CHAPTER 7: RICCI FLOW

DANNY CALEGARI

Abstract. These are notes on Ricci Flow on 3-Manifolds after Hamilton and Perelman, which are being transformed into Chapter 7 of a book on 3-Manifolds. These notes are based on a graduate course taught at the University of Chicago in Fall 2019.

Contents

1. The Hamilton–Perelman program

In this section we give a very informal overview of the Hamilton–Perelman program proving the Poincaré Conjecture and the Geometrization Conjecture for 3-manifolds.

1.1. **What is Ricci flow?** There's lots of different ways to answer this question, depending on the point you want to emphasize. There's no getting around the precision and economy of a formula, but for now let's see how far we can get with mostly words.

First of all, what is curvature? To be differentiable is to have a good linear approximation at each point: the derivative. To be smooth is for successive derivatives to be themselves differentiable; for example, the deviation of a smooth function from its derivative has a good quadratic approximation at each point: the Hessian. A Riemannian manifold is a space which is Euclidean (i.e. flat) to first order. Riemannian manifolds are smooth, so there is a well-defined second order deviation from flatness, and that's Curvature.

In Euclidean space of dimension n a ball of radius r has volume r^n times a (dimension dependent) constant. The *scalar curvature* R (a function) measures the leading order deviation of this quantity in a Riemannian manifold. Explicitly, if we pick a point p in a manifold M, we denote the ball of radius r about p in M by $B_M(p, r)$. We want to

Date: November 8, 2023.

compare the geometry of this ball to that of $B_{\mathbb{E}^n}(0,r)$, the ball of radius r about the origin in Euclidean space of dimension n. Then

$$\frac{\mathrm{vol}(B_M(p,r))}{\mathrm{vol}(B_{\mathbb{E}^n}(0,r))} = 1 - R\,\frac{r^2}{6(n+2)} + O(r^3)$$

at least if r is small. In words: when the scalar curvature is positive (resp. negative), volume of metric balls grows slower (resp. faster) than in Euclidean space.

The *Ricci curvature* Ric measures the deviation of the volume *in a particular direction*. At a point p, we can choose a unit vector v and look at the volume growth of a tightly focussed cone starting at p in the direction v. When the Ricci curvature is positive (resp. negative) in the direction v, volume growth 'in the direction v' is slower (resp. faster) than in Euclidean space. The deviation is second order, so the Ricci curvature is a (symmetric) quadratic form; in other words, it has the same units as the Riemannian metric.

Ricci flow is a differential equation for the evolution of a family of metrics on a smooth manifold; it says that the time derivative of the Riemannian metric is -2 times the Ricci curvature; i.e. distances contract in directions where the volume grows slower than Euclidean space, and distances expand in directions where the volume grows faster than Euclidean space. Where does this formula come from? It turns out that in harmonic local coordinates (i.e. coordinates which are harmonic functions for the metric), the Ricci curvature is $-1/2$ times the Laplacian of the metric, up to lower order terms. Thus the Ricci flow might be thought of as a natural geometric flow modeled on the heat flow for the metric. Under such a 'heat flow', one imagines the metric will average out and become homogeneous. But just from the definition it's not at all obvious that Ricci flow is even defined for short time (it is) or that it does not become singular in finite time (it does). A proper analysis of its properties, including a classification of finite time singularities, surgery, and long-time behavior, is far beyond the scope of this survey. Our aim in this chapter is to give an introduction to the subject, and to explain enough of the recent developments to sketch how Ricci flow can be used to prove the Poincaré Conjecture and (with more work) the Geometrization Conjecture.

1.2. **Ricci flow.** The Ricci curvature is a symmetric 2-tensor, i.e. a section of the symmetric square of the cotangent bundle. In other words, it's a tensor of the same kind as the Riemannian metric tensor g.

Ricci flow, introduced by Richard Hamilton in his 1982 paper [12], is a differential equation for a 1-parameter family of Riemannian metrics g_t on a manifold M, specifically

$$\partial_t g = -2\,\mathrm{Ric}$$

Ricci flow will typically grow or shrink the volume; when the scalar curvature is positive the manifold will shrink, and when it is negative the volume will expand.

Ricci flow enjoys two fundamental symmetries: *diffeomorphism invariance* and *parabolic rescaling*.

First, for any diffeomorphism $\varphi : M \to N$ we have $\mathrm{Ric}(\varphi^* g) = \varphi^*\mathrm{Ric}(g)$. In particular, Ricci flow commutes with the group of self-diffeomorphisms of M. As a corollary Ricci flow preserves any symmetries of the initial metric.

Second, scaling g by a positive number λ stretches distances by $\sqrt{\lambda}$ and sectional curvatures by λ^{-1}. For the Ricci tensor these two factors cancel, and Ric is unchanged as a tensor (although its norm, which depends on the metric, is scaled by λ^{-1}). Thus the Ricci flow of λg is obtained by rescaling the Ricci flow of g, but proceeds with time stretched by the same factor λ. In other words, solutions to Ricci flow are preserved by *parabolic rescaling* of space and time $ds \to \sqrt{\lambda}\, ds, dt \to \lambda\, dt$.

1.2.1. *Short time existence and uniqueness.* The Ricci curvature depends linearly on the second order derivatives of the metric, and nonlinearly on lower order terms. The equation of Ricci flow is weakly parabolic — the symbol of $-2d\text{Ric}$, thought of as a quadratic form on the tangent space to the space of Riemannian metrics, is non-negative but not definite. This degeneracy is a result of diffeomorphism invariance. Nevertheless, Hamilton [12] proved short-term existence and uniqueness of Ricci flow on a compact manifold. A shorter proof due to DeTurck [10] explicitly breaks the degeneracy by adding a term that comes from comparison to a fixed background metric. The deformed flow is parabolic, and turns out to be equivalent to Ricci flow up to (time-dependent) diffeomorphism.

1.2.2. *Fixed points.* The simplest solutions to Ricci flow are when $\text{Ric} = 0$ so that g is constant with t. A manifold with this property is said to be *Ricci flat*. Any real $2n$-manifold with $\text{SU}(n)$ holonomy (a *Calabi-Yau* manifold) is Ricci flat. In 3 dimensions or less any Ricci flat manifold is a Euclidean space form, but for $n = 4$ the $K3$ surfaces are interesting examples.

A manifold with $\text{Ric} = \lambda g$ for some constant λ is called *Einstein*. Multiplying the metric by a constant leaves Ric unchanged, so if g_0 is Einstein, the family $g_t = (1 - 2\lambda t)g_0$ is a solution to Ricci flow. In 3 dimensions or less an Einstein manifold has constant curvature $\lambda/(n-1)$. If $\lambda \le 0$ the metric g_t is defined for all $t \ge 0$ but for $\lambda > 0$ it becomes singular, and the manifold vanishes to a point, at $t = 1/2\lambda$. The key example in three dimensions is the shrinking round sphere S^3. A 3-sphere of radius 1 has constant Ricci curvature equal to 2 and shrinks homothetically to a point at time $t = 1/4$.

If M is a product $M = A \times B$ with product metric $g_M := g_A \oplus g_B$ then the product of a geodesic in A with a geodesic in B is a totally flat 2-plane in M, so the Ricci curvature of M is $\text{Ric}_M = \text{Ric}_A \oplus \text{Ric}_B$. It follows that under Ricci flow, M evolves by the product of Ricci flows on the factors. The key example in three dimensions is the shrinking round cylinder $S^2 \times \mathbb{R}$. This has constant Ricci curvature equal to 1 in the S^2 direction and 0 in the $\mathbb{R}$ direction. Thus the $\mathbb{R}$ factor is unchanged, and the spheres shrink homothetically to points at time $t = 1$.

1.2.3. *Solitons.* A vector field X on M generates a flow ψ_t, and a flow of the form $\partial_t g = \mathcal{L}_X g$ (where $\mathcal{L}_X$ denotes Lie derivative) defines a family of metrics g_t which are obtained from g_0 by pullback $g_t = \psi_t^* g_0$ and are therefore all isometric.

Thus it's natural to consider generalizations of the Einstein condition, namely metrics satisfying

$$\text{Ric} = \lambda g - \frac{1}{2}\mathcal{L}_X g$$

For such an initial metric, Ricci flow scales the metric infinitesimally at the constant speed -2λ, while simultaneously flowing by X. Thus the metrics g_t are all self-similar, in the sense that each is related to the initial metric by rescaling plus a diffeomorphism. Such a metric is called a *soliton*. The soliton is said to be *expanding, steady* or *shrinking* according to whether λ is negative, zero or positive. A shrinking soliton becomes singular at $t = 1/2\lambda$. Examples of shrinking solitons include the round sphere S^3 and the round cylinder $S^2 \times \mathbb{R}$.

A *gradient soliton* is one for which $X = \text{grad} f$ for some function f. Note that there is an equation $\mathcal{L}_{\text{grad} f} g = 2\,\text{Hess}(f)$ so a metric determines a gradient soliton if there is f such that $\text{Ric} + \text{Hess}(f) = \lambda g$.

Hamilton's *cigar soliton* is given by the metric $g = (dx^2 + dy^2)/(1 + x^2 + y^2)$ on $\mathbb{R}^2$ which evolves under Ricci flow by pullback under the radial vector field $X = -2(x\partial_x + y\partial_y)$. One could think of this as an infinite nearly cylindrical cigar whose tip is rounded; under Ricci flow the tip 'burns away' leaving a new cigar isometric to the first. This is very similar to the *grim reaper* soliton for mean curvature flow; see § 2.1.3.

Bryant's *bowl soliton* exists on $\mathbb{R}^n$ for any $n \geq 3$, and is given in polar coordinates by a radially symmetric metric $g = dr^2 + a(r)^2 g_{S^{n-1}}$ for suitable $a(r)$ asymptotic to $\sqrt{r}$ as r gets large. These are much like the bowl solitons for mean curvature flow.

1.2.4. *Berger spheres.* The three-sphere is a Lie group if we identify it with the unit quaternions. The round metric is left-invariant. Let e_1, e_2, e_3 be a left-invariant orthogonal frame for the tangent bundle, and let $\omega^1, \omega^2, \omega^3$ be the dual frame for the cotangent bundle. A *Berger sphere* is a Riemannian manifold diffeomorphic to S^3, with metric of the form $g = A\omega^1 \otimes \omega^1 + \omega^2 \otimes \omega^2 + \omega^3 \otimes \omega^3$. Geometrically, the vector field e_1 has flowlines the (totally geodesic) circles of a Hopf fibration; a Berger metric is obtained from the round metric by scaling these circles by a factor $\sqrt{A}$. For such a metric, the eigenvectors of the curvature operator Rm are the coordinate planes $e_i \wedge e_j$ and one can compute their sectional curvatures as

$$K(e_1 \wedge e_2) = A, \quad K(e_1 \wedge e_3) = A, \quad K(e_2 \wedge e_3) = 4 - 3A$$

In other words, for the family of metrics with $A \to 0$ the sectional curvatures stay *bounded* while the volume goes to zero, and the manifold 'collapses' to a round 2-sphere of radius 1/2.

The metric $g = A\omega^1 \otimes \omega^1 + B\omega^2 \otimes \omega^2 + B\omega^3 \otimes \omega^3$ is homothetic to a Berger sphere, so it has sectional curvatures

$$K(e_1 \wedge e_2) = A/B^2, \quad K(e_1 \wedge e_3) = A/B^2, \quad K(e_2 \wedge e_3) = 4/B - 3A/B^2$$

and Ricci curvature

$$\text{Ric} = 2A^2/B^2\, \omega^1 \otimes \omega^1 + (4 - 2A/B)\omega^2 \otimes \omega^2 + (4 - 2A/B)\omega^3 \otimes \omega^3$$

In particular, Ric is *diagonal* with respect to our chosen frame. Actually, this follows from the fact that the metric (and hence Ric) has an SO(2) family of symmetries fixing every point with axis in the e_1 direction.

Under Ricci flow we evidently get a family of homothetically scaled Berger spheres, parameterized by (time-dependent) functions A and B that satisfy

$$A' = -4A^2/B^2, \quad B' = -8 + 4A/B$$

This system of ODEs becomes singular in finite time, but as it does so the ratio A/B evolves by $(A/B)' = 8A(B - A)/B^3$ which is positive if $B > A$ and negative if $B < A$ so that asymptotically $A/B \to 1$ and the collapsing spheres converge after rescaling to the round S^3. See Figure 1.

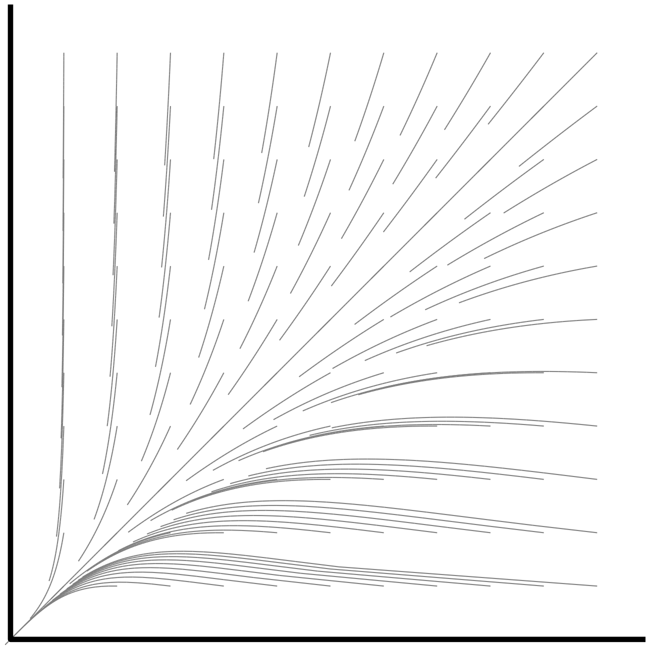

FIGURE 1. Evolution of A, B parameters for Berger spheres under Ricci flow. All flowlines become asymptotic to the diagonal near the origin.

1.3. **Geometrization.** Hamilton's program, as successfully completed by Perelman, has spectacular applications to 3-manifold topology of which the most famous is a proof of the Geometrization Conjecture and (as a special case) the Poincaré Conjecture. The details of this program lie far beyond the scope of this chapter; it's challenging even to give an overview. The following is just a cartoon.

1.3.1. *Short time behavior.*

(1) Definition of Ricci flow $\partial_t g = -2\text{Ric}$
(2) Ricci flow exists and is unique for short time
(3) Uniform curvature bounds give control over higher derivatives of curvature
(4) Monotonicity: certain geometric inequalities (pertaining to curvature or volume or both) persist or improve with time; most importantly:
   (a) non-negative (positive) sectional curvature is preserved
   (b) in dimension 3 non-negative (positive) Ricci curvature is preserved
   (c) pinching: there is a function $\phi(s)$ that goes to 0 as $s \to \infty$ so that
   $$\text{Rm} \ge -\phi(R)R + C$$
   where the estimate holds pointwise/timewise
   (d) κ-noncollapsing in finite time: for Ricci flow on a compact manifold with normalized initial conditions, if we rescale the metric at some finite time so that $R = 1$ at some point, there is a lower bound κ on the volume — and hence injectivity radius — of the ball of (rescaled) radius 1 around this point

It's important to quantify the implicit estimates. First, by rescaling the original metric we can assume we have *normalized initial conditions*: i.e. $|\text{Rm}| \le 1$, and every ball of radius 1 has volume at least half of the volume of a unit ball in Euclidean space. Second, the κ in the definition of κ-noncollapsing depends on the time t at which we are doing the rescaling.

1.3.2. *Structure of finite time singularities.*

(1) When a singularity develops in time, the effect of doing a parabolic rescaling near a point where R blows up is to obtain a new flow which is ϵ-close to a *κ-solution*
(2) A κ-solution satisfies the following properties:
   (a) it's ancient (flow is defined on $(-\infty, t]$)
   (b) the curvature is non-negative $\mathrm{Rm} \geq 0$
   (c) the curvature norms $|\mathrm{Rm}|$ are bounded on each time slice
   (d) the scalar curvature R is positive everywhere
   (e) it's κ-noncollapsed
   (f) normalized volume controls normalized curvature and vice versa
(3) The set of pointed κ-solutions is compact in the sense of Cheeger-Gromov-Hamilton convergence
(4) A compact κ-solution is diffeomorphic to a spherical space form
(5) In every non-compact κ-solution defined at time t, there's a scale D and a point x so that outside the ball of radius $DR(x,t)^{-1/2}$ about x every point is the center of an *ϵ-neck*; this means that after rescaling to have $R = 1$, the manifold is ϵ-close to a round product $S^2 \times \mathbb{R}$ on a ball of radius $1/\epsilon$
(6) Unless M is a round cylindrical flow with $S^2 \times \mathbb{R}$ geometry, the ball promised by the previous bullet is either a 3-ball or a punctured $\mathbb{RP}^3$

Again, the estimates must be quantified: D depends on κ and ϵ, and in order to apply these structure theorems to finite time singularity, we are only assured some κ which in turn depends on time.

1.3.3. *Surgery.*

(1) Just before a singularity, the large curvature part of the manifold (where the curvature is at least some R_0 depending on ϵ and t) consists of entire components which are ϵ-close to shrinking space forms or manifolds with $S^2 \times \mathbb{R}$ geometry, or they have a canonical geometric fibration by almost round 2-spheres, with high curvature ends capped off by 3-balls or punctured $\mathbb{RP}^3$s
(2) Near each frontier region of the large curvature part of the manifold we perform *surgery*:
   (a) each closed component evolves under Ricci flow and shrinks to an asymptotically round or $S^2 \times \mathbb{R}$ geometry point in finite time; in particular, these summands admit a geometric structure
   (b) at a scale where the curvature is at least as big as a certain threshold $R_0/\sqrt{\delta}$ (but where $\delta(t) \to 0$ as $t \to \infty$), we cut off the (nearly product) neck, warp the metric slightly to round the end, and cap off with a round B^3
   (c) at the topological level, this has the effect of undoing finitely many connect sums or self-connect sums
   (d) at the geometric level, this can be done in such a way that if we restart the flow, the pinching inequality still holds, and all relevant geometric quantities

(κ, D, R_0, δ etc.) can still be controlled and deteriorate with time in an *a priori* specified way

(3) the estimates on the geometric quantities can be arranged so that when one performs Ricci flow with surgery, the surgery times do not accumulate, and the flow can be continued until $t = \infty$

1.3.4. *Finite time extinction.*

(1) If Ricci flow with surgery becomes *extinct* (i.e. the manifold is empty after finite time) then it is obtained from spherical space forms and manifolds with $S^2 \times \mathbb{R}$ geometry by *finitely many* connect sums and self-connect sums; in particular, it satisfies the Geometrization Conjecture
(2) Ricci flow with surgery becomes extinct under the following circumstances:
    (a) the scalar curvature R is positive everywhere
    (b) the scalar curvature R is non-negative everywhere and M is not Euclidean
    (c) if every prime summand of M has non-trivial π_2 or π_3
(3) In more detail, under Ricci flow with surgery the area of a minimal S^2 representing a nontrivial element of π_2, or a minimax S^2 associated to a nontrivial element of π_3, shrinks at a definite rate; in particular, a component containing such a sphere becomes extinct in finite time.

A homotopy 3-sphere has $\pi_3(M) = \mathbb{Z}$; thus under Ricci flow with surgery it becomes extinct in finite time, and consequently it is homeomorphic to S^3. This proves the Poincaré Conjecture.

1.3.5. *Long time behavior.*

(1) when t gets big, there is a 'thick-thin' decomposition: a point is in the thick part if the ball around it of radius $\sqrt{t}$ has (after rescaling to radius 1) controlled curvature and injectivity radius
(2) at points in the thin part, there is some scale $r \leq \sqrt{t}$ so that rescaled balls have $\text{Rm} \geq -1$ and *small* normalized volume
(3) the evolution equation for scalar curvature implies that the ratio $\text{vol}(t)/(t+1/4)^{3/2}$ is non-increasing; if the limit is zero, the entire manifold is thin and the theory of collapsing with one-sided (lower) curvature bounds implies that M has the structure of a graph manifold
(4) otherwise the thick part stays non-empty, and the scalar curvature becomes very close to its (scaled) infimum throughout the thick part; thus the rescaled balls converge to a hyperbolic metric
(5) because balls in the thick part of radius $\sqrt{t}$ have volume comparable to that of the entire manifold, we can cover the thick part with boundedly many standard balls in which the metric is closer and closer to hyperbolic; in particular, the thick part admits a hyperbolic structure, and the manifold satisfies the Geometrization Conjecture

2. Mean curvature flow: a comparison

Riemannian metrics on 3-manifolds are hard to visualize. Fortunately, there is another domain — mean curvature flow of surfaces in $\mathbb{R}^3$ — that displays many of the same qualitative features as Ricci flow on 3-manifolds, but where pictures are much easier to draw.

Our survey of mean curvature flow is extremely brief, and meant to highlight a few key features (monotonicity of curvature, singularity formation and local models, entropy functionals) which closely parallel Ricci flow.

