathcal{M}$ for $t \in (-\epsilon, \epsilon)$ is a time-dependent diffeomorphism of $\mathcal{M}$ to $\mathcal{M}$.

B PROOF FOR THE RICCI FLOW

B.1 PROOF FOR LEMMA 1

Definition 8 *The linearization of the Ricci curvature tensor is given by*

$$D[\mathrm{Ric}](h)_{ij} = -\frac{1}{2}g^{pq}(\nabla_p\nabla_q h_{ij} + \nabla_i\nabla_j h_{pq} - \nabla_q\nabla_i h_{jp} - \nabla_q\nabla_j h_{ip}).$$

Proof. Based on Appendix A, we have

$$\nabla_q\nabla_i h_{jp} = \nabla_i\nabla_q h_{jp} - R^r_{qij}h_{rp} - R^r_{qip}h_{jm}.$$

Combining with Definition 8, we can obtain the deformation equation because of $\nabla_k g_{ij} = 0$,

$$\begin{aligned} D[-2\mathrm{Ric}](h)_{ij} =& g^{pq}\nabla_p\nabla_q h_{ij} + \nabla_i\left(\frac{1}{2}\nabla_j h_{pq} - \nabla_q h_{jp}\right) + \nabla_j\left(\frac{1}{2}\nabla_i h_{pq} - \nabla_q h_{ip}\right) + O(h_{ij}) \\ =& g^{pq}\nabla_p\nabla_q h_{ij} + \nabla_i V_j + \nabla_j V_i + O(h_{ij}). \end{aligned}$$

□

B.2 DESCRIPTION OF THE DETURCK TRICK

Using a time-dependent diffeomorphism $\varphi(t)$, we express the pullback metrics $\bar{g}(t)$:

$$g(t) = \varphi^*(t)\bar{g}(t)$$

is a solution of the Ricci flow. Based on the chain rule for the Lie derivative in Appendix A, we can calculate

$$\begin{aligned}
\frac{\partial}{\partial t} g(t) &= \frac{\partial\left(\varphi^*(t)\bar{g}(t)\right)}{\partial t} \\
&= \left(\frac{\partial\left(\varphi^*(t+\tau)\bar{g}(t+\tau)\right)}{\partial \tau}\right)_{\tau=0} \\
&= \left(\varphi^*(t)\frac{\partial \bar{g}(t+\tau)}{\partial \tau}\right)_{\tau=0} + \left(\frac{\partial\left(\varphi^*(t+\tau)\bar{g}(t)\right)}{\partial \tau}\right)_{\tau=0} \\
&= \varphi^*(t)\frac{\partial}{\partial t}\bar{g}(t) + \varphi^*(t)\mathcal{L}_{\frac{\partial \varphi(t)}{\partial t}}\bar{g}(t).
\end{aligned}$$

With the help of Equation (1), for the reparameterized metric, we have

$$\frac{\partial}{\partial t} g(t) = \varphi^*(t)\frac{\partial}{\partial t}\bar{g}(t) + \varphi^*(t)\mathcal{L}_{\frac{\partial \varphi(t)}{\partial t}}\bar{g}(t) = -2\operatorname{Ric}(\varphi^*(t)\bar{g}(t)) = -2\varphi^*(t)\operatorname{Ric}(\bar{g}(t)).$$

The diffeomorphism invariance of the Ricci curvature tensor is used in the last step. The above equation is equivalent to

$$\frac{\partial}{\partial t}\bar{g}(t) = -2\operatorname{Ric}(\bar{g}(t)) - \mathcal{L}_{\frac{\partial \varphi(t)}{\partial t}}\bar{g}(t).$$