2.1. Definitions and Basic Examples.

Let S be a hypersurface in $\mathbb{R}^n$. The *mean curvature* is the trace of the second fundamental form. Recall: if e_i are linearly independent vector fields on S near a point p then $\mathrm{II}(e_i, e_j) := e_i(e_j(x))(p)^\perp$. Then $H(p) := \sum_i \mathrm{II}(e_i, e_i)$ where e_i runs over an orthonormal basis for T_pS.

If we fix coordinates x_j on an abstract surface S and an immersion $F : S \to \mathbb{R}^n$ then on S the metric g and second fundamental form h can be expressed in coordinates as

$$g_{ij} := \langle \partial_i F, \partial_j F \rangle, \quad h_{ij} := -\langle \nu, \partial_i \partial_j F \rangle$$

where $\nu(x)$ is the (outer) unit normal to the surface at $F(x)$. With this notation one also defines the *scalar* mean curvature h to be the trace of h_{ij}; i.e. $h = g^{ij} h_{ij}$. Thus $H = -h\nu$ (the sign is chosen so that for a mean convex surface $h \ge 0$).

A family of hypersurfaces $F_t : S \to \mathbb{R}^n$ is said to evolve by *mean curvature flow* (abbreviated MCF) if it satisfies

$$\partial_t F = H = -h\nu$$

or more generally

$$(\partial_t F)^\perp = H$$

Flows satisfying the second equation differ from the first only by reparameterization of the surface.

Stationary solutions to mean curvature flow are minimal surfaces, and in fact one can think of mean curvature flow as gradient flow for the area functional on the space of smooth maps.

2.1.1. *Self-shrinkers.*

The simplest non-static examples of MCF are shrinking spheres and cylinders. In $\mathbb{R}^3$ examples are a family of 2-spheres of radius $\sqrt{-4t}$ and a family of cylinders of radius $\sqrt{-2t}$ for $t < 0$.

At $t = 0$ the family of shrinking spheres becomes singular, and collapses to a point, whereas the family of shrinking cylinders collapses to a straight line.

Both are examples of *self-shrinkers*:

Definition 2.1 (Self-Shrinker)**.** A family F_t, $t \in [-1, 0)$ evolving by MCF is a *self-shrinker* if $F_t = \sqrt{-t} F_{-1}$.

By slight abuse of notation we also call F_t a self-shrinker if it satisfies this equation after translation (in time and/or space) and/or parabolic rescaling $(x, t) \to (\lambda x, \lambda^2 t)$.

2.1.2. *Barriers and the maximum principle.* If R_t and S_t are two (complete) hypersurfaces evolving by mean curvature, then if R_0 and S_0 are disjoint, then R_t and S_t continue to be disjoint for all $t > 0$ where defined. To see this, suppose that R_0 is on the 'outside' of S_0, and suppose at some first time t the surfaces R_t and S_t become tangent at p. Since R_t is still on the outside of S_t, the mean curvature of S_t is bigger than the mean curvature of R_t in the direction pointing into the interior; but this means that S_t is moving into the interior at p faster than R_t is; running time backwards slightly this implies that $R_{t-\epsilon}$ already intersected $S_{t-\epsilon}$ for small ϵ, contrary to the definition of t as the first time the surfaces intersect.

It follows that *every closed hypersurface becomes singular under MCF in finite time.* Indeed, any S is in the interior of a round ball B of some finite radius r. Since ∂B shrinks to a point in time $r^2/4$, the surface S must become singular before this time.

2.1.3. *The grim reaper.* The *grim reaper* is a noncompact translating solution to MCF in the plane. At some initial time it's given by the graph $y = -\log\cos(x)$ for $x \in (-\pi/2, \pi/2)$. Under MCF it translates upwards at constant speed. Taking products with Euclidean space gives examples in any dimension. One can also find ($O(n-1)$-)rotationally symmetric solutions, which look roughly like paraboloids, called *bowl solitons*. Grim reapers are often used in barrier arguments to get a priori bounds.

2.1.4. *The Dumbell.* A *dumbell* is a surface obtained by taking two round spheres (the 'bells') and tubing them together by a narrow neck, and then rounding the corners at the ends of the tube. Typically one takes the two spheres to be the same radius (otherwise the dumbell is 'lopsided'). A round sphere concentrically placed inside the bells puts a lower bound on how long it takes for this part of the surface to shrink to nothing. Meanwhile, a sufficiently small Angenent doughnut around the tube puts an *upper* bound on the time to the first singularity. If the bells are big enough compared to the thickness of the tube, we can deduce that a singularity develops in finite time without the diameter going to zero; see Figure 2.

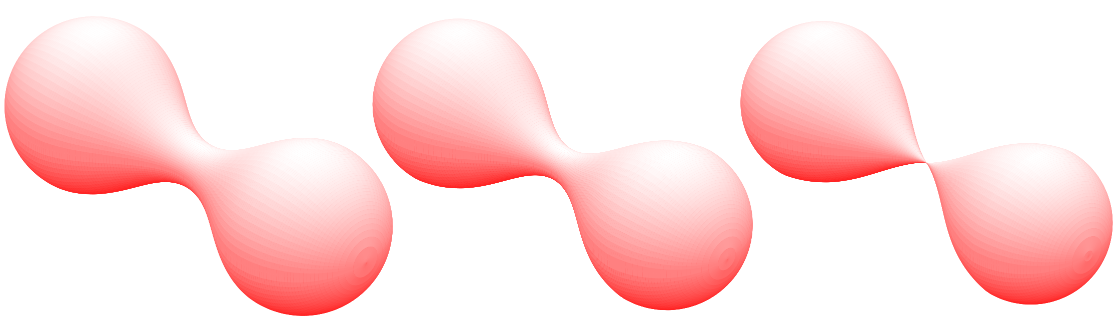

FIGURE 2. A shrinking dumbell pinches off a neck

This kind of singularity is called a *neckpinch*. Near the singular time, the neck converges after parabolic rescaling to the round shrinking cylinder.

2.1.5. *Lopsided Dumbell.* A *lopsided dumbell* is a dumbell with two bells of different radii. For such a dumbell one of the bells shrinks faster than the other, and for a judicious choice of bell and tube radius it's plausible that the bell could shrink to a point at exactly the same time that the neck becomes singular. At this singularity we are left precisely with the larger bell and a metric which is convex and nonsingular except at exactly one 'cone' point where the curvature has become infinite. It turns out one can then evolve this singular sphere by MCF; it instantly becomes smooth and convex and thereafter by Huisken's Theorem 2.2 shrinks to a round point in finite time. This kind of singularity is called a *degenerate neckpinch*.

2.2. **Self-shrinkers as minimal surfaces.** It turns out that a self-shrinker is just a minimal surface for a suitable metric on $\mathbb{R}^3$. Let's consider a generalized MCF of the form $\partial_t F = H + X$ for some (time-dependent) vector field X always tangent to F. The normalized self-shrinker condition is $F(t) = \sqrt{-t}F(-1)$ so that at time -1 we have $F' = -F/2 + X$ or equivalently $\langle H + F/2, \nu\rangle = 0$.

For any smooth function ϕ on $\mathbb{R}^3$ there's an associated functional S on surfaces defined by $S(F) := \int_F \phi \, d\text{area}$. This functional is nothing but the area of F in the conformally Euclidean metric $ds^2 = \phi(x)dx^2$. When is F a minimal surface for such a metric? Let's vary F by moving it infinitesimally in the normal direction by $f\nu$ where f is some smooth function. Then

$$\delta S = \int_F \delta\phi \, d\text{area} + \int_F \phi \, \delta d\text{area} = \int_F \langle f\nu, \nabla\phi\rangle + \phi\langle f\nu, -H\rangle d\text{area}$$

The radially symmetric function $\phi(x) = e^{-|x|^2/4}$ satisfies $\nabla\phi(x) = -(x/2)\phi$ so that δS vanishes identically for all f if and only if F is a self-shrinker. The functional S is due to Huisken.

2.2.1. *Angenent's shrinking doughnuts.* Using Huisken's S functional, Angenent [1] constructed a smooth embedded torus in $\mathbb{R}^3$ which is a self-shrinker. The torus is obtained as a surface of revolution (about the x-axis), with cross-section a circle $\gamma : S^1 \to x$–z plane. In order for the resulting torus to be a critical point for S, it's necessary and sufficient for γ to be a geodesic for the (incomplete) metric

$$ds^2 = z^2 e^{-\frac{x^2+z^2}{4}}(dx^2 + dz^2)$$

on the upper half plane $z > 0$.

For simplicity, one looks for a geodesic invariant under the symmetry $x \to -x$, in which case one can normalize the initial condition so that $\gamma(0) = (0, s)$ and $\gamma'(0) = (1, 0)$. Continue this initial condition until the first time γ runs into the z axis again (if it does), at $\gamma(t(s)) = (0, z(s))$ with $\gamma'(t(s)) = (\alpha(s), \beta(s))$.

If $\alpha(s) < 0$ and $\beta(s) = 0$ then reflection of the arc $\gamma([0, t])$ in the z axis gives the desired geodesic. The existence of such an s can be proved numerically; it turns out $s \sim 3.3151$.

2.3. **Convexity and Huisken's theorem.** The following key theorem was proved by Huisken in 1984:

Theorem 2.2 (Huisken [19] 1.1)**.** *Let S_0 be a uniformly convex hypersurface in $\mathbb{R}^n$ with $n \geq 3$. Then MCF has a smooth solution on a maximal time interval $[0, T)$ and the surfaces*

S_t converge to a single point as $t \to T$. Furthermore if the surfaces S_t are homothetically rescaled to have constant area they converge to a round sphere of that area in the C^∞ topology as $t \to T$.

One says informally that a convex surface shrinks to a 'round point' in finite time.

There is a strong analogy between this theorem, and the theorem of Hamilton (to be proved in the sequel) that a 3-manifold with positive Ricci curvature converges by (rescaled) Ricci flow to a spherical space-form, but actually Hamilton's result came earlier and was the direct inspiration for Huisken.

2.3.1. *Evolution of geometric quantities.* By direct calculation one obtains formulae for the evolution of key geometric quantities. Let's denote the metric on our surface by g_{ij}. The second fundamental form is either denoted by h_{ij} (if we want to emphasize coordinates) or A (if we just want to think of it as a tensor). This notation is standard, and frees up h to denote the trace of h, i.e. the scalar mean curvature.

Proposition 2.3. *Under MCF one has the formulae for the evolution of the metric g_{ij}:*

$$\partial_t g_{ij} = -2hh_{ij}$$

for the second fundamental form h_{ij}:

$$\partial_t h_{ij} = \Delta h_{ij} - 2hh_{il}g^{lm}h_{mj} + |A|^2 h_{ij}$$

for the norm of the second fundamental form $|A|^2$:

$$\partial_t |A|^2 = \Delta |A|^2 - 2|\nabla A|^2 + 2|A|^4$$

and for the scalar mean curvature h:

$$\partial_t h = \Delta h + |A|^2 h$$

Taking traces of the first equation, one sees that the area form $\mu := \sqrt{\det g_{ij}}$ evolves by $\partial_t \mu = -h^2 \mu$; i.e. total area is decreasing. Furthermore from the third equation and the maximum principle it follows that if the mean curvature is non-negative (resp. strictly positive) for any t_0, then it stays non-negative (resp. strictly positive) for all $t > t_0$.

By the Gauss equation, the Ricci curvature R_{ij} of a hypersurface satisfies

$$R_{ij} = hh_{ij} - g^{kl}h_{kj}h_{li}$$

Thus under MCF the metric evolves by $\partial_t g_{ij} = -2\text{Ric} + 2g^{kl}h_{kj}h_{li}$. This demonstrates a family resemblance between MCF and Ricci flow.

2.3.2. *Convexity and the tensor maximum principle.* A more subtle analysis of the evolution of the second fundamental form h_{ij} shows that not only mean convexity, but honest *convexity* is preserved by MCF. The term $-2hh_{il}g^{lm}h_{mj}$ in the formula for $\partial_t h_{ij}$ reflects the contribution from the change in the metric, so in a suitable evolving orthonormal frame e_a, e_b etc. this term goes away, and the evolution equation for h_{ij} simplifies to

$$\partial_t h_{ab} = \Delta h_{ab} + |h_{ab}|^2 h_{ab}$$

(see Hamilton [16] Thm. 2.3; note that a, b etc. do *not* denote coordinate indices).

Since h_{ab} is a tensor, the ordinary maximum principle does not apply. However, Hamilton proved a *tensor maximum principle*, to be discussed in detail in § 3.6, which does apply directly to equations of this sort. Roughly speaking, Hamilton's principle applies to tensors T which are sections of a vector bundle V evolving by a PDE of the form $\partial_t T = \Delta T + \Psi(T)$ where Ψ has order 0. Suppose we want to prove that solutions to the PDE stay in some closed subspace K of V satisfying suitable conditions on K (fiberwise convexity, invariance under parallel transport). For each fiber V_x consider the associated ODE $\partial_t T_x = \Psi(T_x)$. Hamilton's tensor maximum principle says that if every solution of a fiberwise ODE which starts in K_x must stay in K_x, then any solution of the PDE which starts in K must stay in K.

The conditions of the theorem apply to the evolution equation for the second fundamental form, where we can take K to be the subspace where the eigenvalues are non-negative. Thus convexity is preserved by MCF. A strong version of the principle implies that strict convexity is preserved, and actually a spacewise uniform lower bound on the eigenvalues can only increase with time.

One consequence is that when the surface becomes singular in finite time (as it must) it can only collapse to a single point.

2.3.3. *Curvature pinching and convergence to a round point.* The eigenvalues of the second fundamental form are the principal curvatures λ and μ. Let's define the function $f := h^{-2}(|A|^2 - h^2/2) = h^{-2}(\lambda-\mu)^2/2$, a scale-invariant measure of how close these eigenvalues are to each other. We've already seen that strict mean convexity $h > 0$ is preserved under MCF, so f is nonsingular while MCF is. Evidently, $f \geq 0$ and is equal to zero at umbilical points — those where the principal curvatures are equal.

From the evolution equations for h and for h_{ij} one derives the evolution equation (c.f. [19], Lem. 5.2)

$$\partial_t f = \Delta f + \frac{2}{h}\langle \nabla f, \nabla h\rangle - \frac{2}{h^4}|h\nabla_i h_{lk} - h_{lk}\nabla_i h|^2$$

At a local spatial pointwise maximum, we must have $\nabla f = 0$ and $\Delta f \leq 0$. It follows that the maximum of f is monotonically nonincreasing in time, and with more work Huisken is able to show it must actually decrease and go to zero everywhere as the surface shrinks to a point. A similar evolution equation gives an *a priori* bound on the norm of ∇h_{ij} after rescaling, so as we approach a singularity the rescaled surfaces become more and more umbilical everywhere. But a totally umbilical surface in $\mathbb{R}^3$ is a round sphere or plane, by a classical theorem of Meusnier. In words, a mean convex surface shrinks by MCF to a round point in finite time.

The fact that $f \to 0$ monotonically is true in all dimensions; however, Meusnier's theorem only holds for hypersurfaces in dimension at least 3. The analog of Huisken's theorem for *curves* in the plane is true (actually, the initial hypothesis of convexity is superfluous) but the proof is completely different.

2.4. **Singularities of MCF.** By focussing at the first place a singularity develops, and rescaling so that $|A|^2 = 1$ we obtain a noncompact surface. Let's translate the family F_t in space and time so that the singularity is at the origin $(0,0)$ and then take a sequence of parabolic dilations $(x,t) \to (\lambda x, \lambda^2 t)$ with $\lambda \to \infty$ to obtain MCFs F_t^λ. It turns out that some subsequence of the F_t^λ necessarily converges weakly to a limiting 'tangent flow' F_t^∞. Just as the usual tangent cone construction produces an object with a dilational symmetry, it turns out that the tangent flow is a self-shrinker; i.e. that $F_t^\infty = \sqrt{-t}F_{-1}^\infty$.

This was proved for so-called type I 'rapidly forming' singularities by Huisken [20] (we'll see his argument in a momemnt) and in full generality by Ilmanen [22].

2.4.1. *Entropy functionals.* To prove that some limit F_t^∞ exists one needs two-sided control over the norm of the second fundamental form A and its spatial derivatives $\nabla^m A$ for F_t^λ for each fixed compact interval $a \le t \le b < 0$ independent of λ. Because the surfaces we are considering are embedded in $\mathbb{R}^3$ lower bounds on injectivity radius come for free from curvature bounds.

The time derivative of $|\nabla^m A|^2$ is equal to $\Delta|\nabla^m A|^2 + 2|\nabla^{m+1}A|^2$ plus a polynomial in the various $\nabla^* A$ of order at most m. This lets one use a bootstrapping argument to control the norm of $\nabla^m A$ in the rescaled flows in terms of the norm of A. So we are reduced to getting normalized control on $|A|^2$ as we approach the singularity.

Huisken imposes this control by fiat. First of all by the maximum principle, the evolution equation

$$\partial_t |A|^2 = \Delta|A|^2 - 2|\nabla A|^2 + 2|A|^4$$

implies that the spacewise maximum $\max_{F_t} |A|^2$ grows at least like $\max_{F_t} |A|^2 \ge 1/2(T-t)$. A singularity (i.e. a first time T when the surface becomes singular) is said to be *rapidly forming* (one also says *type I*) if this estimate is sharp up to a constant; i.e.

$$\max_{F_t} |A|^2 \le \frac{C}{2(T-t)}$$

Examples include convex surfaces and cylinders, and rotationally symmetric shrinking necks. For such a surface one has a priori uniform geometric control on the rescaled flows F_t^λ, and some subsequence converges to a limit flow F_t^∞ called the *tangent flow.*

To prove that the tangent flow is a self-shrinker there is a further ingredient. Huisken introduced the first examples of what have become known as *entropy functionals* in [20] § 3. Recall in § 2.2 we defined the function S for surfaces F in $\mathbb{R}^3$ by

$$S(F) := \frac{1}{4\pi}\int_F e^{-x^2/4}\text{darea}$$

(up to a multiple of 4π) and observed that F is a critical point for S if and only if it's a self-shrinker.

Let's suppose $F(t)$ becomes singular at $(0,0)$, and for the sake of clarity, let's make the dimension dependent quantities explicit by working with hypersurfaces in arbitrary $\mathbb{R}^{n+1}$. For a family $F(t)$ evolving by MCF, let's consider the time-dependent functional

$$S_t(F) := \frac{1}{(4\pi(-t))^{n/2}}\int_{F_t} e^{-x^2/(-4t)}\text{dvol}_t$$

Write $\phi(x,t) := e^{-x^2/(-4t)}/(4\pi(-t))^{n/2}$ and $\tau := -t$. With this notation,

Proposition 2.4 (Huisken, [20] Thm. 3.1)**.** *The time derivative of $S_t(F(t))$ satisfies*

$$\frac{d}{dt}S_t(F_t) = -\int_{F(t)} \phi\Big|H + \frac{1}{2\tau}F^\perp\Big|^2 \text{dvol}_t$$

In particular, this quantity is non-increasing with t, and is stationary if and only if F is a self-shrinker.

Proof. Because F is evolving by mean curvature, the time derivative of the volume form at time t is $-H^2\text{dvol}_t$ so we have

$$\frac{d}{dt}S_t(F(t)) = \int_{F(t)} \langle \nabla\phi, H\rangle + \phi' - \phi H^2 \,\text{dvol}_t$$

Now, $\nabla\phi = -(F/2)\phi$ and $\phi' = (n/2\tau - |F|^2/4\tau^2)\phi$. Therefore

$$\begin{aligned}\frac{d}{dt}S_t(F_t) &= \int_{F(t)} -\phi\left(H^2 - \frac{n}{2\tau} + \frac{1}{2\tau}\langle F, H\rangle + \frac{|F|^2}{4\tau^2}\right)\text{dvol}_t \\ &= \int_{F(t)} -\phi\left|H + \frac{1}{2\tau}F\right|^2 \text{dvol}_t + \int_{F(t)} \frac{\phi}{2\tau}\langle F, H\rangle \text{dvol}_t + \int_{F(t)} \frac{n\phi}{2\tau}\text{dvol}_t\end{aligned}$$

Now, let's consider the variation of the volume of $F(t)$ in the direction $Y := \phi F/2\tau$. On the one hand, this is $\int_{F(t)} -\langle Y, H\rangle \text{dvol}_t$. On the other hand, if we restrict attention to the tangent plane to F at some point, then in directions tangent to the level sets of ϕ the derivative of distance is $\phi/2\tau$ whereas in the direction of $\nabla\phi$ the derivative of distance is $\phi/2\tau + \langle \nabla\phi/2\tau, F^\top\rangle$. It follows that we obtain an identity

$$\int_{F(t)} \frac{\phi}{2\tau}\langle F, H\rangle \text{dvol}_t = \int_{F(t)} \phi\left(-\frac{n}{2\tau} + \frac{|F^\top|^2}{4\tau^2}\right)\text{dvol}_t$$

Making this substitution proves the identity and the proposition. □

Huisken's theorem follows from this. Monotonicity of $S_t(F_t)$ implies that for each fixed t the function $\lambda \to S_t(F_t^\lambda)$ is strictly decreasing as a function of λ unless F_t is a self-shrinker. The infimum is achieved for the limit flow F_t^∞; thus the limit flow is a self-shrinker.

Ilmanen removes the type I hypothesis and proves the existence of a self-shrinking tangent flow in full generality; see [22] Lemma 8 for a precise statement. His arguments use geometric measure theory rather than the PDE methods of Huisken, and his techniques apply to the more general world of *Brakke flows*, where in place of hypersurfaces one works with a family of Radon measures on $\mathbb{R}^n$ satisfying MCF only in a weak distributional sense.

2.4.2. *Classification of generic singularities.* Near a singularity the parabolic blow-ups of a MCF family converge to a self-shrinker. The example of Angenent's doughnut shows that the geometry, and even the topology of a self-shrinker can be rather complicated. However this leaves open the possibility that for *generic* surfaces the only self-shrinkers that arise as limits of singularities are spheres and cylinders.

This is in fact accomplished by Colding-Minicozzi [9]. To explain the argument, let's consider a family of functionals of the form

$$S_{x_0,t_0}(F) := \int_F (4\pi t_0)^{-n/2} e^{-|x-x_0|^2/4t_0}\text{dvol}$$

This is Huisken's time-dependent entropy functional, centered at an arbitrary point x_0. Define a new functional λ of a hypersurface $\lambda(F)$ to be the supremum of $S_{x_0,t_0}(F)$ over all x_0, t_0. This functional is non-negative, invariant under similarities of Euclidean space, non-increasing under MCF, and the critical points are self-shrinkers. It's customary to refer to this functional simply as *entropy*.

Now Colding–Minicozzi's argument has two main ingredients. The first is an analysis of *stability* for self-shrinkers: they show that the only 'stable' self-shrinkers in dimension 3 are spheres, planes and cylinders. For every other self-shrinker one can find a small graphical perturbation whose entropy is strictly smaller. The second is a *compactness* theorem: for any fixed upper bound on area and genus, the space of self-shrinkers is compact. Because of compactness, when you perturb an unstable self-shrinker the entropy goes down by a definite amount.

So: start with an arbitrary evolving surface, and zoom in right before it becomes singular. If the tangent flow is stable, there's nothing to show. Otherwise it can be perturbed a very small amount so that the entropy is reduced. Repeat the process for the perturbed surface: i.e. zoom in near an evolving singularity, perturb if necessary, and so on. Since entropy is always positive, and the entropy of the original surface was finite, we only need to perform finitely many perturbations before the tangent flow becomes stable.

2.4.3. *MCF with surgery.* Suppose F develops a singularity with tangent flow a shrinking round cylinder. One can zoom in to just before the singularity develops and perform *surgery* — cut off the neck where it starts to get large, and replace the interpolating cylinder by a pair of round hemispheres to cap off the exposed ends. This operation is called a *surgery*. Topologically, it has the effect of undoing a connect sum or self-connect sum of F. If this operation is performed judiciously, we can restart MCF on the surgered surface until the next singularity develops. Near a singularity with tangent flow a shrinking round sphere there is an even simpler operation: we can zoom in to just before the singularity develops and simply throw the (nearly) round surface away. For generic initial F this MCF with surgery makes sense for all time: there are finitely many times where we undo a connect sum, and finitely many times when some component shrinks to a point and disappears. After every component has disappeared the 'flow' proceeds statically on the empty surface.

Huisken and Sinestrari [21] developed this procedure rigorously in any dimension: near a singularity with tangent flow is a shrinking round $S^{n-1} \times \mathbb{R}$, cut off the neck and replace with round hemispheres; near a singularity with tangent flow a shrinking round S^n, throw the hypersurface away. If these are the only singularities that develop, one deduces (by reversing the topological operations) that the original hypersurface F was diffeomorphic either to S^n or to a finite connected sum of $S^{n-1} \times S^1$s.

Say that a hypersurface is *two-convex* if the sum $\lambda_1 + \lambda_2$ of the two smallest eigenvalues of the second fundamental form is non-negative everywhere. An application of the tensor maximum principle shows that two-convexity is preserved by MCF. The main result of [21] is that if $n \geq 3$ (so F is a hypersurface in $\mathbb{R}^{\geq 4}$) and F is two-convex, the only singularities that develop have tangent flows a round sphere or cylinder, and furthermore one can perform surgery near these singularities in such a way as to preserve the condition of two-convexity. As a purely topological conclusion they deduce that any two-convex hypersurface in $\mathbb{R}^{\geq 4}$ is diffeomorphic to S^n or a finite connected sum of $S^{n-1} \times S^1$s.

3. Curvature evolution and pinching

In this section we compute formulae for the evolution of various geometric quantities under Ricci flow. We prove short time existence and uniqueness of the flow after Hamilton and DeTurck, and obtain various monotonicity and pinching estimates for curvature as applications of the maximum principle. These estimates are crucial for obtaining a priori control on the structure of the singularities that form in finite time.

3.1. **Formulae after all.** Let's turn now to formulae. In Riemannian geometry there's typically a trade off between economy of notation and ease of calculation, and in order to facilitate the latter it's crucial to be able to work in local coordinates. Unfortunately this also means using several notational conventions that can obscure the literal meaning of a formula, especially as certain natural operations involving differential operators are neither commutative nor associative. Thus in this section we spell out the meaning of the various formulae we will use throughout the rest of the chapter.

3.1.1. *Local coordinates.* Let's work in a local chart with smooth coordinates x^i. We use abbreviations $\partial_i := \partial/\partial x^i$ and $\nabla_i := \nabla_{\partial_i}$. For a tensor, lower indices are covariant and upper indices are contravariant, so a tensor $T \in \Gamma(\otimes^k T^*M \otimes^l TM)$ is written as

$$T = T_{a_1 a_2 \cdots a_k}^{b_1 b_2 \cdots b_l} dx^{a_1} \otimes \cdots \otimes dx^{a_k} \otimes \partial_{b_1} \otimes \cdots \otimes \partial_{b_l}$$

Usually the dx^i and ∂_j terms are omitted, so that the tensor is denoted just as $T_{a_1 a_2 \cdots a_k}^{b_1 b_2 \cdots b_l}$. Sometimes we use a single letter to denote a multi-index, e.g. $\alpha := a_1 a_2 \cdots a_k$ and write T_α^β. We use the Einstein summation convention that repeated indices (one upper, one lower) indicate summation, e.g. $X^i Y_i$ really means $\sum_i X^i Y_i$.

3.1.2. *The metric tensor g.* The metric g is a symmetric 2-form, i.e. a section of $S^2 T^*M$. At each point p it determines a positive definite inner product $g(X,Y)_p$, also written $\langle X, Y\rangle_p$. Typically the point p is omitted. In local coordinates,

$$g = g_{ij} dx^i \otimes dx^j$$

where $g_{ij} = g_{ji}$. We write the inverse of the matrix g_{ij} as g^{ij}; i.e. $g^{ij} g_{jk} = \delta_k^i$ (remember the summation convention). The metric gives a canonical identification between TM and T^*M, which we can use to raise or lower the indices of a tensor. Thus

$$g^{ij} T_{\alpha i}^\beta = T_\alpha^{\beta j} \text{ and } g_{ij} T_\alpha^{\beta i} = T_{\alpha j}^\beta$$

In the special case of a vector field $X = X^i \partial_i$ we denote the dual 1-form by $X^\flat$, so that $X^\flat = g_{ji} X^i dx^j = X_j dx^j$. Likewise for a 1-form $\alpha = \alpha_i dx^i$ the dual vector field is $\alpha^\sharp = g^{ij} \alpha_i \partial_j = \alpha^j \partial_j$. For a function f the gradient $\text{grad} f$ is by definition $\text{grad} f = (df)^\sharp$.

Partial derivatives of the g_{ij} and g^{ij} are related by

$$0 = \partial_l (g^{ij} g_{jk}) = (\partial_l g^{ij}) g_{jk} + g^{ij} (\partial_l g_{jk})$$

3.1.3. *The Levi-Civita connection ∇.* A *connection* ∇ on a smooth vector bundle V over M is a rule that takes a vector field X on M and a section σ of V and produces another section $\nabla_X \sigma$ of V which is tensorial in X, and satisfies a Leibniz rule $\nabla_X f\sigma = X(f)\sigma + f\nabla_X \sigma$ for any smooth function f.

Given g there is a unique connection ∇ on TM called the Levi-Civita connection which preserves the metric and is torsion-free; i.e.

$$Z(g(X,Y)) = g(\nabla_Z X, Y) + g(X, \nabla_Z Y) \text{ and } \nabla_X Y - \nabla_Y X = [X,Y]$$

for all vector fields X, Y, Z. It satisfies the *Koszul formula* (which can be taken as a definition):

$$\langle \nabla_X Y, Z\rangle = \frac{1}{2}\{Xg(Y,Z)+Yg(Z,X)-Zg(X,Y)+g([X,Y],Z)-g([Y,Z],X)-g([X,Z],Y)\}$$

The connection is not a tensor, but the difference of two connections on the same bundle is a tensor. Local coordinates define a 'trivial' connection $\tilde{\nabla}$ satisfying $\tilde{\nabla}_i \partial_j = 0$ and the difference between ∇ and $\tilde{\nabla}$ is expressed locally with the Christoffel symbols

$$\nabla_i \partial_j = \Gamma_{ij}^k \partial_k \text{ where } \Gamma_{ij}^k = \frac{1}{2} g^{kl}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})$$

Note that $\Gamma_{ij}^k = \Gamma_{ji}^k$ since ∇ is torsion-free.

3.1.4. *Connections on other bundles.* The connection ∇ on TM determines a connection on T^*M that we also denote ∇, by the formula

$$Y(\alpha(X)) = (\nabla_Y \alpha)(X) + \alpha(\nabla_Y X)$$

By the Leibniz rule this gives a connection on every $\otimes^k T^*M \otimes^l TM$. In coordinates, $\nabla_i dx^j = -\Gamma_{ik}^j dx^k$.

If $T = T_\beta^\alpha dx^\beta \otimes \partial_\alpha$ is a tensor (where α, β are multi-indices), we typically want to compute the coefficients of ∇T. Here we use the potentially misleading, but common convention that

$$\nabla_i T_\beta^\alpha := (\nabla T)_{i\beta}^\alpha = \partial_i T_\beta^\alpha + \sum_k \Gamma_{il}^{\alpha_k} T_\beta^{\alpha_1 \cdots l \cdots \alpha_{|\alpha|}} - \sum_k \Gamma_{i\beta_k}^l T_{\beta_1 \cdots l \cdots \beta_{|\beta|}}^\alpha$$

With this convention, taking higher covariant derivatives is 'associative'; i.e.

$$\nabla_i \nabla_j T_\alpha^\beta := (\nabla\nabla T)_{ij\alpha}^\beta$$

and so on. Since ∇ is a metric connection, $\nabla_i g_{jk} := (\nabla g)_{ijk} = 0$ so taking covariant derivatives commutes with contraction of indices.

The alternative convention is to use $\nabla_i T$ to denote the result of contracting the tensor ∇T with the vector field ∂_i. If this convention is meant we denote it with brackets. Hence

$$\nabla_i \nabla_j T = \nabla_i(\nabla_j T) - \nabla_{\nabla_i \partial_j} T$$

Fortunately the commutator $(\nabla_i \nabla_j - \nabla_j \nabla_i)T$ is the same in either convention.

At a point p at the center of normal coordinates we have $\Gamma_{ij}^k = 0$ so that $\nabla_i T_\beta^\alpha = \partial_i T_\beta^\alpha$. If the quantity we are computing is a tensor, an equality which holds in special coordinates at a point holds everywhere. This tremendously simplifies several formulae, as we shall see, particularly in § 3.2.

3.1.5. *Hessian and Laplacian.* The *Hessian* of a tensor T is the second covariant derivative $\text{Hess}(T) := \nabla\nabla T$ and the *rough Laplacian* is the trace of the Hessian. That is,

$$\Delta T := \operatorname{tr} \text{Hess}(T) = \nabla^* \nabla T = g^{ij} \nabla_i \nabla_j T$$

For a function f we have $\text{Hess}(f) = \nabla df$, and Δf agrees with the usual Hodge–de Rham Laplacian applied to f. Note that this is the 'analyst's Laplacian', with nonpositive spectrum.

We shall use the same convention for the components of the rough Laplacian applied to tensors as we do with covariant derivatives, e.g.

$$\Delta T^\alpha_\beta := (\Delta T)^\alpha_\beta = g^{ij}\nabla_i\nabla_j T^\alpha_\beta$$

3.1.6. *Lie derivative.* A vector field X on a closed manifold generates a flow φ_t, and for any covariant (resp. contravariant) tensor field T we can push forward (resp. pull back) T under this flow and compute the derivative with respect to t at $t = 0$. The result is called the *Lie derivative of* T *in the direction* X, denoted $\mathcal{L}_X T$.

For a function f we have $\mathcal{L}_X f = X(f)$ and for a vector field Y we have $\mathcal{L}_X Y = [X, Y]$. For a k-form α Cartan's 'magic formula' says $\mathcal{L}_X\alpha = \iota_X d\alpha + d\iota_X\alpha$ where ι_X is contraction with X. For other tensors one can compute Lie derivative by the Leibniz rule. For instance, if g denotes the metric, then for any vector field $X = X^i\partial_i$,

$$g(\nabla_X\partial_i, \partial_j) + g(\partial_i, \nabla_X\partial_j) = X(g(\partial_i, \partial_j)) = (\mathcal{L}_X g)(\partial_i, \partial_j) + g([X, \partial_i], \partial_j) + g(\partial_i, [X, \partial_j])$$

so that

$$\mathcal{L}_X g_{ij} := (\mathcal{L}_X g)_{ij} = g(\nabla_i X, \partial_j) + g(\partial_i, \nabla_j X) = g_{kj}\nabla_i X^k + g_{ki}\nabla_j X^k = \nabla_i X_j + \nabla_j X_i$$

Equivalently, for a 1-form α we have $\mathcal{L}_{\alpha^\sharp} g_{ij} = \nabla_i\alpha_j + \nabla_j\alpha_i$. As a special case,

$$\mathcal{L}_{\text{grad} f}\, g_{ij} = 2\text{Hess}(f)$$

3.1.7. *The Riemann curvature tensor* $\mathcal{R}$. The *curvature tensor* $\mathcal{R}$ is defined by

$$\mathcal{R}(X, Y)Z := \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$$

for all vector fields X, Y, Z. The fact that it is tensorial in all three entries X, Y and Z follows from a calculation. By abuse of notation we sometimes write

$$\mathcal{R}(X, Y, Z, W) := \langle\mathcal{R}(X, Y)Z, W\rangle$$

In local coordinates,

$$\mathcal{R}(\partial_i, \partial_j)\partial_k = R^l_{ijk}\partial_l \text{ where } R^l_{ijk} = \partial_i\Gamma^l_{jk} - \partial_j\Gamma^l_{ik} + \Gamma^p_{jk}\Gamma^l_{ip} - \Gamma^p_{ik}\Gamma^l_{jp}$$

We also write

$$\langle\mathcal{R}(\partial_i, \partial_j)\partial_k, \partial_l\rangle = R_{ijkl} = g_{lm}R^m_{ijk}$$

It satisfies several symmetries, most prominently

$$R_{ijkl} = R_{klij}, \quad -R_{ijlk} = R_{ijkl} = -R_{jikl}, \text{ and } R_{ijkl} + R_{jkil} + R_{kijl}$$

The first two symmetries together imply that the curvature can be thought of as a section of $S^2\Lambda^2T^*M$; i.e. as a symmetric quadratic form on Λ^2T^*M. The third symmetry is called the first, or *algebraic* Bianchi identity. The second, or *differential* Bianchi identity is

$$\nabla_m R^l_{ijk} + \nabla_i R^l_{jmk} + \nabla_j R^l_{mik}$$

(remember that a term like $\nabla_m R^l_{ijk}$ means the ∂_l component of $(\nabla_m\mathcal{R})(\partial_i, \partial_j)\partial_k$, and so on).

Recall that $\nabla_i dx^j = -\Gamma^j_{ik}dx^k$. Thus

$$(\nabla_i\nabla_j - \nabla_j\nabla_i)dx^k = R^k_{jil}dx^l$$

Thus for vector fields $X = X^k\partial_k$ and 1-forms $\alpha = \alpha_k dx^k$ we obtain formulae

$$\nabla_i\nabla_j X^k - \nabla_j\nabla_i X^k = R^k_{ijl}X^l$$

and

$$\nabla_i\nabla_j\alpha_k - \nabla_j\nabla_i\alpha_k = R^l_{jik}\alpha_l = g^{lm}R_{jikm}\alpha_l$$

and so on by the Leibniz rule for other tensors.

3.1.8. *Curvatures K,* Rm*,* Ric *and R.* The *sectional curvature* of the 2-plane spanned by non-parallel vectors X, Y is the number

$$K(X,Y) := \langle \mathcal{R}(X,Y)Y, X\rangle / \|X\wedge Y\|^2$$

We denote by Rm the curvature tensor thought of as a symmetric quadratic form on $\Lambda^2 T^*M$, with the convention that $\mathrm{Rm}(X\wedge Y, Z\wedge W) := \langle \mathcal{R}(X,Y)W, Z\rangle$. Since $X\wedge Y := X\otimes Y - Y\otimes X$ the eigenvalues of Rm are equal to twice the sectional curvatures.

The *Ricci curvature* is obtained by taking the trace, i.e.

$$\mathrm{Ric}(X,Y) = \sum_i \langle \mathcal{R}(e_i, X)Y, e_i\rangle$$

for any orthonormal basis e_i. If v is a unit vector, $\mathrm{Ric}(v,v)$ is equal to $(n-1)$ times the average of K over all 2-planes containing v. In local coordinates,

$$\mathrm{Ric} = R_{ij}dx^i\otimes dx^j \text{ where } R_{ij} = R^k_{kij} = \partial_k\Gamma^k_{ij} - \partial_i\Gamma^k_{kj} + \Gamma^m_{ij}\Gamma^k_{km} - \Gamma^m_{kj}\Gamma^k_{im}$$

$$= \frac{1}{2}g^{kl}(\partial_j\partial_k g_{il} - \partial_l\partial_k g_{ij} - \partial_i\partial_j g_{kl} + \partial_i\partial_l g_{kj}) + \text{lower order derivatives}$$

The *scalar curvature* R is obtained by taking a further trace

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

Note that Ric is a section of S^2T^*M while R is a function.

3.1.9. *Sign errors.* There are many opportunities for sign errors in these formulae. One major source of such errors are the multiple competing conventions for the definitions of $\mathcal{R}$ and R_{ijkl}. Everyone should agree on the signs of R and K and the eigenvalues of the symmetric quadratic form Ric. The eigenvalues of Rm depend on the choice of normalization of wedge product of 1-forms; different normalizations result in a factor of 2.

3.2. **Evolution of Curvature.** Let's suppose we have a manifold and a family $g_{ij}(t)$ of smooth metrics. Write $h_{ij} = \partial_t g_{ij}$, and observe that $0 = \partial_t(g^{ij}g_{jk}) = (\partial_t g^{ij})g_{jk} + g^{ij}h_{jk}$ so that $\partial_t g^{ij} = -g^{jk}g^{il}h_{lk}$.

Recall that the Christoffel symbols are defined by

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

The connection ∇ is not a tensor, but its time derivative is. Thus we can compute the derivative $\partial_t\Gamma^k_{ij}$ at a point p in normal coordinates where $\partial_i g_{jk} = 0$. At such a point

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

$$= \frac{1}{2}g^{kl}(\nabla_i h_{jl} + \nabla_j h_{il} - \nabla_l h_{ij})$$

and since both sides are components of tensors, equality holds generally (remember our notational convention for the components of the covariant derivatives of a tensor; see § 3.1.4). Using the equality

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

and once more computing at a point in normal coordinates where $\Gamma^k_{ij} = 0$ we obtain

$$\partial_t R^l_{ijk} = \frac{1}{2} g^{lp} \left(\nabla_i \nabla_j h_{kp} + \nabla_i \nabla_k h_{jp} - \nabla_i \nabla_p h_{jk} - \nabla_j \nabla_i h_{kp} - \nabla_j \nabla_k h_{ip} + \nabla_j \nabla_p h_{ik} \right) \tag{3.1}$$

Now if we suppose the metrics $g(t)$ are a solution of Ricci flow, then $h_{ij} = -2R_{ij}$ and therefore

$$\partial_t \Gamma^k_{ij} = -g^{kl}(\nabla_i R_{jl} + \nabla_j R_{il} - \nabla_l R_{ij})$$

and

$$\partial_t R^l_{ijk} = g^{lp} \left(-\nabla_i \nabla_j R_{kp} - \nabla_i \nabla_k R_{jp} + \nabla_i \nabla_p R_{jk} + \nabla_j \nabla_i R_{kp} + \nabla_j \nabla_k R_{ip} - \nabla_j \nabla_p R_{ik} \right)$$

Remarkably, the right hand side can be re-written as ΔR^l_{ijk} plus a term which is quadratic in the curvature:

Proposition 3.1. *Let $g(t)$ be a solution to Ricci flow. Then*

$$\begin{aligned} \partial_t R^l_{ijk} = \Delta R^l_{ijk} &+ g^{pq}(R^r_{ijp}R^l_{rqk} - 2R^r_{pik}R^l_{jqr} + 2R^l_{pir}R^r_{jqk}) \\ &- R^p_i R^l_{pjk} - R^p_j R^l_{ipk} - R^p_k R^l_{ijp} + R^l_p R^p_{ijk} \end{aligned}$$

where $R^p_i := g^{pj}R_{ij}$ and so on.