B.3 PROOF FOR THEOREM 3

Considering any $x \in \mathcal{M}$, $t_0 \in [0, T)$, $V \in T_x\mathcal{M}$, we have

$$\begin{aligned}
\left|\log\left(\frac{g_x(t_0)(V,V)}{g_x(0)(V,V)}\right)\right| &= \left|\int_0^{t_0} \frac{\partial}{\partial t}\left[\log g_x(t)(V,V)\right] dt\right| \\
&= \left|\int_0^{t_0} \frac{\frac{\partial}{\partial t} g_x(t)(V,V)}{g_x(t)(V,V)} dt\right| \\
&\leq \int_0^{t_0} \left|\frac{\partial}{\partial t} g_x(t)\left(\frac{V}{|V|_{g(t)}}, \frac{V}{|V|_{g(t)}}\right)\right| dt \\
&\leq \int_0^{t_0} \left|\frac{\partial}{\partial t} g_x(t)\right|_{g(t)} dt \\
&\leq C.
\end{aligned}$$

By exponentiating both sides of the above inequality, we have

$$e^{-C} g_x(0)(V,V) \leq g_x(t_0)(V,V) \leq e^{C} g_x(0)(V,V).$$

This inequality can be rewritten as

$$e^{-C} g_x(0) \leq g_x(t_0)(V,V) \leq e^{C} g_x(0)(V,V)$$

because it holds for any V. Thus, the metrics $g(t)$ are uniformly equivalent to $g(0)$.

Now, we have the well-defined integral:

$$g_x(T) - g_x(0) = \int_0^T \frac{\partial}{\partial t} g_x(t) dt.$$

We say that this integral is well-defined because of two reasons. Firstly, as long as the metrics are smooth, the integral exists. Secondly, the integral is absolutely integrable. Based on the norm inequality induced by $g(0)$, one has

$$|g_x(T) - g_x(t)|_{g(0)} \leq \int_t^T \left|\frac{\partial}{\partial t} g_x(t)\right|_{g(0)} dt.$$

For each $x \in \mathcal{M}$, the above integral will approach to zero as t approaches T. Because $\mathcal{M}$ is compact, the metrics $g(t)$ converge uniformly to a continuous metric $g(T)$ which is uniformly equivalent to $g(0)$ on $\mathcal{M}$. Moreover, we can show that

$$e^{-C} g_x(0) \leq g_x(T) \leq e^{C} g_x(0).$$

C Proof for the Information Geometry

C.1 Proof for Theorem 7

The linearly nearly Euclidean divergence can be defined between two nearby points $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$, where the first derivative of the linearly nearly Euclidean divergence w.r.t. $\boldsymbol{\xi}'$ is:

$$\begin{aligned}
&\partial_{\boldsymbol{\xi}'} D_{LNE}[\boldsymbol{\xi}' : \boldsymbol{\xi}] \\
&= \sum_i \left[\partial_{\boldsymbol{\xi}'} \frac{1}{\tau^2} \log\cosh(\tau \xi_i') - \partial_{\boldsymbol{\xi}'} \frac{1}{\tau^2} \log\cosh(\tau \xi_i) - \frac{1}{\tau} \partial_{\boldsymbol{\xi}'} (\xi_i' - \xi_i) \tanh(\tau \xi_i) \right] \\
&= \sum_i \partial_{\boldsymbol{\xi}'} \frac{1}{\tau^2} \log\cosh(\tau \xi_i') - \frac{1}{\tau} \tanh(\tau \boldsymbol{\xi}).
\end{aligned}$$

The second derivative of the linearly nearly Euclidean divergence w.r.t. $\boldsymbol{\xi}'$ is:

$$\partial^2_{\boldsymbol{\xi}'} D_{LNE}[\boldsymbol{\xi}' : \boldsymbol{\xi}] = \sum_i \partial^2_{\boldsymbol{\xi}'} \frac{1}{\tau^2} \log\cosh(\tau \xi_i').$$

We deduce the Taylor expansion of the linearly nearly Euclidean divergence at $\boldsymbol{\xi}' = \boldsymbol{\xi}$:

$$\begin{aligned}
D_{LNE}[\boldsymbol{\xi}' : \boldsymbol{\xi}] \approx\ & D_{LNE}[\boldsymbol{\xi} : \boldsymbol{\xi}] + \left(\sum_i \partial_{\boldsymbol{\xi}'} \frac{1}{\tau^2} \log\cosh(\tau \xi_i') - \frac{1}{\tau} \tanh(\tau \boldsymbol{\xi}) \right)^{\top} \Bigg|_{\boldsymbol{\xi}'=\boldsymbol{\xi}} d\boldsymbol{\xi} \\
& + \frac{1}{2} d\boldsymbol{\xi}^{\top} \left(\sum_i \partial^2_{\boldsymbol{\xi}'} \frac{1}{\tau^2} \log\cosh(\tau \xi_i') \right) \Bigg|_{\boldsymbol{\xi}'=\boldsymbol{\xi}} d\boldsymbol{\xi} \\
&= 0 + 0 + \frac{1}{2\tau^2} d\boldsymbol{\xi}^{\top} \partial \left[\frac{\partial \cosh(\tau \boldsymbol{\xi})}{\cosh(\tau \boldsymbol{\xi})} \right] d\boldsymbol{\xi} \\
&= \frac{1}{2\tau^2} d\boldsymbol{\xi}^{\top} \frac{\partial^2 \cosh(\tau \boldsymbol{\xi}) \cosh(\tau \boldsymbol{\xi}) - \partial \cosh(\tau \boldsymbol{\xi}) \partial \cosh(\tau \boldsymbol{\xi})^{\top}}{\cosh^2(\tau \boldsymbol{\xi})} d\boldsymbol{\xi} \\
&= \frac{1}{2\tau^2} d\boldsymbol{\xi}^{\top} \left(\frac{\partial^2 \cosh(\tau \boldsymbol{\xi})}{\cosh(\tau \boldsymbol{\xi})} - \tau^2 \left[\frac{\sinh(\tau \boldsymbol{\xi})}{\cosh(\tau \boldsymbol{\xi})} \right] \left[\frac{\sinh(\tau \boldsymbol{\xi})}{\cosh(\tau \boldsymbol{\xi})} \right]^{\top} \right) d\boldsymbol{\xi} \\
&= \frac{1}{2} \sum_{i,j} \delta_{ij} - \left[\tanh(\tau \boldsymbol{\xi}) \tanh(\tau \boldsymbol{\xi})^{\top} \right]_{ij} d\xi_i d\xi_j.
\end{aligned}$$

C.2 Proof for Lemma 5

We would to know in which direction minimizes the loss function with the constraint of the linearly nearly Euclidean divergence, so that we do the minimization:

$$d\boldsymbol{\xi}^* = \underset{d\boldsymbol{\xi} \text{ s.t. } D_{LNE}[\boldsymbol{\xi} : \boldsymbol{\xi} + d\boldsymbol{\xi}] = c}{\arg\min} \mathbb{L}(\boldsymbol{\xi} + d\boldsymbol{\xi})$$

where c is the constant. The loss function descends along the manifold with constant speed, regardless the curvature.

Now, we write the minimization in Lagrangian form. Combined with Theorem 7, the linearly nearly Euclidean divergence can be approximated by its second order Taylor expansion. Approximating $\mathbb{L}(\boldsymbol{\xi} + d\boldsymbol{\xi})$ with it first order Taylor expansion, we get:

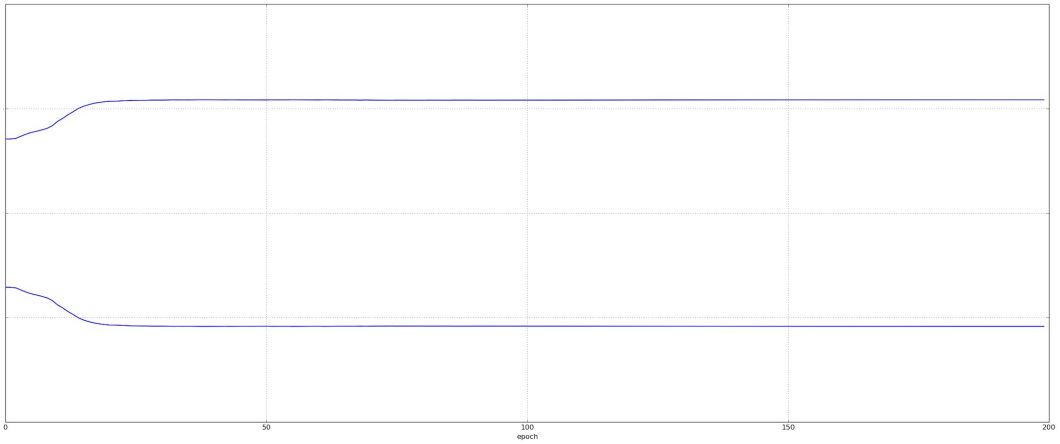
$$\begin{aligned}
d\boldsymbol{\xi}^* &= \arg\min_{d\boldsymbol{\xi}} \mathbb{L}(\boldsymbol{\xi} + d\boldsymbol{\xi}) + \lambda \left(D_{LNE}[\boldsymbol{\xi} : \boldsymbol{\xi} + d\boldsymbol{\xi}] - c \right) \\
&\approx \arg\min_{d\boldsymbol{\xi}} \mathbb{L}(\boldsymbol{\xi}) + \partial_{\boldsymbol{\xi}} \mathbb{L}(\boldsymbol{\xi})^{\top} d\boldsymbol{\xi} + \frac{\lambda}{2} d\boldsymbol{\xi}^{\top} \bar{g}_{ij}(t) d\boldsymbol{\xi} - c\lambda.
\end{aligned}$$