Proof. First observe that we can rewrite

$$-\nabla_i \nabla_j R_{kp} + \nabla_j \nabla_i R_{kp} = g^{qm}(R_{ijkm}R_{qp} + R_{ijpm}R_{qk})$$

Recall the definition of the rough Laplacian as

$$\Delta R^l_{ijk} = g^{mq}\nabla_m \nabla_q R^l_{ijk} = g^{mq}\nabla_m(-\nabla_i R^l_{jqk} - \nabla_j R^l_{qik})$$

by the second Bianchi identity. The difference between $\nabla_m \nabla_i$ and $\nabla_i \nabla_m$ is a curvature term, so we can write

$$\Delta R^l_{ijk} = g^{mq}(\text{curvature term} - \nabla_i \nabla_m R^l_{jqk} - \nabla_j \nabla_m R^l_{qik})$$

Now $R^l_{jqk} = g^{lp}R_{jqkp} = g^{lp}R_{kpjq}$ and $R^l_{qik} = g^{lp}R_{qikp} = g^{lp}R_{kpqi}$ so applying the second Bianchi identity again and then contracting,

$$\begin{aligned} \Delta R^l_{ijk} + \text{curvature term} &= g^{mq}g^{lp}(-\nabla_i \nabla_m R_{kpjq} - \nabla_j \nabla_m R_{kpqi}) \\ &= g^{mq}g^{lp}(\nabla_i \nabla_k R_{pmjq} + \nabla_i \nabla_p R_{mkjq} + \nabla_j \nabla_k R_{pmqi} + \nabla_j \nabla_p R_{mkqi}) \\ &= g^{lp}(-\nabla_i \nabla_k R_{pj} + \nabla_i \nabla_p R_{kj} + \nabla_j \nabla_k R_{pi} - \nabla_j \nabla_p R_{ki}) \end{aligned}$$

Collecting curvature terms gives the result. □

Now, $\partial_t R_{ijkl} = \partial_t g_{lm} R^m_{ijk} = -2R_{lm}R^m_{ijk} + g_{lm}\partial_t R^m_{ijk}$ so we have

$$\partial_t R_{ijkl} = \Delta R_{ijkl} + \text{quadratic curvature term}$$

and likewise for $\partial_t R_{ij}$ and $\partial_t R$. Explicitly, one has the following formulae:

Theorem 3.2 (Hamilton, [12] 7.1, 7.3, 7.5)**.** *Let $g(t)$ be a solution to Ricci flow. Define the tensor $B_{ijkl} := g^{pr}g^{qs}R_{pijq}R_{rkls}$*

Then we have the following evolution formulae for curvature. For R_{ijkl}:

$$\partial_t R_{ijkl} = \Delta R_{ijkl} - 2(B_{ijkl} - B_{ijlk} - B_{iljk} + B_{ikjl}) \\ - g^{pq}(R_{pjkl}R_{qi} + R_{ipkl}R_{qj} + R_{ijpl}R_{qk} + R_{ijkp}R_{ql})$$

For R_{ij}:

$$\partial_t R_{ij} = \Delta R_{ij} + 2g^{pr}g^{qs}R_{pijq}R_{rs} - 2g^{pq}R_{pi}R_{qk}$$

For R:

$$\partial_t R = \Delta R + 2g^{ij}g^{kl}R_{ik}R_{jl} = \Delta R + 2|\mathrm{Ric}|^2$$

3.3. **The Ricci flow is not parabolic.** The Ricci curvature is a 2nd order differential operator from metrics to symmetric 2-forms. Although it's nonlinear, it is *semilinear* — i.e. linear in the derivatives of highest order. If we write $\Gamma(S^2_+T^*M)$ for positive definite symmetric 2-forms, then $\mathrm{Ric} : \Gamma(S^2_+T^*M) \to \Gamma(S^2T^*M)$. The derivative $d\mathrm{Ric}$ at any specific metric g is therefore a linear map $d\mathrm{Ric}_g : \Gamma(S^2T^*M) \to \Gamma(S^2T^*M)$. By contracting indices in equation 3.1 we get the formula

$$d\mathrm{Ric}_g(h_{ij}) = \frac{1}{2}g^{pq}(\nabla_q\nabla_i h_{jp} + \nabla_q\nabla_j h_{ip} - \nabla_q\nabla_p h_{ij} - \nabla_i\nabla_q h_{jp} - \nabla_i\nabla_j h_{pq} + \nabla_i\nabla_p h_{qj})$$

$$= \frac{1}{2}g^{pq}(\nabla_q\nabla_i h_{jp} + \nabla_q\nabla_j h_{ip} - \nabla_q\nabla_p h_{ij} - \nabla_i\nabla_j h_{pq}) \tag{3.2}$$

(the 4th and 6th term cancel after contraction with g^{pq}).

If P is a linear differential operator of order k between sections of bundles E and F, the *symbol* of P (denoted $\sigma(P)$) is the homogeneous term of highest order in the Fourier transform of P. In other words, $\sigma(P)$ is a tensor, i.e. a $C^\infty(M)$-linear map

$$\sigma(P) : \Gamma(S^kT^*M \otimes E) \to \Gamma(F)$$

In local coordinates, we replace each differential operator ∂_j by a formal dual variable ξ_j which is a coordinate on the cotangent bundle, and take the homogeneous polynomial in ξ of highest order. In our case we have

$$\sigma(d\mathrm{Ric}_g)(\xi) : \Gamma(S^2T^*M) \to \Gamma(S^2T^*M)$$

given by

$$\sigma(d\mathrm{Ric}_g)(\xi)(h_{ij}) = \frac{1}{2}g^{pq}(\xi_q\xi_i h_{jp} + \xi_q\xi_j h_{ip} - \xi_q\xi_p h_{ij} - \xi_i\xi_j h_{pq})$$

Now let's specialize to the case of a 2nd order (possibly nonlinear) differential operator $P : \Gamma(E) \to \Gamma(E)$ and let's fix a metric on (the fibers of) E. The equation $\partial_t\theta = P\theta$ is said to be *parabolic* if for every point p and every ξ nonzero at p, the inner product $\langle\sigma(P)(\xi)(h), h\rangle$ is positive definite on E_p (at least if M is compact — the only case we consider).

It will turn out for $P = -2\mathrm{Ric}_g$ that the symbol $\sigma(-2d\mathrm{Ric}_g)$ is degenerate. Let's see why. The reason is the diffeomorphism invariance of the Ricci flow. This gives an enormous (infinite dimensional) family of symmetries of the flow, and deformations tangent to these symmetries will be degenerate.

More precisely: the metric and curvature both pull back under any diffeomorphism $\varphi : M \to M$; i.e. $\text{Ric}(\varphi^* g) = \varphi^* \text{Ric}(g)$. If φ_t is a 1-parameter family of diffeomorphisms generated by a vector field X, we can differentiate this equality to get

$$d\text{Ric}_g(\mathcal{L}_X g) = \mathcal{L}_X \text{Ric}_g$$

Define an operator $Q : \Gamma(T^*M) \to \Gamma(S^2T^*M)$ by $Q(\alpha) := \mathcal{L}_{\alpha^\sharp} g_{ij} = \nabla_i \alpha_j + \nabla_j \alpha_i$. Then

$$d\text{Ric}_g \circ Q(\alpha) = \mathcal{L}_{\alpha^\sharp} \text{Ric}_g$$

The right hand side is first order in α, whereas the left hand side is *a priori* third order, so (when thought of as a third order operator!) its symbol vanishes. Taking symbols commutes with composition, so

$$0 = \sigma(d\text{Ric}_g \circ Q) = \sigma(d\text{Ric}_g)\sigma(Q)$$

In particular, $\sigma(d\text{Ric}_g)(\xi)$ vanishes on expressions of the form $(\xi_i \alpha_j + \xi_j \alpha_i)$, so that Ricci flow is not parabolic.

3.4. **The DeTurck trick.** Nevertheless, Hamilton [12] demonstrated short time existence and uniqueness for Ricci flow with arbitrary smooth initial metric on a compact manifold. In fact, he obtained a lower bound on the lifetime of a maximal solution of the form const./ $\max |\text{Rm}|$ where the constant depends only on dimension.

Hamilton's proof is technically difficult, and relies on the Nash–Moser Inverse Function Theorem. DeTurck [10] gave a much simpler proof. As we have seen, the degeneracy of the symbol comes from the naturalness of the flow. DeTurck's trick is to modify Ricci flow by adding a suitable Lie derivative term, cancelling the degeneracy. The resulting flow is parabolic and enjoys short term existence and uniquess. On the other hand, if we evolve the manifold by this modified flow together with a family of diffeomorphisms which undo the effect of the Lie derivative term, we recover ordinary Ricci flow and prove existence (uniqueness requires a little more work).

Let's take another look at the symbol of $-2d\text{Ric}_g$. The term $g^{pq}\nabla_p\nabla_q h_{ij} = \Delta h_{ij}$ contributes $|\xi|^2 h_{ij}$ to the symbol. This is promising, but we have still to understand the highest order contribution of the other terms.

We can switch the order of covariant derivatives at the cost of introducing curvature terms. However, these curvature terms are tensors — i.e. 0th order — and therefore make no difference to the symbol. So

$$-2d\text{Ric}_g(h_{ij}) = \Delta h_{ij} - \nabla_i g^{pq}\nabla_q h_{jp} - \nabla_j g^{pq}\nabla_q h_{ip} + g^{pq}\nabla_i\nabla_j h_{pq} + \text{lower order terms}$$

If we define a 1-form $V := V_k dx_k$ by

$$V_k := g^{pq}\nabla_p h_{qk} - \frac{1}{2}\nabla_k(g^{pq}h_{pq})$$

then by substitution,

$$-2d\text{Ric}_g(h_{ij}) = \Delta h_{ij} - \nabla_i V_j - \nabla_j V_i + \text{lower order terms}$$

Now, $\nabla_i V_j + \nabla_j V_i$ is nothing but $\mathcal{L}_{V^\sharp} g$. This suggests that we should try to find a natural differential operator from metrics to vector fields $\rho : \Gamma(S^2_+ T^*M) \to \Gamma(TM)$ for which $d\rho(h_{ij}) = V^\sharp$, and then the leading term of $-2\text{Ric}_g + \mathcal{L}_{\rho(g)} g$ will be Δh_{ij} with symbol $|\xi|^2 h_{ij}$.

The formula for V_k looks very similar to the contraction of a Christoffel symbol, with the metric term g_{ij} replaced by h_{ij}. Now, the Christoffel symbol is not itself a well-defined tensor, but if we fix a background metric $\tilde{g}$ with Christoffel symbols $\tilde{\Gamma}_{ij}^k$, then for any other metric g the difference $\Gamma_{ij}^k - \tilde{\Gamma}_{ij}^k$ *is* an honest tensor, and we can define the vector field $W^k := g^{pq}(\Gamma_{pq}^k - \tilde{\Gamma}_{pq}^k)$.

Now, let's define the modified flow $\partial_t g = -2\mathrm{Ric} + \mathcal{L}_W g$. Define $P(g) := -2\mathrm{Ric}_g + \mathcal{L}_W g$ so that

$$\begin{aligned} d(\mathcal{L}_W g)(h_{ij}) &= \nabla_i g_{jk} g^{pq} d\Gamma_{pq}^k(h_{ij}) + \nabla_j g_{ik} g^{pq} d\Gamma_{pq}^k(h_{ij}) \\ &= \nabla_i g_{jk} g^{pq} \left(\frac{1}{2} g^{kl} (\nabla_p h_{ql} + \nabla_q h_{pl} - \nabla_l h_{pq}) \right) + \text{ similar term} \\ &= \nabla_i g^{pq} (\nabla_q h_{pj} - \frac{1}{2} \nabla_j h_{pq}) + \nabla_j g^{pq} (\nabla_q h_{pi} - \frac{1}{2} \nabla_i h_{pq}) \end{aligned}$$

so that $\sigma(-2d\mathrm{Ric}_g + d\mathcal{L}_W g)(\xi)(h_{ij}) = |\xi|^2 h_{ij}$. Thus the modified flow is parabolic, and has short time existence and uniqueness. Composing modified flow with the inverse of the flow generated by the (time-dependent) vector field W recovers ordinary Ricci flow.

3.5. **Scalar Maximum Principle.** Fix a compact manifold M, a (time-dependent) vector field V, and a function ψ of a real variable. A *heat equation* is an equation of the form

$$\partial_t f = \Delta f + V(f) + \psi(f)$$

for some smooth function f. The simplest case is that $V = \psi = 0$; i.e. $\partial_t f = \Delta f$. This says that the value of f at each point evolves by moving in the direction of the average of nearby values. Solutions to this equation satisfy a *maximum principle* which we state in the following way. Suppose $\partial_t f = \Delta f$, and suppose at time 0 the values of f all lie in a closed convex set $K \subset \mathbb{R}$. Then the values of $f(t)$ lie in K for all positive t.

Now let's consider $\partial_t f = \Delta f + \psi$. Suppose $f(t) \in K$ for all t up to some first time t_0 when there is some p with $f(t_0)(p) \in \partial K$. Then $f(t_0) \in K$ so $\Delta f(t_0)(p) = 0$. Suppose that there is a relationship between f and ψ (in many important cases ψ is a function of f) so that ψ points into the interior of K whenever $f \in \partial K$. Then the maximum principle applies, and we deduce $f(t) \in K$ for all t. We can even let K depend on t, in which case we need ψ to dominate $\partial_t \partial K(t)$ (in the obvious sense) whenever $f \in \partial K(t)$.

The first place we want to apply this is to the evolution equation for scalar curvature

$$\partial_t R = \Delta R + 2|\mathrm{Ric}|^2$$

Evidently $|\mathrm{Ric}|^2 \geq 0$ so we can take K to be any set of the form $[r, \infty)$ where the obvious choice for r is $R_{\min}(0)$, the spatial minimum of R at time 0 (achieved, since M is compact). If $R < 0$ somewhere at time 0 we can do better. R is the trace of Ric, so Cauchy–Schwarz implies $|\mathrm{Ric}|^2 \geq R^2/n$. Thus we conclude:

Proposition 3.3 (Monotonicity of $R_{\min}$)**.** *The spatial minimum $R_{min}(t)$ is monotone non-decreasing under Ricci flow and satisfies*

$$R_{min}(t) \geq \frac{R_{min}(0)}{1 - 2tR_{min}(0)/n}$$

In particular, if $R_{min}(0)$ is positive, the metric becomes singular in time $O(1/R_{min}(0))$.

3.5.1. *Higher derivatives of curvature.* The maximum principle can be used to give control over the spatial derivatives of curvature as a function of time.

Theorem 3.2 gives a formula for the evolution of the curvature tensor of the form

$$\partial_t \mathrm{Rm} = \Delta \mathrm{Rm} + O(\mathrm{Rm}^2)$$

where $O(\mathrm{Rm}^2)$ denotes an unspecified term quadratic in Rm. From this we can compute

$$\begin{aligned}\partial_t |\mathrm{Rm}|^2 &= 2\langle \partial_t \mathrm{Rm}, \mathrm{Rm}\rangle = 2\langle \Delta \mathrm{Rm}, \mathrm{Rm}\rangle + O(|\mathrm{Rm}|^3) \\ &= \Delta |\mathrm{Rm}|^2 - 2|\nabla \mathrm{Rm}|^2 + O(|\mathrm{Rm}|^3)\end{aligned}$$

where now $O(\cdot)$ is the usual big-O notation for a function. From this one can give an a priori estimate on the rate of blow-up of $|\mathrm{Rm}|$. Denote by $|\mathrm{Rm}|_{\max}(t)$ the maximum of $|\mathrm{Rm}|$ at time t. At a spatial maximum for $|\mathrm{Rm}|^2$ we have $\Delta|\mathrm{Rm}|^2 \le 0$ so the time derivative of $|\mathrm{Rm}|^2_{\max}(t)$ is bounded above by $2C|\mathrm{Rm}|^3_{\max}(t)$ for some C. Therefore we obtain the following counterpart to Proposition 3.3:

Proposition 3.4 (Curvature blow-up rate). *There is a constant C depending only on the dimension so that*

$$|\mathrm{Rm}|_{max}(t) \le \frac{|\mathrm{Rm}|_{max}(0)}{1 - Ct|\mathrm{Rm}|_{max}(0)}$$

One application is a doubling-time estimate: if the norm of the curvature is $\le K$ at time 0 it stays $\le 2K$ up to time at least $1/2CK$.

For any tensor T we have $\Delta T = \mathrm{tr}\nabla^2 T$ so commuting ∇ with Δ gives rise to an identity of the form

$$\nabla \Delta T = \Delta \nabla T + O(T, \nabla \mathrm{Rm}) + O(\nabla T, \mathrm{Rm})$$

where $O(A, B)$ denotes a term linear in each of A and B. Thus

$$\nabla \Delta \mathrm{Rm} = \Delta \nabla \mathrm{Rm} + O(\nabla \mathrm{Rm}, \mathrm{Rm})$$

Likewise, the effect of commuting ∂_t with ∇ contributes an $O(T, \nabla \mathrm{Rm})$ term coming from the time derivative of the metric (and hence the connection). Putting these contributions together for $T = \mathrm{Rm}$ gives rise to an identity of the form

$$\partial_t \nabla \mathrm{Rm} = \Delta \nabla \mathrm{Rm} + O(\nabla \mathrm{Rm}, \mathrm{Rm})$$

and consequently

$$\partial_t |\nabla \mathrm{Rm}|^2 = \Delta |\nabla \mathrm{Rm}|^2 - 2|\nabla^2 \mathrm{Rm}|^2 + O(|\nabla \mathrm{Rm}|^2 |\mathrm{Rm}|)$$

Inductively, one obtains an estimate of the form

$$\partial_t |\nabla^m \mathrm{Rm}|^2 = \Delta |\nabla^m \mathrm{Rm}|^2 - 2|\nabla^{m+1} \mathrm{Rm}|^2 + O\Big(\sum_{i+j=m} |\nabla^i \mathrm{Rm}||\nabla^j \mathrm{Rm}||\nabla^m \mathrm{Rm}|\Big)$$

Using this we obtain the following estimate:

Theorem 3.5 (Curvature derivative bounds). *Suppose $|\mathrm{Rm}| \le K$ for all $x \in M$ and all times in the interval $t \in (0, t_0/K]$. Then for each integer m there is a constant C depending on m, on t_0 and on the dimension of M, so that for any time t in the same interval there is an estimate*

$$|\nabla^m \mathrm{Rm}| \le \frac{CK}{t^{m/2}}$$

Notice that there is *no* a priori estimate on any $|\nabla^m \mathrm{Rm}|$ with $m > 0$ at time $t = 0$, and that the control gets *better* as time increases. Furthermore, by the doubling time estimate, if $|\mathrm{Rm}| \le K$ at time 0 then $|\mathrm{Rm}| \le 2K$ up to time $1/2CK$; so the hypothesis of the theorem is always satisfied for some K and t_0.

Proof. Let's examine the case $m = 1$. We want to prove an estimate of the form $|\nabla \mathrm{Rm}| \le CKt^{-1/2}$. For $t \to 0$ we have no control over $\nabla\mathrm{Rm}$, so instead we try to control an expression of the form

$$f(x,t) := t|\nabla\mathrm{Rm}|^2 + \alpha|\mathrm{Rm}|^2$$

for some α to be determined. We compute

$$\begin{aligned}\partial_t f &\le |\nabla\mathrm{Rm}|^2 + t(\Delta|\nabla\mathrm{Rm}|^2 + C|\nabla\mathrm{Rm}|^2|\mathrm{Rm}|) + \alpha(\Delta|\mathrm{Rm}|^2 - 2|\nabla\mathrm{Rm}|^2 + C|\mathrm{Rm}|^3)\\ &= \Delta f + |\nabla\mathrm{Rm}|^2(1 + Ct|\mathrm{Rm}| - 2\alpha) + C\alpha|\mathrm{Rm}|^3\end{aligned}$$

By hypothesis $t|\mathrm{Rm}| \le t_0$ and $|\mathrm{Rm}| \le K$ so

$$\partial_t f \le \Delta f + |\nabla\mathrm{Rm}|^2(1 + Ct_0 - 2\alpha) + C\alpha K^3$$

for some constant C depending only on the dimension. Thus if we take $\alpha = (1 + Ct_0)/2$ we can ignore the second term, and estimate $\partial_t f \le \Delta f + CK^3$ where now the constant depends on t_0. Since $f(0) \le \alpha K^2$, the maximum principle shows that for all time

$$f \le \alpha K^2 + CtK^3 \le CK^2$$

at the cost of adjusting constants, so that $|\nabla\mathrm{Rm}| \le (f/t)^{1/2} \le CKt^{-1/2}$.