To solve this minimization, we set its derivative w.r.t. $d\boldsymbol{\xi}$ to zero:

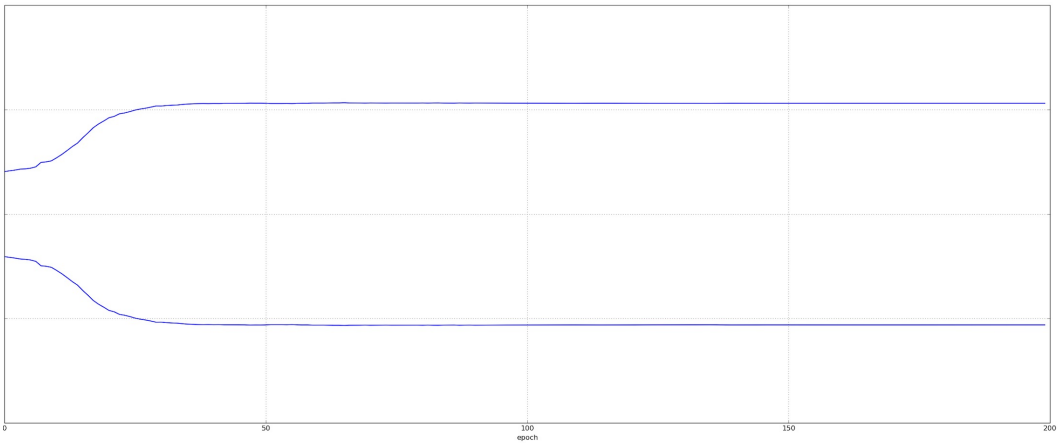
$$\begin{aligned} 0 &= \frac{\partial}{\partial d\boldsymbol{\xi}}\mathbb{L}(\boldsymbol{\xi}) + \partial_{\boldsymbol{\xi}}\mathbb{L}(\boldsymbol{\xi})^\top d\boldsymbol{\xi} + \frac{\lambda}{2} d\boldsymbol{\xi}^\top \left[\delta_{ij} - \tanh(\tau\boldsymbol{\xi}(t))\tanh(\tau\boldsymbol{\xi}(t))^\top\right] d\boldsymbol{\xi} - c\lambda \\ &= \partial_{\boldsymbol{\xi}}\mathbb{L}(\boldsymbol{\xi}) + \lambda\left[\delta_{ij} - \tanh(\tau\boldsymbol{\xi}(t))\tanh(\tau\boldsymbol{\xi}(t))^\top\right] d\boldsymbol{\xi} \\ d\boldsymbol{\xi} &= -\frac{1}{\lambda}\left[\delta_{ij} - \tanh(\tau\boldsymbol{\xi}(t))\tanh(\tau\boldsymbol{\xi}(t))^\top\right]^{-1} \partial_{\boldsymbol{\xi}}\mathbb{L}(\boldsymbol{\xi}) \end{aligned}$$

where a constant factor $1/\lambda$ can be absorbed into learning rate. Up to now, we get the optimal descent direction, i.e., the opposite direction of gradient which takes into account the local curvature defined by $\left[\delta_{ij} - \tanh(\tau\boldsymbol{\xi}(t))\tanh(\tau\boldsymbol{\xi}(t))^\top\right]^{-1}$.