The case case of higher spatial derivatives follows in a similar way. □

A more subtle argument due to Shi allows one to control the norm of $|\nabla^m\mathrm{Rm}|$ pointwise from only *local* control over $|\mathrm{Rm}|$. Explicitly, suppose we have an open subset U and we have a bound $|\mathrm{Rm}| \le K$ for all $p \in U$ and $t \in (0, t_0]$. Suppose that at time 0 the ball of radius r around p is contained in U. Then $|\nabla\mathrm{Rm}|^2 \le CK^2(1/r^2 + 1/t_0 + K)$ and similarly for higher derivatives.

3.6. **Hamilton's Tensor Maximum Principle.** The evolution formulae in Theorem 3.2 are all of the form $\partial_t T = \Delta T +$ lower order term. This is a tensorial version of the scalar heat equations considered in § 3.5. Hamilton [13] Thm 4.3 proved a version of the maximum principle for tensor equations that we state in the following way.

Suppose V is a tensor bundle over M. We want to give V a (fiberwise) metric and a connection, so that it makes sense to take the Laplacian of a section of V, to talk about convexity of subsets etc. The manifold M should admit an evolving Riemannian metric. There should be a relationship between V and M as follows. Although the metric on V is fixed, the evolving metric on M manifests itself by an evolving *connection* on V. We say that V is *natural* if the (fixed) fiberwise metric is parallel for the connection at all time.

Theorem 3.6 (Hamilton's Maximum Principle)**.** *Let V be a natural tensor bundle over M, and let Ψ be a vertical vector field on V. Suppose that there is a closed subset K of V which satisfies*

(1) the set K is fiberwise convex;
(2) for all t the set K is invariant under parallel transport; and

(3) for any $x \in M$ any solution to the ODE $\partial_t T(x) = \Psi(T(x))$ which starts in K must stay in K.

Suppose $T(t)$ is a (time-dependent) section of V which evolves by

$$\partial_t T = \Delta T + \Psi(T)$$

and satisfies $T(0) \subset K$. Then $T(t) \subset K$ for all t.

Most geometrically natural subsets K that arise in practice will be invariant under parallel transport.

The idea is to reduce the statement to the scalar maximum principle. A closed subset of a vector space is convex if it is the sublevel set of a convex function. So if we could find a parallel fiberwise convex function u with $u \le 0$ on K, to show that $T \subset K$ we just need to check that $u(T) \le 0$. The key inequality which lets us work with $u(T)$ in place of T is the following:

Lemma 3.7 (Laplacian inequality)**.** *Let V be a natural tensor bundle over M, and let $u : V \to \mathbb{R}$ be a function which is fiberwise convex, and invariant under parallel transport (at any time); one natural choice is to take u to be equal to the distance to K in each fiber. Then for any section T of V, at each fixed time t, the following inequality holds pointwise in M:*

$$\Delta u_p(T) \ge du_p(T_p)(\Delta T)$$

Proof. Fix a time t, and a point $p \in M$. We denote the connection on V at time t by ∇. Fix an orthonormal frame $e_i(p)$ for V_p, and extend it locally to an orthonormal frame e_i near p by parallel transport along radial geodesics (in M). Then at the point p, $\nabla e_i = 0$ and $\nabla^2 e_i$ is antisymmetric, since ∇ preserves the (fiberwise) metric. Thus $\Delta e_i := \text{tr}\nabla^2 e_i$ vanishes at p. If we write T locally as $T := \sum \tau_i e_i$ then $\nabla T = \sum (d\tau_i)e_i + \tau_i \nabla e_i$ and

$$\nabla^2 T = \sum (\nabla d\tau_i)e_i + d\tau_i \nabla e_i + \tau_i \nabla^2 e_i$$

so we deduce that $\Delta T = \sum (\Delta \tau_i) e_i$ at p.

Since the e_i are parallel along radial geodesics, and u is invariant under parallel transport, it follows that $u_q(e_i(q)) = u_p(e_i(p))$ for q near p. Thus

$$u(T)(q) := u_q\left(\sum \tau_i(q)e_i(q)\right) = u_p\left(\sum \tau_i(q)e_i(p)\right)$$

so if we differentiate,

$$d(u(T))(q) = du_p\left(\sum \tau_i(q)e_i(p)\right)\left(\sum d\tau_i(q)e_i(p)\right)$$

Differentiate again and evaluate at p to get

$$\nabla d(u(T))(p) = \text{Hess}(u_p)(T_p)\left(\nabla T, \nabla T\right) + du_p(T_p)\left(\sum (\nabla d\tau_i)e_i\right)$$

Here $\text{Hess}(u_p)$ means the Hessian of the function u_p in the vector space V_p; this is a symmetric quadratic form on V_p, and we are evaluating it at $T_p \in V_p$ on $(\nabla T, \nabla T)$. The result is a quadratic form *on* T_pM, whose value on a pair of vectors $X, Y \in T_pM$ is $\text{Hess}(u_p)(T_p)(\nabla_X T, \nabla_Y T)$.

Taking the trace of the left hand side gives $\Delta u_p(T)$. Furthermore,

$$\operatorname{tr}\operatorname{Hess}(u_p)(T_p)\,(\nabla T, \nabla T) = \sum \operatorname{Hess}(u_p)(T_p)\,(\nabla_i T, \nabla_i T)$$

and because u_p is convex, each term (and therefore the sum) is ≥ 0. Finally,

$$\operatorname{tr} du_p(T_p)\left(\sum(\nabla d\tau_i)e_i\right) = du_p(T_p)\left(\sum(\Delta\tau_i)e_i\right) = du_p(T_p)(\Delta T)$$

and the lemma is proved. □

Since u is time-invariant, $\partial_t du(T) = du(\partial_t T)$, and therefore by Lemma 3.7, the tensor PDE implies a scalar differential *inequality* $\partial_t u(T) - \Delta u(T) \leq du(\Psi(T))$ to which we can apply the ordinary scalar maximum principle.

3.6.1. *The Uhlenbeck trick.* In order to apply the tensor maximum principle, we must deal with the fact that the tensors we are interested in (e.g. Ric) are living in a bundle whose (fiberwise) metric is time-dependent. To get around this, we use a bookkeeping trick due to Karen Uhlenbeck. An orientable 3-manifold M is parallelizable, so for any metric $g(0)$ we can find a global orthonormal frame e_1, e_2, e_3. Under Ricci flow, we evolve this frame by $g(\partial_t e_a, e_b) = \operatorname{Ric}(e_a, e_b)$ and so on (note: it's safer to use letters like a, b, c for indices so as not to confuse them for coordinate directions i, j, k). In other words, $\partial_t e_a = \operatorname{Ric}(e_a, \cdot)^\sharp$.

Lemma 3.8. *The evolving frame e_a stays orthonormal under Ricci flow.*

Proof. Just differentiate

$$\partial_t(g(e_a, e_b)) = (\partial_t g)(e_a, e_b) + g(\partial_t e_a, e_b) + g(e_a, \partial_t e_b) = 0$$

□

The sections e_a at any given time determine a family of isomorphisms $\iota(t) : V \to TM$ where V is a rank 3 trivial vector bundle. Pulling back the metric gives V a fiberwise constant metric; pulling back the connection gives it a time-dependent family of connections, which nevertheless preserve the fiberwise metric. Thus V is natural in the sense of Theorem 3.6. One may construct in this way natural bundles isomorphic to T^*M, to S^2T^*M and so on where the evolution of some tensor of interest is subject to the maximum principle.

3.6.2. *Evolution of the Einstein tensor.* The Einstein tensor G, as arises in the theory of general relativity, is the symmetric 2-tensor $\operatorname{Ric} - Rg/2$. We denote by E its negative; i.e. $E := Rg/2 - \operatorname{Ric}$. In 3-dimensions if the eigenvalues of the curvature operator Rm are λ, μ, ν (say) then the eigenvalues of Ric are $\mu+\nu, \lambda+\nu, \lambda+\mu$ and $R = 2(\lambda+\mu+\nu)$. It follows that the Einstein tensor has eigenvalues λ, μ, ν so in three dimensions E corresponds to Rm under the isomorphism $\Lambda^2 T^*M \cong T^*M$, at least up to a constant.

From the evolution equations for Ric and R, one derives an equation of the form

$$\partial_t E = \Delta E + \Psi(E)$$

In an evolving orthonormal frame where E is diagonal with entries λ, μ, ν, the matrix $\Psi(E)$ is also diagonal with entries $\lambda^2 + \mu\nu$ and so forth; i.e. under the ODE $\partial_t E = \Psi(E)$ the eigenvalues evolve by

$$\lambda' = 2(\lambda^2 + \mu\nu), \quad \mu' = 2(\mu^2 + \lambda\nu), \quad \nu' = 2(\nu^2 + \lambda\mu)$$

This may be proved by a calculation, although it is worth remarking that invariance under parabolic rescaling implies that Ψ is homogeneous of degree 2, and equivariance under the orthogonal group shows that it has the same eigenspace decomposition as E.

Let's suppose that we have $R > 0$, which (as we have already seen) is preserved by Ricci flow. Since the derivatives of the eigenvalues are homogeneous of order 2, we may recover the flowlines of the ODE and their orientation (though not their parameterization) by projecting to the hyperplane $\lambda + \mu + \nu = 1$. Figure 3 shows how these quantities evolve with time.

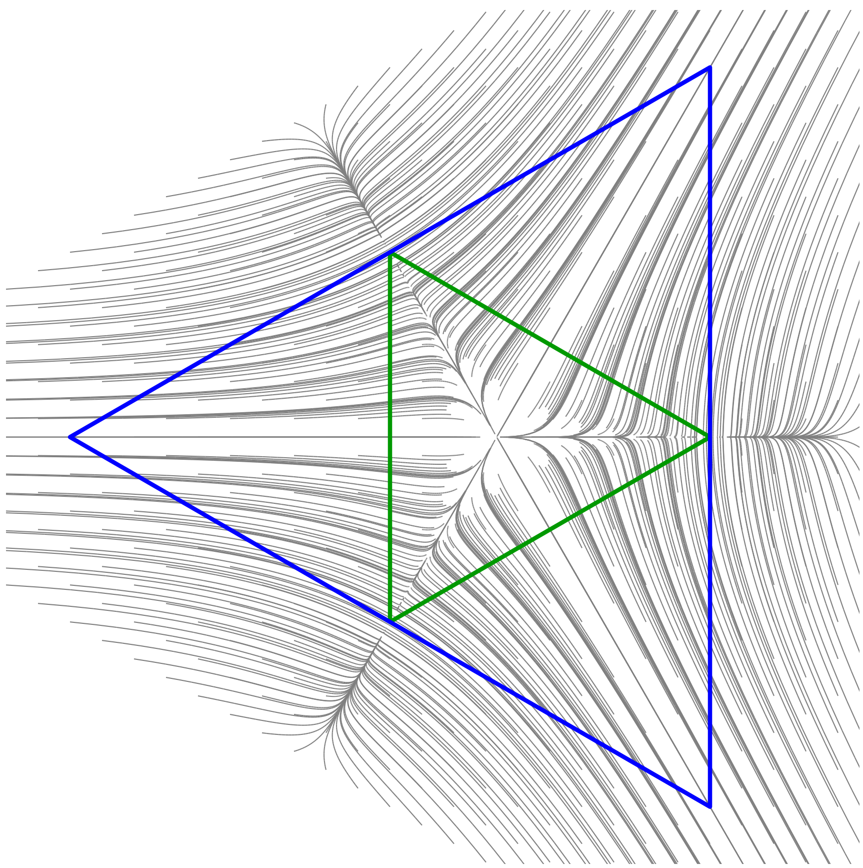

FIGURE 3. Projected flowlines of $\partial_t E = \Psi(E)$. The green triangle is the region where $\mathrm{Rm} \geq 0$ and the blue triangle is the region where $\mathrm{Ric} \geq 0$.

The convex region with $\mathrm{Rm} \geq 0$ is indicated by the green triangle, and the convex region with $\mathrm{Ric} \geq 0$ is indicated by the blue triangle. Evidently both these regions are preserved by $\partial_t E = \Psi(E)$. The vertices of the green triangle correspond to (projective) fixed points, where $\mu = \nu = 0, \lambda = 1$ corresponding to the gradient shrinking soliton $S^2 \times \mathbb{R}$. Hamilton's tensor maximum principle applied to these convex sets implies:

Theorem 3.9 (Non-negativity). *Let M be a closed 3-manifold, and suppose $M, g(t)$ satisfies Ricci flow. If* $\mathrm{Rm} \geq 0$ *(resp.* $\mathrm{Ric} \geq 0$*) for $t = 0$ then* $\mathrm{Rm} \geq 0$ *(resp.* $\mathrm{Ric} \geq 0$*) for all $t > 0$.*

It is evident from Figure 3 that concentrically scaled copies of the blue triangle are all taken inside themselves under the ODE, at least for $R \geq 0$. These are the level sets where the projective inequality $\text{Ric} \geq \epsilon R$ holds pointwise. If $\epsilon \in [0, 1/3)$ then $\text{Ric} \geq \epsilon R$ implies that $R \geq 0$ and we deduce the following:

Theorem 3.10 (Positive pinching)**.** *For any $\epsilon \in [0, 1/3)$ the inequality $\text{Ric} \geq \epsilon R$ is preserved by Ricci flow.*

3.6.3. *Roundness.* When Ric is strictly positive (corresponding to the interior of the blue triangle in Figure 3) every solution to the ODE $\partial_t E = \Psi(E)$ projectively converges to the origin where all eigenvalues are equal. We would like to conclude from the tensor maximum principle that the same is true for solutions to the PDE.

Let's suppose we can find a subset K of the bundle of symmetric 2-forms satisfying the hypotheses of Theorem 3.6 and with the property that pointwise, the level sets $K \cap \{\text{tr}(T) = C\}$ projectively converge to the origin as $C \to \infty$. A solution to the PDE which starts in K must stay there.

Since R is strictly positive, the spatial infimum of R must blow up in finite time, and therefore the ratio of the eigenvalues of Ric must tend to 1; colloquially this phenomenon is called *roundness*.

Here's a precise statement:

Theorem 3.11 (Roundness)**.** *Let M be a closed 3-manifold, and suppose $M, g(t)$ satisfies Ricci flow. Suppose there are positive constants $\alpha < \beta$ so that at time $t = 0$ we have $\alpha \leq \text{Ric} \leq \beta$ pointwise in the sense of operators. Then for any positive γ there is a constant C so that*

$$|\text{Ric} - Rg/3| \leq \gamma R + C$$

Proof. Let's order the eigenvalues of Rm as $\lambda \geq \mu \geq \nu$ so that at time $t = 0$ we have $\alpha \leq \mu + \nu$ and $\lambda + \mu \leq \beta$. Evidently $|\text{Ric} - Rg/3| \leq \lambda - \nu$, so we just need to control the right hand side.

The inequality $\mu + \nu \geq \alpha$ is convex, satisfied at $t = 0$, and preserved by the ODE, and therefore holds for all time for the PDE. Likewise the condition $\text{Ric} \geq \epsilon R$ for $\epsilon := (\alpha/3\beta)$ holds at time 0 and is preserved by the flow by Theorem 3.10; equivalently, $\mu + \nu \geq \delta\lambda$ for $\delta := 2\epsilon/(1 - 2\epsilon)$. If we define $\theta := 1/(1 + \delta/2)$ then $\theta \in (1/2, 1)$, and for some $A \gg 1$ we can ensure that at $t = 0$,

$$\lambda - \nu \leq A(\mu + \nu)^\theta$$

The eigenvalue λ is the max of linear functions, and is therefore convex. Likewise, ν and $\mu + \nu$ are concave. Thus $(\lambda - \nu) - A(\mu + \nu)^\theta$ is convex, and evidently invariant under parallel transport. We claim that the inequality $\lambda - \nu \leq A(\mu + \nu)^\theta$ is preserved by the ODE, and therefore also the PDE. From this the theorem follows, since it implies

$$|\text{Ric} - Rg/3| \leq \lambda - \nu \leq A(\mu + \nu)^\theta \leq A(R/2)^\theta \leq \gamma R + C$$

for any fixed $\gamma > 0$ and for sufficiently large C.

To prove the claim, it suffices to show that the ratio $(\lambda - \nu)(\mu + \nu)^{-\theta}$ is decreasing as a function of time. We compute logarithmic derivatives

$$\log(\lambda - \nu)' = 2(\lambda - \mu + \nu) \text{ and } \log(\mu + \nu)' = 2\left(\lambda - \mu + \nu + 2\mu^2/(\mu + \nu)\right)$$

We know $\mu + \nu \geq \delta\lambda$ and therefore

$$\delta(\lambda - \mu + \nu) \leq \mu + \nu \leq \mu^2/(\mu + \nu)$$

so that $\log(\mu + \nu)' \geq (2 + \delta)(\lambda - \mu + \nu)$. From this the claim follows. □

This theorem strongly suggests that a 3-manifold M with $\mathrm{Ric} > 0$ shrinks to a round point in finite time, and in fact this is one of the main theorems proved by Hamilton in his first paper on Ricci flow [12]. We shall give a proof of this theorem in § 4.3 when we have discussed the process of taking limits, modulo a crucial point about injectivity radius that is established in § 5 using Perelman's so-called $\mathcal{W}$-functional.

3.6.4. *Strong maximum principle and product metrics.* A strong version of the maximum principle says that if a non-negative function f evolves by the heat equation, then if $f(p, t_0) > 0$ for some point p and time t_0, we have $f > 0$ for every point, and for all $t > t_0$. Contrapositively, this implies that if there is a positive time t_0 with $f(p, t_0) = 0$ then f is *identically* zero on $[0, t_0]$. For a non-negative function f on a compact conneced domain K evolving by the heat equation with Dirichlet boundary conditions $f|_K = 0$, one concludes only that $f > 0$ for every point in the interior.

Applying this to the evolution equation for R we conclude:

Lemma 3.12. *Let $M, g(t)$ be a 3-manifold (not necessarily compact or complete) satisfying Ricci flow. Suppose* $\mathrm{Rm} \geq 0$ *for $t = 0$ and suppose there is a point p and positive time t_0 so that $R(p, t_0) = 0$. Then M is flat on $t \in [0, t_0]$.*

The only subtlety is that M is not assumed to be compact — this is important for applications, where M might be a limit of parabolic rescalings near a singularity that we want to show is e.g. a neck pinch.

Proof. Suppose not, so that there is $t < t_0$ and a point q where $R(q, t) > 0$. Let N be a compact submanifold containing p and q, and let f be a non-negative function, positive at q with $R \geq f$ and with $f|_{\partial N} = 0$, and then evolve f so that it solves the heat equation on N with Dirichlet boundary conditions. Then $\partial_t(R - f) = \Delta(R - f) + 2|\mathrm{Ric}|^2$ so $(R - f) \geq 0$. But $f(t) > 0$ in the interior of N. □

Theorem 3.13 (Product metric)**.** *Let $M, g(t)$ be a 3-manifold (not necessarily compact or complete) satisfying Ricci flow. Suppose* $\mathrm{Rm} \geq 0$ *for $t = 0$ and M is not flat. Suppose further that for some p and t that* Ric *has a zero eigenvalue at (p, t). Then for $t > 0$ the metric splits (locally) as a product of a surface of positive curvature and a line.*

Proof. Since M is not flat, Lemma 3.12 implies the strict inequality $R > 0$ for $t > 0$. Let $s(x, t)$ be the function equal at each point to the sum of the two smallest eigenvalues of Rm. Thus $s \geq 0$, and by hypothesis $s(p, t) = 0$. The function s is convex and invariant under parallel translation; thus an application of the strong maximum principle implies that $s = 0$ identically. It follows that for all $t > 0$ at every point there is a unique vector V (up to sign) with $\|V\| = 1$ and $\mathrm{Ric}(V) = 0$. Work in a local coordinate patch where we can choose the sign of V consistently. We claim that V is parallel; this claim evidently proves the theorem.

Now, since $\mathrm{Rm} \geq 0$ we can only have $\mathrm{Ric}(V) = 0$ if the sectional curvature vanishes on every 2-plane containing V; i.e. if $\mathcal{R}(V, X, X, V) = 0$ for all X. We claim that in 3 dimensions, for any 4-tensor q with the symmetries of $\mathcal{R}$, if there is a nonzero vector V so that $q(V,X,X,V)=0$ for all X, then $q(V,\cdot,\cdot,\cdot)=0$. We prove this as follows. First of all, for any fixed X the symmetric quadratic form $q_X(Y):=q(Y,X,X,Y)$ vanishes on X and V, and therefore (because we are in 3 dimensions) $q_X(V,Y):=q(V,X,X,Y)=0$ for any Y. In other words, $q(V,\cdot,\cdot,\cdot)$ is antisymmetric in the first two indices, and therefore antisymmetric in all three indices — i.e. it is a 3-form. On the other hand, $q(V,X,Y,V)$ is symmetric in X and Y, and since it is also antisymmetric in these terms, it is zero. But any nonzero 3-form in 3 dimensions is of the form $X\wedge Y\wedge V$, so $q(V,\cdot,\cdot,\cdot)$ is identically zero. As a special case, we conclude that $\mathcal{R}(V,\cdot,\cdot,\cdot)=0$ (and V is evidently the unique vector with this property, up to scale).

Fix a point p and a curve γ through p and let $\tilde{V}$ be obtained from $V(p)$ by parallel transport along γ. Let X be any other parallel vector field along γ. Then $\mathcal{R}(\tilde{V},X,X,\tilde{V})\ge 0$ (because $\mathrm{Rm}\ge 0$) and vanishes at p, so that $\nabla_{\gamma'}(\mathcal{R}(\tilde{V},X,X,\tilde{V}))$ vanishes at p. Because $\tilde{V}$ and X are both parallel along γ,

$$\nabla_{\gamma'}(\mathcal{R}(\tilde{V},X,X,\tilde{V}))=(\nabla_{\gamma'}\mathcal{R})(\tilde{V},X,X,\tilde{V})$$

and since X is an arbitrary parallel field, it follows that $(\nabla_{\gamma'}\mathcal{R})(V,X,X,V)=0$ at the point p for any X. By the algebraic fact we proved above, it follows that $(\nabla_{\gamma'}\mathcal{R})(V,\cdot,\cdot,\cdot)=0$ at p, and since p is arbitrary, everywhere.

Now let X,Y,Z be any three parallel vector fields along γ. We compute

$$0=\nabla_{\gamma'}(\mathcal{R}(V,X,Y,Z))=\mathcal{R}(\nabla_{\gamma'}V,X,Y,Z)$$

Since the values of X,Y,Z at p are arbitrary, this implies that $\nabla_{\gamma'}V$ is in the kernel of $\mathcal{R}$, which is to say, it is proportional to V. Since $\|V\|=1$ it follows that $\nabla_{\gamma'}V=0$, and since γ' is arbitrary, V is parallel. From this the theorem follows. □

3.6.5. *Hamilton–Ivey curvature pinching.* Let $f(x):=x\log x-x$. This is convex and strictly increasing for $x\ge 1$, and we let $f^{-1}(y)$ (which is increasing and concave for $y\ge -1$) denote the inverse function. Note that $f^{-1}(y)/y\to 0$ as $y\to\infty$.

Theorem 3.14 (Hamilton–Ivey pinching)**.** *Let M be a compact 3-manifold and let $g(t)$ be a solution of Ricci flow which satisfies $R\ge -1$ and $\mathrm{Rm}+f^{-1}(R)\ge 0$ in the sense of operators, at $t=0$. Then these inequalities are both defined and satisfied for $t\ge 0$. In particular, for any initial metric there is a function $\phi(y)$ that goes to zero as $y\to\infty$ so that $\mathrm{Rm}\ge -\phi(R)R+C$.*

Proof. Since $f^{-1}\ge 1$, any metric can be rescaled to satisfy the hypothesis of the theorem. We have already seen that the spatial infimum of R can only increase, so $R\ge -1$ is preserved under Ricci flow. This means that $f^{-1}(R)$ is defined.

Writing the eigenvalues of Rm as $\lambda\ge\mu\ge\nu$, we are reduced to showing that for $R\ge -1$, the inequality $\nu+f^{-1}(\lambda+\mu+\nu)\ge 0$. The functions ν and f^{-1} are concave, and $\lambda+\mu+\nu$ is linear so the subset K where these inequalities are both satisfied is closed, convex and parallel. It suffices to show that for $R\ge -1$, this inequality is preserved fiberwise by the ODE.

If ν is non-negative then so is Rm, and the inequality is vacuous. So we assume $\nu < 0$ and we can rewrite the inequality as $\lambda + \mu \geq (-\nu)\log(-\nu)$. We compute

$$(\lambda + \mu + \nu\log(-\nu))' = 2\left(\lambda^2 + \mu\nu + \mu^2 + \lambda\nu + (\nu^2 + \lambda\mu)(1 + \log(-\nu)\right)$$

We need to show this derivative is non-negative on ∂K where $\nu + f^{-1}(\lambda + \mu + \nu) = 0$. Since $f^{-1} \geq 1$ we must have $\nu \leq -1$ on ∂K. Since $\lambda + \mu = (-\nu)\log(-\nu)$ we must have $\lambda + \mu \geq 0$ on ∂K. Thus we are reduced to showing that

$$\lambda^2 + \mu\nu + \mu^2 + \lambda\nu + (\nu^2 + \lambda\mu)\left(1 - \frac{\lambda + \mu}{\nu}\right) \geq 0$$

for $\lambda + \mu \geq 0$ and $\nu \leq -1$. This inequality is homogeneous in the eigenvalues, so we can rescale to $\nu = -1$, and we just need to show

$$\lambda^2 + \mu^2 + 1 + \lambda\mu(1 + \lambda + \mu) \geq 0$$

when $\lambda + \mu \geq 0$ and $\mu \geq -1$. This is straightforward.

The last claim follows if we take $\phi(x) := f^{-1}(x)/x$, where C depends on $|\text{Rm}|$ at time $t = 0$. □

This theorem is especially useful in the analysis of finite-time singularities. At a singularity $|\text{Rm}|$ must blow up. Since any lower bound on R is preserved under Ricci flow, we can't have $R \to -\infty$. Furthermore, the theorem implies that any upper bound on R puts a lower bound on Rm and therefore an upper bound on $|\text{Rm}|$; thus, the only way for a singularity to occur is for $R \to \infty$. But in this case, any negative eigenvalue of Rm is very small in comparison to R. In particular, if we do a parabolic rescaling near a finite-time singularity so that $R = 1$, we must have $\text{Rm} \geq -\epsilon$ for any positive ϵ. In particular, any geometric limit near a finite-time singularity has *non-negative sectional curvature*. This fact, in conjunction with Theorem 3.13, put extremely strong constraints on the geometry near a finite-time singularity.

3.6.6. *Hamilton's Harnack inequality.* A *Harnack inequality* controls the values at different points of a non-negative bounded harmonic function on a domain. Roughly speaking, it says that when the value of the function is small, it can't increase too quickly.

Hamilton [15], Thm. 1.1 obtained a tensor Harnack inequality for Ricci flow. He introduces tensors

$$P_{ijk} := \nabla_i R_{jk} - \nabla_j R_{ik} \text{ and } M_{ij} := \Delta R_{ij} + \frac{R_{ij}}{2t} - \frac{1}{2}\nabla_i\nabla_j R + \text{curvature term}$$

and defines an operator

$$Z(U, W) := M_{ij}W^iW^j + 2P_{ijk}U^{ij}W^k + \text{curvature term}$$

for a vector field W and a 2-vector field U. Hamilton shows that if $g(t)$ solves Ricci flow for $t > 0$ and has non-negative curvature operator $\text{Rm} \geq 0$, then for any U, W, we have $Z(U, W) \geq 0$ for all t. This is proved using the maximum principle. First of all, the condition $\text{Rm} \geq 0$ implies $\text{Ric} \geq 0$ so when t is small enough, $R_{ij}/2t$ dominates the other terms and shows $Z \geq 0$. Then for U, W at a point and time where $Z(U, W)$ becomes zero, Hamilton shows how to extend U and W locally in such a way that one obtains a formula for $(\partial_t - \Delta)Z$ which is ≥ 0 when $Z \geq 0$.

This inequality has some special cases which are extremely useful, and will reappear when we come to discuss Perelman's reduced length and reduced volume in § 5.4.

Using $\partial_t R_{ij} = \Delta R_{ij} +$ curvature term we can trade the ΔR_{ij} for a $\partial_t R_{ij}$ term, modulo curvature terms. If we fix vector fields X and Y we can take $W = Y$ and $U = X \wedge Y = \frac{1}{2}(X^i Y^j - X^j Y^i)$ and define $H(X,Y) = Z(X \wedge Y, Y)$. Then

$$H(X,Y) = -\mathrm{Hess}_R(Y,Y) - 2\langle R(Y,X)Y, X\rangle + 4(\nabla_X \mathrm{Ric}(Y,Y) \tag{3.3}$$

$$- \nabla_Y \mathrm{Ric}(Y,X)) + 2\partial_t \mathrm{Ric}(Y,Y) + 2|\mathrm{Ric}(Y,)|^2 + \frac{1}{t}\mathrm{Ric}(Y,Y) \tag{3.4}$$

Summing Y over an orthonormal basis gives

$$H(X) := \sum_i H(X, e_i) = \partial_t R + \frac{1}{t}R + 2\langle \nabla R, X\rangle + 2\mathrm{Ric}(X,X) \tag{3.5}$$

Theorem 3.15 (Hamilton trace Harnack inequality; [15] Cor. 1.2)**.** *Let $M, g(t)$ be a complete solution to Ricci flow with bounded curvature and non-negative curvature operator* $\mathrm{Rm} \geq 0$ *on some time interval* $0 < t < T$*. Then for any vector field X,*

$$H(X) := \partial_t R + R/t + 2\langle \nabla R, X\rangle + 2\mathrm{Ric}(X,X) \geq 0$$

One particularly important applications is to *ancient flows* — those defined on a time interval $(-\infty, T)$. In this case the R/t term goes away, and we get the inequality

$$\partial_t R + 2\langle \nabla R, X\rangle + 2\mathrm{Ric}(X,X) \geq 0$$

Applying this to $X = 0$ we see that for an ancient solution the scalar curvature R is *pointwise* non-decreasing!

3.6.7. *Time-dependent K.* It turns out that one can generalize Theorem 3.6 to the situation where the fiberwise convex sets K are time-dependent. Hamilton's argument only allows this under special circumstances, namely that K should be fiberwise convex in both space *and* time. However, Chow–Lu [8] showed that it's enough to consider K merely spacewise fiberwise convex. In other words, Theorem 3.6 remains true verbatim if we allow K to depend on time, and insist only that $K(t)$ is fiberwise convex and invariant under parallel transport for each fixed t. See [8] Thm 3 for details.

4. Singularities and Limits

We have seen that under Ricci flow singularities frequently must develop in finite time. It's crucial to understand the geometry near such a finite-time singularity. If we parabolically rescale near such a singularity, it will turn out that some subsequence of the rescaled flows converges (in a suitable sense) to a 'limit flow' whose geometry is especially easy to analyze. Making sense of this convergence, and proving structure theorems for the limit, depends on some fundamental comparison theorems in Riemannian geometry that we now discuss.

4.1. **Bishop–Gromov inequality.** In this section we prove the Bishop–Gromov inequality. This says the following. Suppose we are in an n-dimensional manifold M with $\mathrm{Ric} \geq (n-1)\kappa$; i.e. the Ricci curvature is at least as large as in a space of constant sectional curvature κ. Fix a point p. The inequality relates the volume of $\mathrm{vol}(B_r(p))$, the ball of radius r about p in M, and vol_r^κ, the volume of a ball of radius r in the space of constant curvature κ. It says that the ratio $\mathrm{vol}(B_r(p))/\mathrm{vol}_r^\kappa$ is *non-increasing* as a function of r.

In § 5.4 we shall see that monotonicity of the *reduced volume* $\tilde{V}$ for Ricci flow can be proved using a kind of parabolic substitute for length — Perelman's so-called *$\mathcal{L}$-length*, and we shall emphasize the similarity to the proof of the Bishop–Gromov theorem.

4.1.1. *Jacobi fields and the Index Form.* Let's recall that for each $p \in M$ the exponential map $\exp_p : T_pM \to M$ takes a vector v to $\gamma_v(1)$ where γ_v is the unique geodesic in M with $\gamma_v(0) = p$ and $\gamma_v'(0) = v$. Lines in T_pM map to geodesics γ in M, and linear vector fields along such lines map to *Jacobi fields* along γ, satisfying the *Jacobi equation*

$$\nabla_{\gamma'}\nabla_{\gamma'}X = R(X,\gamma')\gamma'$$

Now, fix a geodesic $\gamma : [a,b] \to M$. Let $\mathcal{V}$ denote the space of normal vector fields along γ and $\mathcal{V}_0$ the space of normal vector fields that vanish at both endpoints. The *index form*

$$\begin{aligned} I(V,W) :&= \int_a^b \langle \nabla_{\gamma'}V, \nabla_{\gamma'}W\rangle - \langle R(W,\gamma')\gamma', V\rangle dt \\ &= \langle \nabla_{\gamma'}V, W\rangle|_a^b - \int_a^b \langle \nabla_{\gamma'}\nabla_{\gamma'}V - R(\gamma',V)\gamma', W\rangle dt \end{aligned}$$

is a symmetric bilinear form on $\mathcal{V}$. Up to the cut locus, it is non-negative on $\mathcal{V}_0$, and vanishes exactly on the Jacobi fields. For a Jacobi field $V \in \mathcal{V}$ vanishing at $\gamma(a)$ we have $I(V,V) = \langle V', V\rangle_{\gamma(b)}$ and again, up to the cut locus, for any other W with $W(a) = V(a) = 0$ and $W(b) = V(b)$ we have $I(W,W) \geq I(V,V)$.

4.1.2. *Ricci curvature and spherical coordinates.* If we choose spherical coordinates r, Θ on T_pM then for each i the vector field $d\exp_p \partial_{\theta_i}$ is a Jacobi field along each radial geodesic $\Theta = \mathrm{const}$. For each i there is a 2-plane in T_pM spanned by $\partial r, \partial_{\theta_i}$ with sectional curvature K_i; the circle of radius r in this plane maps under $\exp_p$ to a curve of length $2\pi\,(r - K_ir^3/6 + O(r^4))$.

Let's write the volume form on M near p in radial coordinates as $\mu(r,\Theta)drd\Theta$. Thinking of trace as the derivative of determinant we get

$$\mu(r,\Theta) = r^{n-1}\left(1 - \mathrm{Ric}(\Theta,\Theta)r^2/6 + O(r^3)\right)$$

Here we are interpreting Θ both as a coordinate on the unit $(n-1)$-sphere, and as a unit-length vector in T_pM. Said in words, $\mathrm{Ric}(v)$ gives the leading order to which the volume of M grows slower than in Euclidean space along radial geodesics in the direction v.

We can relate this to the Index Form as follows. Let V_i be a collection of normal Jacobi fields along a radial geodesic in the direction Θ where $V_i(0) = 0$ and $V_i(r)$ are an orthonormal basis (such a family exists and is unique before we reach the cut locus). Then the logarithmic derivative of μ can be computed from the Index Form by the formula

$$\sum_i I(V_i, V_i) = \mathrm{tr}(V') = \frac{\mu'(r,\Theta)}{\mu(r,\Theta)}$$

4.1.3. *Bishop–Gromov inequality.* Our formula for μ shows that a lower bound for Ric (actually, R) gives an upper bound for the volume growth near p. The *Bishop-Gromov* inequality is a global version of this observation.

Theorem 4.1 (Bishop–Gromov)**.** *Suppose* $\mathrm{Ric} \geq (n-1)\kappa$ *for some constant* κ*. Let* vol_r^κ *denote the volume of the ball of radius* r *in the* n*-dimensional space of constant curvature* κ*. Then for an arbitrary point* p*, the function*

$$r \to \frac{\mathrm{vol}(B_r(p))}{\mathrm{vol}_r^\kappa}$$

is non-increasing as a function of r*, and tends to* 1 *as* $r \to 0$*.*

Proof. This is proved by integrating an inequality between the logarithmic derivative $\mu'(r,\Theta)/\mu(r,\Theta)$ for M, and the analogous quantity $\mu_\kappa'(r)/\mu_\kappa(r)$ in a space of constant curvature κ.

First let's assume that r is smaller than the distance to the cut locus of p in the direction Θ. As we saw in § 4.1.2 the logarithmic derivative of μ can be computed as

$$\frac{\mu'(r,\Theta)}{\mu(r,\Theta)} = \sum_{j=1}^{n-1} I(V_j, V_j)$$

where I is the index form, and the V_j are normal Jacobi fields along a radial geodesic γ with $V_j(0) = 0$ and $V_j(r)$ an orthonormal basis. Let $H_j(t) = (s_\kappa(t)/s_\kappa(r))e_j(t)$ be another collection of vector fields, where e_j is a parallel orthonormal frame, and

$$s_\kappa(t) = \begin{cases} \sin(\sqrt{\kappa}t)/\sqrt{\kappa} \text{ if } \kappa > 0 \\ t \text{ if } \kappa = 0 \\ \sinh(\sqrt{-\kappa}t)/\sqrt{-\kappa} \text{ if } \kappa < 0 \end{cases}$$

chosen so that $H_j(0) = V_j(0) = 0$ and $H_j(r) = V_j(r)$. Up to the cut locus the index form is positive definite in the space $\mathcal{V}_0$ of vector fields vanishing at the endpoints, so $I(V_j, V_j) \leq I(H_j, H_j)$.

Now let H_j^κ be normal Jacobi fields along a radial geodesic in a space of constant curvature κ, also defined by the formula $H_j^\kappa(t) = (s_\kappa(t)/s_\kappa(r))e_j^\kappa(t)$ where now e_j^κ is a parallel orthonormal frame in the constant curvature space. By a direct computation,

$$\sum I(H_j, H_j) = \sum I(H_j^\kappa, H_j^\kappa) + \int_0^r \left(\frac{s_\kappa(t)}{s_\kappa(r)}\right)^2 \big((n-1)\kappa - \mathrm{Ric}(\gamma', \gamma')\big)dt$$

so

$$\frac{\mu'(r,\Theta)}{\mu(r,\Theta)} = \sum I(V_j, V_j) \leq \sum I(H_j, H_j) \leq \sum I(H_j^\kappa, H_j^\kappa) = \frac{\mu_\kappa'(r)}{\mu_\kappa(r)}$$

This comparison is valid up to points in the cut locus of p. But when we are at or beyond the cut locus the inequality *still holds* — either because $\mu(r,\Theta)$ vanishes, or because the exponential map is no longer injective, so there is no further contribution to $\mathrm{vol}(B_r(p))$. $\square$

4.2. **Injectivity radius and volume.** The following theorem shows that in the presence of curvature bounds, a lower bound on volume implies a lower bound on injectivity radius, on every scale. This is crucial to obtain geometric limits under Ricci flow, since the derivative of volume is directly related to (scalar) curvature. This theorem is a restatement of Cheeger–Gromov–Taylor [6] Thm. 4.3 in the form most suited to applications. Sometimes in the literature this theorem is confused with Cheeger's Propellor Lemma, which is a slightly different statement about *global* bounds on the injectivity radius, and depends on working at a place where there is a smooth nontrivial closed geodesic.

Theorem 4.2 (Volume controls injectivity radius)**.** *For every dimension n and every $\epsilon > 0$ there is a $\delta(n, \epsilon) > 0$ so that if M is a complete Riemannian manifold of dimension n, and p is any point so that*

(1) $|\mathrm{Rm}| \le r^{-2}$ *on $B_r(p)$; and*

(2) $\mathrm{vol}(B_r(p)) \ge \epsilon r^n$,

then the injectivity radius at p is at least δr.

Proof. Note that the statement of the theorem is scale-invariant, so let's rescale so that $r = 1$, and denote $B_1(p)$ simply by B. Then $|\mathrm{Rm}| \le 1$ on B so the distance from p to its first conjugate point in any direction is at least π. Thus the theorem is proved if we can get a lower bound on the length of the shortest nontrivial geodesic γ from p to itself in terms of ϵ. Note that γ will almost certainly make a definite angle at p where it meets itself, unless p is quite special. Let's suppose the length of γ is ℓ.

Now, the geodesic γ determines a non-trivial element $[\gamma]$ in $\pi_1(B, p)$ (even if its image in $\pi_1(M, p)$ is trivial). We claim that the order of $[\gamma]$ is at least $N := \lfloor 2/\ell \rfloor$. To see this, let's lift to the universal cover $\tilde{B}$ of B. Let $\tilde{p}$ be the lift of p at the center of a ball $\hat{B} := B_1(\tilde{p})$ of radius 1, let $\tilde{\gamma}$ be a geodesic segment starting at $\tilde{p}$ and lifting γ, and let τ be the element of the deck group taking $\tilde{p}$ to the other end of $\tilde{\gamma}$. Write $\tilde{p}_0 := p$ and $\tilde{p}_k = \tau^k(\tilde{p})$ and suppose $N < \lfloor 2/\ell \rfloor$. Then the $\tilde{p}_k$ are all within distance 1 of $\tilde{p}_0$ which is to say they're in $\hat{B}$. Because $|\mathrm{Rm}| \le 1$ and $1 < \pi/2$ the ball $\hat{B}$ in $\tilde{B}$ is convex. Thus the convex function $x \to \sum_i d(x, \tilde{p}_i)^2$ has a unique minimum point $q \in \hat{B}$, and the point q is necessarily fixed by τ, contrary to the fact that τ is a nontrivial element of the deck group.