C.3 Proof for Corollary 4



(a) ResNet20 on CIFAR10



(b) ResNet56 on CIFAR100

Figure 2: The evolution of metrics $\bar{g}(t)$ with the epoch of training process.

Definition 9 *For $\boldsymbol{A} \in \mathcal{R}^{n\times n}$, $\boldsymbol{A}$ is called as the strictly diagonally-dominant matrix when it satisfies*

$$\left|a_{ii}\right| > \sum_{j=1, j\neq i}^{n} \left|a_{ij}\right|, \quad i = 1, 2, \ldots, n.$$

Definition 10 *If $\boldsymbol{A} \in \mathcal{R}^{n\times n}$ is a strictly diagonally-dominant matrix, then $\boldsymbol{A}$ is a nonsingular matrix together.*

Subsequently, we can consider the inverse matrix of the metric $\bar{g}(t)$. Due to the strictly diagonally-dominant feature in Definition 9 and Definition 10, we can approximate $\left[\delta_{ij} - \tanh(\tau\boldsymbol{\xi}(t))\tanh(\tau\boldsymbol{\xi}(t))^\top\right]^{-1}$. As for we can also ignore the fourth-order small quantity $\sum O(\rho_a\rho_b\rho_c\rho_d)$ because of the characteristic of the strictly diagonally-dominant, so that

$$
\begin{aligned}
&\left[\delta_{ij} - \tanh(\tau\boldsymbol{\xi}(t))\tanh(\tau\boldsymbol{\xi}(t))^\top\right]\left[\delta_{ij} + \tanh(\tau\boldsymbol{\xi}(t))\tanh(\tau\boldsymbol{\xi}(t))^\top\right] \\
&= \begin{bmatrix} 1-\rho_1\rho_1 & -\rho_1\rho_2 & \cdots \\ -\rho_2\rho_1 & 1-\rho_2\rho_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} 1+\rho_1\rho_1 & \rho_1\rho_2 & \cdots \\ \rho_2\rho_1 & 1+\rho_2\rho_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \\
&= \begin{bmatrix} 1-\sum O(\rho_a\rho_b\rho_c\rho_d) & \rho_1\rho_2 - \rho_1\rho_2 - \sum O(\rho_a\rho_b\rho_c\rho_d) & \cdots \\ -\rho_2\rho_1 + \rho_2\rho_1 - \sum O(\rho_a\rho_b\rho_c\rho_d) & 1-\sum O(\rho_a\rho_b\rho_c\rho_d) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \approx \boldsymbol{I}.
\end{aligned}
$$

Note that the Euclidean metric δ_{ij} is equal to the identity matrix $\boldsymbol{I}$.

C.4 Network Manifold Evolution on Classification Datasets

For CIFAR datasets (Krizhevsky et al., 2009), we set total training epochs as 200 where the learning strategy is lowered by 10 times at epoch 80, 150, and 190, with the initial 0.1. The learning strategy is a weight decay of 0.0001, a batch size of 128, SGD optimization.

Here, we apply ResNet-20 and ResNet-56 models (He et al., 2016) to observe the evolution of network manifold, i.e., the convergence of metrics depended on time. As far as we define the metric $\bar{g}(t)$, we can use the length $|ds^2| = \sqrt{\sum_{i,j}\bar{g}_{ij}(t)d\xi_i d\xi_j}$ to intuitively reflect the curvature of the manifold. More specifically, we define a ball whose radius is equal to $|ds^2|$:

$$
B_t(x,r) := \left\{ x \in \mathcal{M}^n : r = \sqrt{\sum_{i,j}\bar{g}_{ij}(t)d\xi_i d\xi_j} \right\}.
$$

By observing the change of the ball in Figure 2, we can know the change of the metric. Actually, the change in metric is not so obvious. In this case, we have zoomed in the vertical coordinate in order to make the changes in the metric look more obvious.