Using this lower bound on the order of $[\gamma]$ we can obtain an upper bound on $\mathrm{vol}(B_r(p))$. Let $U \subset B_{1/2}(p)$ be the intersection with the complement of the cut locus to p, so that U is star shaped, and every point in U is joined to p by a radial geodesic of length $\le 1/2$. Lift U to $\tilde{U} \subset \hat{B} \subset \tilde{B}$ containing $\tilde{p}$. Then the translates $\tau^k \tilde{U}$ for $-\lfloor 1/2\ell \rfloor \le k < \lfloor 1/2\ell \rfloor$ are all distinct and contained in $\hat{B}$ so

$$\begin{aligned} \mathrm{vol}(B_{1/2}(p)) = \mathrm{vol}(U) &\le \mathrm{vol}(\hat{B})/2\lfloor 1/2\ell \rfloor \\ &\le \mathrm{vol}_1^{-1}/2\lfloor 1/2\ell \rfloor = \mathrm{vol}_{1/2}^{-1} \frac{\mathrm{vol}_1^{-1}}{(\mathrm{vol}_{1/2}^{-1})(2\lfloor 1/2\ell \rfloor)} \end{aligned}$$

where remember vol_s^κ denotes the volume of a ball of radius s in the n-dimensional space of constant curvature κ, and the inequality $\mathrm{Rm} \ge -1$ plus Bishop–Gromov implies $\mathrm{vol}(\hat{B}) \le \mathrm{vol}_1^{-1}$. Finally, the monotonicity property of Bishop–Gromov gives $\mathrm{vol}(B_1(p)) \le C.\ell$ for some uniform constant C at least when $\ell < 1/2$. Thus a lower bound on volume gives a lower bound on ℓ and hence injectivity radius at p. ∎

4.3. **Geometric limits.** Let (M_i, g_i, p_i) be a sequence of pointed complete Riemannian manifolds of fixed dimension; i.e. for each i we have a point p_i in the manifold M_i with metric g_i. We say that this sequence *converges* in the sense of Cheeger–Gromov to (M, g, p) (of the same dimension) if there is an exhaustion of M by compact sets K_i containing p and a sequence of maps $\phi_i : K_i \to M_i$ taking p to p_i that are diffeomorphisms onto their image, and such that the pullback metrics $\phi_i^* g_i$ converge smoothly to g on compact subsets.

It is a fundamental fact that any sequence of pointed complete Riemannian manifolds of fixed dimension has a convergent subsequence providing the following two conditions are satisfied:

(1) Uniform control on derivatives of curvature on compact sets: for every radius r and every m there is a constant $C(r, m)$ so that $|\nabla^m \mathrm{Rm}| \le C(r, m)$ on the ball $B_r(p_i)$ of radius r about p_i in M_i; and
(2) Uniform lower bound on injectivity radius at the basepoint: there is a positive constant C so that the injectivity radius of M_i at p_i is at least C.

The only thing that is not clear is that control on the injectivity radius at p_i gives uniform control throughout the ball $B_r(p_i)$ for any fixed r. But, because we have a lower bound on Rm on $B_r(p_i)$ for any r, if there were points $q_i \in B_r(p_i)$ where $\mathrm{inj}(q_i) \to 0$ then by Bishop–Gromov we would have $\mathrm{vol}(B_r(p_i)) \to 0$. Because we have an upper bound on Rm this would imply also $\mathrm{inj}(p_i) \to 0$, contrary to hypothesis. Thus: for any fixed radius r, there is a uniform positive lower bound on injectivity radius everywhere in $B_r(p_i)$. From this the existence of a convergent subsequence easily follows.

More generally, suppose we have a family of pointed Riemannian flows $(M_i, g_i(t), p_i)$ all defined on a common time interval $a < t < b$. We say this sequence *converges* in the sense of Cheeger–Gromov–Hamilton to a limit $(M, g(t), p)$ if there are K_i and ϕ_i as above so that $\phi_i^* g_i(t)$ converges smoothly to $g(t)$ on compact subsets of M and of (a, b).

Proposition 4.3 (Limit exists)**.** *Let $(M_i, g_i(t), p_i)$ be a sequence of complete pointed Ricci flows defined for t on a common time interval $0 \in (a, b)$. Suppose*

(1) there is a constant C so that $|\mathrm{Rm}| \le C$ for all i for every point in M_i and all $t \in (a, b)$; and
(2) there is a constant C so that for all i the injectivity radius at p_i is at least C at time 0.

Then some subsequence converges smoothly to a limit Ricci flow $(M, g(t), p)$ defined on the same time interval.

We give the sketch of a proof.

Proof. Control over $|\mathrm{Rm}|$ throughout (a, b) gives control over $|\partial_t g(t)|$, so control of injectivity radius at $t = 0$ gives control at any other time.

For compact manifolds, the derivative estimates in Theorem 3.5 give control on $|\nabla^m \mathrm{Rm}|$ in terms of $|\mathrm{Rm}|$; for noncompact manifolds we must use the more general local estimates due to Shi. □

4.4. **κ-solutions.** Now let's suppose $M, g(t)$ satisfies Ricci flow on some maximum time interval $t \in [0, T)$ and that the curvature blows up as $t \to T$. Take a sequence of parabolic rescalings of the flow at points and times approaching the singularity.

We shall see that this sequence of rescaled flows has a convergent subsequence, and that the limit enjoys a host of desirable geometric properties which are summarized by saying that it is a *κ-solution* for some κ depending only on T and the metric $g(0)$.

First we give the definition of κ-noncollapsed, in the sense of Perelman [26], Def. 4.2. (Perelman gives another definition [26] Def. 8.1 which is weaker, but better suited to the analysis of Ricci flow with surgery).

Definition 4.4 (κ-noncollapsed)**.** We say that a metric is *κ-noncollapsed at scales $\le \rho$* for some $\kappa > 0$ if the following is true. Suppose there is $r \le \rho$ and a point p so that $|\mathrm{Rm}(q)| \le r^{-2}$ for all $q \in B_r(p)$. Then $\mathrm{vol}(B_r(p)) \ge \kappa r^n$. A metric is *$\kappa$-noncollapsed* if it is κ-noncollapsed at every scale.

If g is κ-noncollapsed at scales $\le \rho$ then λg is κ-noncollapsed at scales $\le \sqrt{\lambda}\rho$. In particular, the property of being κ-noncollapsed is scale-invariant.

Now we define a κ-solution.

Definition 4.5 (κ-solution)**.** Ricci flow $N, h(t)$ is a *κ-solution* for some κ if it satisfies the following properties:

(1) N is connected, and the metric $h(t)$ is complete for all t;
(2) the flow is *ancient* — i.e. defined on $(-\infty, t_0)$ for some positive $t_0 > 0$;
(3) the curvature norms $|\mathrm{Rm}|$ are bounded on each time slice;
(4) the metric on each time slice is κ-noncollapsed.
(5) the curvature is non-negative $\mathrm{Rm} \ge 0$;
(6) the scalar curvature R is strictly positive everywhere;

We shall discuss κ-noncollapsing in § 5, culminating in Perelman's proof of the 'No local collapsing' Theorem 5.10, using the $\mathcal{W}$-functional. This says that if $M, g(t)$ is Ricci flow on a compact manifold M defined for time in an interval $[0, T)$ then there is a κ depending on $g(0)$ and on T so that each metric $g(t)$ is κ-noncollapsed at scales $\le \sqrt{T}$.

Modulo this result, we are now in a position to analyze blow-up limits of finite time singularities:

Theorem 4.6 (Blow-up)**.** *Let $M, g(t)$ be a compact 3-manifold satisfying Ricci flow on some finite maximum time interval $t \in [0, T)$. Choose points p_i and times $t_i \to T$ so that $\lambda_i := |\mathrm{Rm}|(p_i, t_i) \to \infty$, and $\lambda_i \ge |\mathrm{Rm}|(q, s)$ for all $q \in M$ and $s \le t_i$. Then there is $\kappa > 0$ depending on T and $g(0)$ so that the parabolically rescaled flows $g_i(t) := \lambda_i g(t_i + t/\lambda_i)$ have a subsequence converging to a κ-solution.*

Furthermore this limit solution has the additional property that $|\mathrm{Rm}| \le 1$ for $t \in (-\infty, 0]$.

Proof. Since each $g_i(t)$ has $|\mathrm{Rm}| \le 1$ on $t \le 0$, none of the $g_i(t)$ can become singular too fast by Proposition 3.4. Thus there is a uniform positive constant $t_0 > 0$ so that each $g_i(t)$ is defined on the time interval $(-t_i\lambda_i, t_0)$, and there are uniform bounds on $|\mathrm{Rm}|$ for every $g_i(t)$ on every compact subset of $(-\infty, t_0)$, where defined.

Theorem 5.10 says that $g(t)$ is κ-noncollapsed at scales $\le \sqrt{T}$ for all t. Therefore $g_i(t)$ is κ-noncollapsed at scales $\le \sqrt{\lambda_i T}$. In particular, Theorem 4.2 implies that there is a uniform lower bound on the injectivity radius of $g_i(0)$ at p_i. Thus we can apply Proposition 4.3 and deduce the existence of a limit $(N, h(t), p)$ for some subsequence of the $(M, g_i(t), p_i)$.

Since the $(M, g_i(t), p_i)$ are complete and connected, so is $(N, h(t), p)$. By what we have already argued, this limit is ancient, $|\mathrm{Rm}|$ is bounded on each time slice, and is κ-noncollapsed on every scale.

Hamilton–Ivey pinching (Theorem 3.14) implies that as $|\mathrm{Rm}| \to \infty$ the ratio of the minimum to the maximum eigenvalue of Rm must go to zero. It follows that for every $\epsilon > 0$, we must have $\mathrm{Rm} \geq -\epsilon$ for all $t \leq 0$ for sufficiently large i. Thus for any limit we will have $\mathrm{Rm} \geq 0$ for $t \leq 0$ and therefore $\mathrm{Rm} \geq 0$ for all t. Since $|\mathrm{Rm}|(p_i, 0) = 1$ the same is true of any limit; thus the limit is not flat, and has $R > 0$ everywhere. □

Corollary 4.7 (Hamilton's Uniformization Theorem). *A compact 3-manifold M with* $\mathrm{Ric} > 0$ *shrinks to a round spherical space form in finite time under Ricci flow.*

Proof. Let $M, g(t)$ evolve the initial metric by Ricci flow. Since M is compact and $\mathrm{Ric} > 0$ there must be a finite time singularity. Again by compactness, $0 < \alpha \leq \mathrm{Ric} \leq \beta$ holds throughout M for some α, β so we can apply Theorem 3.11. In other words, for any positive γ there is a constant C so that under Ricci flow, $|\mathrm{Ric} - Rg/3| \leq \gamma R + C$ holds throughout M. Let $g_i(t)$ be parabolically rescaled metrics centered near the singularity where the curvature blows up. For the $g_i(t)$ this inequality becomes $|\mathrm{Ric} - Rg/3| \leq \gamma R + C/\lambda_i$ so the blow-up limit $N, h(t)$ satisfies $|\mathrm{Ric} - Rg/3| \leq \gamma R$ for all $\gamma > 0$. In other words, N is Einstein, and therefore has constant sectional curvature because M is 3-dimensional. Since $h(t)$ has $\mathrm{Rm} \geq 0$ but is not flat, it has constant positive sectional curvature; i.e. it is a spherical space form. In particular, N is compact, and the metrics $g(t)$ on M converge under rescaling in C^∞ to a round metric. □

4.5. **Nonnegative curvature and the Soul Theorem.** Riemannian manifolds with $K \geq 0$ are very special. Non-negative sectional curvature strongly resists noncompactness. The following theorem is fundamental:

Theorem 4.8 (Cheeger–Gromoll Soul Theorem [5]). *Let M be a complete connected Riemannian manifold with sectional curvatures $K \geq 0$. Then then there is a compact totally convex, totally geodesic submanifold S (the 'soul') such that M is diffeomorphic to the normal bundle of S.*

We give the idea of the proof.

Proof. A *ray* is an isometrically embedded copy of $\mathbb{R}^+$. If M is complete and connected but noncompact, then for any point p it contains at least one ray based at p (just take a limit of distance-minimizing geodesics from p to points further and further away). For every ray γ based at p consider the *Busemann function* $b_\gamma(q) := \lim_{t\to\infty} d(\gamma(t), q) - t$. The level sets of b_γ are limits of metric spheres based at points on γ exiting the end. In Euclidean space, these level sets would be flat hyperplanes; in a manifold of non-negative curvature, Busemann functions are concave, so the superlevel sets $b_\gamma \geq C$ are closed and convex. Let A be the subset of M for which $b_\gamma \geq 0$ for every Busemann function associated to a ray based at p. Since $b_\gamma(p) = 0$ for all γ, we have $p \in A$. Then A is closed and convex (because it's an intersection of closed convex sets). We claim that A is compact. For if not, because A is convex, we could find a ray γ based at p and contained in A, and then b_γ would be strictly negative on γ.

Let S^1 be the compact subset of A where $\min_\gamma b_\gamma$ achieves its maximum. One can show that S^1 has no interior, and is therefore a compact submanifold of M of codimension at least 1. If ∂S^1 is nonempty we define S^2 to be the subset of S^1 at maximal distance from ∂S^1. Evidently S^2 is compact, with dimension strictly less than that of S^1. By induction, we form a finite descending chain $S^1 \supset S^2 \supset \cdots \supset S^k$. Then $S := S^k$ is convex and compact without boundary, and of dimension strictly less than that of M.

It turns out that S is a soul. To see this, observe that if there are two minimal geodesics from S to any $q \in M$, these geodesics make an angle of less than $\pi/2$ at q or else they would not be minimal. Thus (e.g. by using a partition of unity) we may form a nonsingular gradient-like vector field transverse to the foliation of $M - S$ by level sets of the distance function to S, and witnessing that this foliation is nonsingular. Flowing along this vector field carries M into S and exhibits a diffeomorphism from M to the normal bundle of S. □

Let's apply this to the underlying manifold M of a κ-solution in the case of a 3-manifold. Since $\text{Rm} \ge 0$ the Soul Theorem applies. If M is compact, the theorem tells us nothing. Otherwise there are three possibilities, depending on the dimension of the soul: S could be a point (in which case M is diffeomorphic to $\mathbb{R}^3$), or S could be a circle or a compact surface.

To analyze these cases we use the so-called *Splitting Theorem*:

Theorem 4.9 (Toponogov Splitting Theorem)**.** *Let M be a complete connected Riemannian manifold with sectional curvatures $K \ge 0$ and suppose M contains a* line *— i.e. an isometrically embedded copy of $\mathbb{R}$. Then M splits isometrically as a product $M = N \times \mathbb{R}$.*

If S is a circle, then M is diffeomorphic to $S^1 \times \mathbb{R}^2$. Any soul S unwraps to a soul in finite covers, and therefore in the universal cover unwraps to a line. Thus the universal cover is isometric to a product $\Sigma \times \mathbb{R}$, where Σ is diffeomorphic to $\mathbb{R}^2$ and has non-negative curvature. In particular, the ends of Σ and hence of M are asymptotically flat; but then M is not non-collapsed. So this case can't occur as a κ-solution.

If S is a surface, then after passing to a double cover if necessary, it is 2-sided and M is diffeomorphic to a product $S \times \mathbb{R}$. This has two ends, and a sequence of distance-minimizing geodesics with endpoints exiting either end has a subsequence converging on compact subsets to a line. Thus M is isometric to a product $S \times \mathbb{R}$ up to taking double covers. Since M has $\text{Rm} \ge 0$ but is not flat, S is a sphere.

We conclude that a noncompact κ-solution is either a shrinking round cylinder up to finite covers, or is diffeomorphic to $\mathbb{R}^3$. It will turn out that compact κ-solutions are diffeomorphic to spherical space-forms. The proof of this in general is rather indirect, and will be sketched in Theorem 5.20. However for compact κ-solutions that arise as the limit of a parabolic blow-up, the proof is much simpler and is given in § 4.6.

4.6. **Compact blow-up solutions.** Let $M, g(t)$ be Ricci flow on a compact 3-manifold becoming singular at time T, and let $N, h(t)$ be a limit flow obtained by blow up as in Theorem 4.6. Suppose N is compact. Then N is a shrinking spherical space form (and is diffeomorphic to M).

Proposition 4.10 (Compact limit is a spherical space-form)**.** *A compact blow-up solution is a shrinking spherical space form.*

Proof. Suppose $\mathrm{Ric} > 0$ everywhere in N. Because N is compact and $g_i \to h$ it follows that $\mathrm{Ric} > 0$ for some $g(t)$. Thus Corollary 4.7 applies and $N, h(0)$ is a spherical space-form.

Otherwise Ric has a zero eigenvalue somewhere and the metric splits locally as a product by Theorem 3.13 so that N (being compact) is finitely covered by $S^2 \times S^1$ for some metric of positive curvature on the S^2 factor. Ricci flow splits as Ricci flow on the S^2 factor and the identity on the S^1 factor. This means that at some very negative time there was a point in one of the S^2 factors where the curvature was very close to zero, and therefore near that point $|\mathrm{Rm}| \le \epsilon$. But if we rescaled near that point to have $|\mathrm{Rm}| = 1$ the S^1 factor (whose length is time-independent) would get arbitrarily short, and the normalized volume would go to zero, contrary to the fact that every $h(t)$ is κ-noncollapsed. So this case can't happen for N compact. □

5. Perelman's monotone functionals

5.1. **The $\mathcal{F}$-functional.** The *total scalar curvature* of a compact Riemannian manifold (M, g) is a functional of g, defined as

$$\mathcal{S}(g) := \int_M R \, d\mathrm{vol}_g$$

In the sequel we suppress the subscript and just write $d$vol. Let's compute the first variation of $\mathcal{S}$. If $v_{ij} := v_{ij} dx^i dx^j$ is a symmetric 2-form, the derivative of $\mathcal{S}$ in the direction of v_{ij} is

$$\delta\mathcal{S} := d\mathcal{S}(v_{ij}) = \int_M (\delta R) d\mathrm{vol} + R(\delta \, d\mathrm{vol})$$

Now, $\delta \, d\mathrm{vol} = v/2 \, d\mathrm{vol}$ where $v = g^{ij} v_{ij}$ is the trace of v_{ij}. Using equation 3.2 gives

$$\begin{aligned}
\delta R = \delta(g^{ij} R_{ij}) &= \delta g^{ij} R_{ij} + g^{ij} \delta R_{ij} \\
&= -g^{ik} g^{jl} v_{kl} R_{ij} + \frac{1}{2} g^{ij} g^{pq} (\nabla_q \nabla_i v_{jp} + \nabla_q \nabla_j v_{ip} - \nabla_q \nabla_p v_{ij} - \nabla_i \nabla_j v_{pq}) \\
&= -g^{ij} g^{pq} (\nabla_i \nabla_j v_{pq} - \nabla_i \nabla_p v_{jq} + v_{ip} R_{jq})
\end{aligned}$$

so that

$$\delta R = -\Delta v + \nabla_i \nabla_j v^{ij} - v^{ij} R_{ij}$$

Now, $-\Delta v$ and $\nabla_i \nabla_j v^{ij}$ are total divergences, and integrate to zero. Thus

$$\delta\mathcal{S} = \int_M \left(-v^{ij} R_{ij} + \frac{v}{2} R \right) d\mathrm{vol}$$

Thus the gradient flow of $\mathcal{S}$ is tantalizingly close to Ricci flow (up to a constant), except for the annoying $Rv/2$ term coming from the variation of $d$vol. One can try to fix this by introducing an auxiliary smooth function f and integrating with respect to a new volume form $dm := e^{-f} d\mathrm{vol}$, and consider simultaneous variations of g and of f which keep this volume form fixed. Writing $\delta f = h$ we get

$$\delta e^{-f} d\mathrm{vol} = \left(\frac{v}{2} - h \right) e^{-f} d\mathrm{vol}$$

so setting $h = v/2$ eliminates the $Rv/2$ term. Of course, the terms $-\Delta v$ and $\nabla_i \nabla_j v^{ij}$ are no longer total divergences with respect to the measure dm, and we must introduce further terms in the integrand to deal with these.

Perhaps with this motivation, Perelman introduces the $\mathcal{F}$-functional in [26], § 1.1:

Definition 5.1 ($\mathcal{F}$-functional)**.** Let M be a smooth manifold. For a metric g and a smooth function f define the functional

$$\mathcal{F}(g,f) := \int_M (R + |\nabla f|^2)e^{-f}d\text{vol}$$

We compute the first variation of this functional.

Proposition 5.2 (First variation of $\mathcal{F}$)**.** *Let $v = v_{ij}dx_i dx_j$ be a symmetric 2-form, and h a smooth function. Also write $v = g^{ij}v_{ij}$, the trace of v. Then*

$$\delta\mathcal{F} := d\mathcal{F}(v_{ij},h) = \int_M \left(-v^{ij}(R_{ij} + \nabla_i\nabla_j f) + \left(\frac{v}{2} - h\right)(2\Delta f - |\nabla f|^2 + R)\right)e^{-f}d\text{vol}$$

Proof. We've already computed the first order variation of R and $e^{-f}d\text{vol}$, namely

$$\delta R = -\Delta v + \nabla_i\nabla_j v^{ij} - v^{ij}R_{ij} \text{ and } \delta(e^{-f}d\text{vol}) = \left(\frac{v}{2} - h\right)e^{-f}d\text{vol}$$

Likewise,

$$\delta|\nabla f|^2 = \delta(g^{ij}\nabla_i f\nabla_j f) = -v^{ij}\nabla_i f\nabla_j f + 2g^{ij}\nabla_i f\nabla_j h$$

Putting this together gives

$$\begin{aligned}\delta\mathcal{F} = \int_M \Big(&-\Delta v + \nabla_i\nabla_j v^{ij} - v^{ij}R_{ij} - v^{ij}\nabla_i f\nabla_j f + 2\langle\nabla f,\nabla h\rangle \\ &+ (R + |\nabla f|^2)\left(\frac{v}{2} - h\right)\Big)e^{-f}d\text{vol}\end{aligned}$$

Now, $\Delta e^{-f} = (|\nabla f|^2 - \Delta f)e^{-f}$ and so

$$\int_M (-\Delta v)e^{-f}d\text{vol} = \int_M -v(\Delta e^{-f})d\text{vol} = \int_M v(\Delta f - |\nabla f|^2)e^{-f}d\text{vol}$$

Integrating by parts twice gives

$$\int_M (\nabla_i\nabla_j v^{ij})e^{-f}d\text{vol} = \int_M v^{ij}(\nabla_i\nabla_j e^{-f})d\text{vol} = \int_M v^{ij}(\nabla_i f\nabla_j f - \nabla_i\nabla_j f)e^{-f}d\text{vol}$$

and similarly,

$$\begin{aligned}\int_M 2\langle\nabla f,\nabla h\rangle e^{-f}d\text{vol} &= \int_M \langle -2\nabla e^{-f},\nabla h\rangle d\text{vol} = \int_M h(2\Delta e^{-f})d\text{vol} \\ &= \int_M h(2|\nabla f|^2 - 2\Delta f)e^{-f}d\text{vol}\end{aligned}$$

so we get

$$\delta\mathcal{F} = \int_M \left(\left(\frac{v}{2} - h\right)(2\Delta f - 2|\nabla f|^2) - v^{ij}(R_{ij} + \nabla_i\nabla_j f) + \left(\frac{v}{2} - h\right)(R + |\nabla f|^2)\right)e^{-f}d\text{vol}$$

as claimed. ∎

Recalling our initial discussion, let's define a measure $dm := e^{-f}d\text{vol}$ and notice that $(v/2-h)=0$ exactly for variations that keep dm fixed. Another way to say this is that for a fixed measure dm, a metric g determines a function f by $f=\log(d\text{vol}/dm)$, and there is a functional $\mathcal{F}^m$ on smooth metrics, defined by $\mathcal{F}^m(g) := \mathcal{F}(g,f)$.

If we define a "metric" on the space of Riemannian metrics by the inner product

$$\langle v_{ij}, v_{ij}\rangle_g := \frac{1}{2}\int_M v^{ij}v_{ij}dm$$

then with this normalization, the gradient flow of the functional $\mathcal{F}^m$ is given by the equations

$$\partial_t g_{ij} = -2(\text{Ric}+\text{Hess}f)$$

and

$$\partial_t f = \partial_t \log\left(\frac{d\text{vol}}{dm}\right) = \frac{1}{2}\,\text{tr}\,\partial_t g_{ij} = -R-\Delta f$$

The evolution of g is Ricci flow composed with the (time-dependent) gradient vector flow by $-\text{grad}f$. On the other hand, the equation $\partial_t f = -R-\Delta f$ is (up to the scalar term R) a *backward* heat equation, and is unlikely to admit a solution for a typical initial measure dm. However, we can solve for Ricci flow of an initial metric $g(t_1)$ on some time interval $[t_1,t_2]$, specify a final value of $f(t_2)$, and solve the heat equation for f in *backward* time from t_2 to t_1 to determine an initial value of f, and thus a measure dm and functional $\mathcal{F}^m$.

Explicitly, after pulling back the metric by the diffeomorphism flow generated by $\text{grad}f$, the equations decouple to

$$\partial_t g_{ij} = -2\text{Ric}, \quad \partial_t f = -R+|\nabla f|^2-\Delta f$$

Writing $u := e^{-f}$ the latter equation becomes the linear equation $\partial_t u = -\Delta u + Ru$, which can be solved in backward time.

5.2. **The $\mathcal{W}$-functional.** To analyze Ricci flow near a developing singularity, it's important to have a scale-invariant version of the $\mathcal{F}$-functional. This is the $\mathcal{W}$-functional, introduced by Perelman in [26], § 3.1:

Definition 5.3 ($\mathcal{W}$-functional)**.** Let M be a smooth manifold. For a metric g, a smooth function f and a scale parameter τ define the functional

$$\mathcal{W}(g,f,\tau) := \int_M \big(\tau(R+|\nabla f|^2)+f-n\big)\,(4\pi\tau)^{-n/2}e^{-f}d\text{vol}$$

for f and τ satisfying $\int_M (4\pi\tau)^{-n/2}e^{-f}d\text{vol} = 1$ and $\tau>0$.

The $\mathcal{W}$-functional is invariant under diffeomorphism $\mathcal{W}(g,f,\tau) = \mathcal{W}(\phi^*g,\phi^*f,\tau)$ and parabolic rescaling $\mathcal{W}(g,f,\tau)=\mathcal{W}(\lambda g,f,\lambda\tau)$, and is therefore constant on gradient shrinking solitons, taking $t=-\tau$ for $t\in(-\infty,0)$.

As before we can compute

Proposition 5.4 (First variation of $\mathcal{W}$)**.** *With v_{ij} and h as in Proposition 5.2 and $\sigma = \delta\tau$ we have*

$$\delta\mathcal{W} := d\mathcal{W}(v_{ij}, h, \sigma) = \int_M \Big(\sigma(R + |\nabla f|^2) - \tau v^{ij}(R_{ij} + \nabla_i\nabla_j f) + h + \left(\frac{v}{2} - h - \frac{n\sigma}{2\tau}\right)\left(\tau(2\Delta f - |\nabla f|^2 + R) + f - n\right)\Big)(4\pi\tau)^{-n/2}e^{-f}\mathrm{dvol}$$

Proof. We have

$$\delta\left((4\pi\tau)^{-n/2}e^{-f}\mathrm{dvol}\right) = \left(\frac{v}{2} - h - \frac{n\sigma}{2\tau}\right)(4\pi\tau)^{-n/2}e^{-f}\mathrm{dvol}$$

and computing the other terms as in Proposition 5.2 gives the result. □

As before, this motivates fixing a smooth measure dm on M with mass 1, and having f and τ depend on g by forcing

$$(4\pi\tau)^{-n/2}e^{-f}\mathrm{dvol} = dm$$

This will make $v/2 - h - n\sigma/2\tau = 0$, so that $\delta\mathcal{W}$ reduces to

$$\delta\mathcal{W} = \int_M \left(\sigma(R + |\nabla f|^2) - \tau v^{ij}(R_{ij} + \nabla_i\nabla_j f) + h\right)(4\pi\tau)^{-n/2}e^{-f}\mathrm{dvol}$$

Thus if we consider a family (g, τ) evolving by

$$\partial_t g_{ij} = -2(\mathrm{Ric} + \mathrm{Hess}(f)) \text{ and } \partial_t\tau = -1$$

then f evolves by

$$\partial_t f = -\Delta f - R + \frac{n}{2\tau}$$

Again, by pulling back the metric under the diffeomorphism flow generated by $\mathrm{grad} f$ the evolution equations decouple to

$$\partial_t g = -2\mathrm{Ric}, \quad \partial_t\tau = -1, \quad \partial_t f = -\Delta f + |\nabla f|^2 - R + \frac{n}{2\tau} \tag{5.1}$$

In any case we can compute

$$\frac{d\mathcal{W}}{dt} = \int_M \left(-(R + |\nabla f|^2) + 2\tau|R_{ij} + \nabla_i\nabla_j f|^2 - \Delta f - R + \frac{n}{2\tau}\right)(4\pi\tau)^{-n/2}e^{-f}\mathrm{dvol}$$

Recall that $\Delta e^{-f} = (|\nabla f|^2 - \Delta f)e^{-f}$ so that $\int_M e^{-f}|\nabla f|^2\mathrm{dvol} = \int_M e^{-f}\Delta f\mathrm{dvol}$. Hence collecting terms gives

$$\begin{aligned}\frac{d\mathcal{W}}{dt} &= \int_M \left(-2(R + |\nabla f|^2) + 2\tau|R_{ij} + \nabla_i\nabla_j f|^2 + \frac{n}{2\tau}\right)(4\pi\tau)^{-n/2}e^{-f}\mathrm{dvol} \\ &= \int_M 2\tau|R_{ij} + \nabla_i\nabla_j f - \frac{1}{2\tau}g_{ij}|^2(4\pi\tau)^{-n/2}e^{-f}\mathrm{dvol}\end{aligned}$$

which is non-negative, and vanishes identically for a shrinking gradient soliton

$$\mathrm{Ric} + \mathrm{Hess}(f) = \frac{1}{2\tau}g$$

in which case τ can be interpreted as the time remaining until extinction.

5.3. **Monotone quantities and noncollapsing.**

Definition 5.5 (λ-functional)**.** For a metric g, let $\lambda := \lambda(g)$ be equal to the infimum of $\mathcal{F}(g, f)$ over all functions f with $\int_M e^{-f} d\text{vol} = 1$.

If we write $\Phi := e^{-f/2}$ then

$$\mathcal{F} = \int_M (4|\nabla\Phi|^2 + R\Phi^2) d\text{vol} = \int_M \Phi(-4\Delta\Phi + R\Phi) d\text{vol}$$

So the infimum is λ, the smallest eigenvalue of $-4\Delta + R$, and the infimum is *achieved* by an eigenvector. As is well-known from elliptic theory, the smallest eigenspace is 1-dimensional, and an eigenfunction does not change sign. So there is a unique eigenfunction $e^{-\bar{f}/2}$ for some smooth function $\bar{f}$ satisfying $\int_M e^{-\bar{f}} d\text{vol} = 1$.

Lemma 5.6 (Monotonicity of λ)**.** *If $g(t)$ evolves by Ricci flow (up to diffeomorphism) then $\lambda(g(t))$ is nondecreasing in t.*

Proof. For some time interval $[t_1, t_2]$ let $\bar{f}(t_2)$ be the minimizer for $g(t_2)$, and put $u(t_2) = e^{-\bar{f}(t_2)}$. Extend u over the domain $[t_1, t_2]$ solving the backwards heat equation $\partial_t u = -\Delta u + Ru$.

Now, if h is any solution to the *forward* heat equation $\partial_t h = \Delta h$ on any time interval $(t', t_2]$ then, since $\partial_t d\text{vol} = -R d\text{vol}$,

$$\begin{aligned} \frac{d}{dt} \int_M u(t)h(t) d\text{vol} &= \int_M (\partial_t u)h + u(\partial_t h) - uhR \, d\text{vol} \\ &= \int_M \big((\partial_t u + \Delta u - Ru)h + u(\partial_t h - \Delta h)\big) d\text{vol} = 0 \end{aligned}$$

Note that taking h constant implies that $\int_M u(t) d\text{vol}$ is independent of t.

If we take h so that the limit of h as $t \to t'$ is a delta function supported at a point x', then $h(t) > 0$ for all $t > t'$, so that

$$u(x', t') = \lim_{t \to t'} \int_M u(t)h(t) d\text{vol} = \int_M u(t_2)h(t_2) d\text{vol} > 0$$

Thus we can extend f by $u(t) = e^{-f(t)}$ and observe that f solves $\partial_t f = -\Delta f + |\nabla f|^2 - R$, so that

$$\lambda(t_1) \le \mathcal{F}(g(t_1), f(t_1)) \le \mathcal{F}(g(t_2), f(t_2)) = \lambda(t_2)$$

□

The analog of λ for the $\mathcal{W}$-functional is the μ-functional:

Definition 5.7 (μ-functional)**.** For a metric g and $\tau > 0$ let $\mu := \mu(g, \tau)$ be equal to the infimum of $\mathcal{W}(g, f, \tau)$ over all functions f with $\int_M (4\pi\tau)^{-n/2} e^{-f} d\text{vol} = 1$.

If we write $\Phi = e^{-f/2}$ as before, then $\mu(g, \tau)$ and Φ solve the equation

$$\tau(-4\Delta + R)\Phi = 2\Phi \log \Phi + (\mu(g, \tau) + n)\Phi$$

Again, such a normalized Φ is smooth and positive, and $f = -2\log\Phi$ is also smooth.

Lemma 5.8 (Monotonicity of μ)**.** *If $g(t)$ evolves by Ricci flow (up to diffeomorphism) then $\mu(g(t), t_0 - t)$ is nondecreasing in t.*

Proof. As before if we define $u(t_2) = (4\pi\tau)^{-n/2}e^{-\bar{f}(t_2)}$ and extend u to $[t_1, t_2]$ solving $\partial_t u = -\Delta u + Ru$. The same argument as in the proof of Lemma 5.6 shows that u is strictly positive, so we can extend $\bar{f}$ to f by $u(t) = (4\pi\tau)^{-n/2}e^{-f(t)}$ and then observe that $\int_M u(t)d\text{vol}$ is independent of t, and f solves $\partial_t f = -\Delta f + |\nabla f|^2 - R + n/(2\tau)$.

The remainder of the argument is the same as in the proof of Lemma 5.6. □

5.3.1. *No local collapsing.* A key application of the monotonicity of the μ functional is to give lower bounds on volumes of rescaled balls when curvature blows up in finite time.

Definition 5.9 (Local collapsing)**.** A family of metrics $g(t)$ evolving by Ricci flow on a time interval $[0, T)$ is said to be *locally collapsing* at T if there are times $t_k \to T$, points p_k and radii r_k so that if B_k denotes the ball of radius r_k centered at p_k in the metric at time t_k, then

(1) r_k^2/t_k is bounded;
(2) $|\text{Rm}| \le r_k^{-2}$ on B_k; and
(3) $\lim_{k\to\infty} \text{vol}(B_k)r_k^{-n} = 0$.

According to Theorem 4.2 this is equivalent to saying that if we rescale the balls B_k to have radius 1, then (because the curvature is bounded) the injectivity radius goes to zero. Thus, local collapsing is the condition we want to exclude in order to take geometric limits of parabolic rescalings near a finite time singularity. In the language of § 4.4, to not locally collapse is essentially equivalent to being κ-noncollapsed on scales $\le \sqrt{T}$ for some κ.

Notice that we can trade off any fixed finite bound on $r_k^2|\text{Rm}|$ on B_k against the bound on r_k^2/t_k by adjusting the r_k, so there is no loss of generality in setting the rescaled curvature bound to 1.

Perelman [26] Thm. 4.1 uses monotonicity of the μ functional to prove no local collapsing at finite times:

Theorem 5.10 (No local collapse)**.** *Let M be closed and suppose Ricci flow is defined on $[0, T)$ where $T < \infty$. Then $g(t)$ is not locally collapsing at T.*

Proof. The idea is to show that if $g(t)$ locally collapses at T, we can find suitable test functions f_k so that $\mathcal{W}(g(t_k), f_k, r_k^2) \to -\infty$. By the definition of μ this implies $\mu(g(t_k), r_k^2) \to -\infty$ and by monotonicity of μ, one has $\mu(g(0), t_k + r_k^2) \to -\infty$. But for a fixed metric $g(0)$, the function μ is continuous in the parameter τ, and therefore $\lim_{k\to\infty} \mu(g(0), t_k + r_k^2)$ is finite. This contradiction will prove the theorem.

It remains to find suitable test functions. For the sake of legibility we'll suppress the subscript k and write $f := f_k$, $r := r_k$ and so on. Since r^2/t is bounded, it makes sense to fix $\tau = r^2$. Let B denote the ball of radius r about p at time t.

As before we can change variables and write $\Phi := e^{-f/2}$ so that

$$\mathcal{W}(g, f, \tau) = \int_M (4\pi\tau)^{-n/2} \left(4\tau|\nabla\Phi|^2 + (\tau R - 2\log\Phi - n)\Phi^2\right) d\text{vol}$$

and μ is the infimum of this functional over all smooth positive Φ with

$$\int (4\pi\tau)^{-n/2}\Phi^2 d\text{vol} = 1$$

In fact, since $x^2 \log x \to 0$ as $x \to 0$ we can take the infimum over non-negative Φ. We will estimate μ by taking Φ to be a suitable bump function, localized near B.

Let ϕ be a function on $\mathbb{R}^+$ equal to 1 on the interval $[0, 1/2]$ and falling off monotonically to 0 at 1. Then $\phi(\text{dist}(p, x)/r)$ is a radial bump function supported near p, and we can consider Φ of the form $\Phi(x) := e^{-c/2}\phi(\text{dist}(p, x)/r)$ where c is a normalization constant.

Now, since $\phi \le 1$ on B and is equal to zero outside, we have

$$\int_M (4\pi r^2)^{-n/2}\Phi^2 d\text{vol} < e^{-c}\text{vol}(B)r^{-n}$$

so to normalize this integral to 1 we must have $c \to -\infty$ under the assumption of local collapse. On the other hand,

$$\mathcal{W} = \int_M (4\pi r^2)^{-n/2} \left(4r^2|\nabla\Phi|^2 + (r^2 R - 2\log\Phi - n)\Phi^2\right) d\text{vol}$$

Let's estimate the contributions of each of these terms. By hypothesis, $|r^2 R|$ is bounded, so that term contributes at most a constant to $\mathcal{W}$, as does the $-n$ term, when integrated against the probability measure $(4\pi r^2)^{-n/2}\Phi^2 d\text{vol}$.

The norm of $\nabla\Phi$ is equal to $e^{-c/2}r^{-1}|\phi'|$, so the $4r^2|\nabla\Phi|^2$ term contributes a bounded term $4|\phi'|^2$ integrated against the measure $(4\pi r^2)^{-n/2}e^{-c}d\text{vol}$ on B. This is not quite a probability measure, though it agrees with the probability measure $(4\pi r^2)^{-n/2}\Phi^2 d\text{vol}$ on the ball of radius $r/2$. However the Bishop–Gromov inequality lets us control the volume on the ball B in terms of the volume of the ball of radius $r/2$, and the conclusion is that $\int_B (4\pi r^2)^{-n/2}e^{-c}d\text{vol}$ is bounded, independent of k. Thus the $4r^2|\nabla\Phi|^2$ term contributes at most a constant to $\mathcal{W}$ too.

Finally, $-2\log\Phi = c - 2\log\phi$. We have already observed that $\phi^2\log\phi \to 0$ where $\phi \to 0$ and is otherwise bounded, so the conclusion is that $\mathcal{W} \le c + \text{const}$. Since $c \to -\infty$ as $k \to \infty$, so does $\mathcal{W}$.

So: $\mu(g(t_k), r_k^2) \to -\infty$ and therefore $\mu(g(0), t_k + r_k^2) \to -\infty$ by Lemma 5.8. This contradicts finiteness of μ for a fixed metric g and bounded τ, and we arrive at the desired contradiction. □

In fact, controlling the entire curvature norm $|\text{Rm}|r_k^2$ on B_k is superfluous in the proof of the theorem (though not for applications to injectivity radius and geometric convergence). Only an upper bound on scalar curvature Rr_k^2 is actually necessary. The use of the Bishop–Gromov inequality appears to require a lower bound on $\text{Ric}\, r_k^2$, but this is only used to control the integral of ϕ' on the outer annulus of B_k — the same control can be obtained by restricting to a smaller ball if necessary.

5.4. **$\mathcal{L}$-length.** Let $M, g(t)$ be Ricci flow. We use the notation $\tau := -t$. For points p, q and times $0 \le \tau_1 < \tau_2$ and a path $\gamma : [\tau_1, \tau_2] \to M$ define the *$\mathcal{L}$-length of* γ, denoted $\mathcal{L}(\gamma)$ to be the quantity

$$\mathcal{L}(\gamma) := \int_{\tau_1}^{\tau_2} \sqrt{\tau}\left(R(\gamma(\tau)) + |\gamma'(\tau)|^2\right) d\tau$$

where $R(\gamma(\tau))$ and $|\gamma'(\tau)|^2$ are computed for the metric $g(\tau)$.

If we fix $(p,0)$ then for any (q,τ) we define the *$\mathcal{L}$-distance* $L(q,\tau) := \inf_\gamma \mathcal{L}(\gamma)$ over all $\gamma : [0,\tau] \to M$ with $\gamma(0) = p$ and $\gamma(\tau) = q$. We also define the *reduced length*

$$l(q,\tau) := L(q,\tau)/2\sqrt{\tau}$$

and the *reduced volume*

$$\tilde{V}(\tau) := \int_M \tau^{-n/2} e^{-l(q,\tau)} d\mathrm{vol}(g(\tau))$$

Example 5.11. Let $M = \mathbb{R}^n$, with constant Euclidean metric. Take $p = 0$. Then for any $(q,\bar{\tau})$ the $\mathcal{L}$-minimizer from $(0,0)$ to $(q,\bar{\tau})$ is $\gamma(\tau) = q\sqrt{\tau/\bar{\tau}}$ which satisfies $L(q,\bar{\tau}) = \int_0^{\bar{\tau}} (q^2/4\bar{\tau})\tau^{-1/2} d\tau = q^2/2\sqrt{\bar{\tau}}$ and $l = q^2/4\bar{\tau}$ so $\tilde{V} = (4\pi)^{n/2}$.

If M is complete, and $|\mathrm{Rm}|$ is bounded on compact time intervals, then for fixed τ and q very far away from p an $\mathcal{L}$-minimizing curve will have $|\gamma'|^2 \gg |R|$ so that $l(q,\tau) \sim d(p,q)^2/4\tau$ and therefore $\tilde{V}$ is finite. Note by the way that l and $\tilde{V}$ are both scale-invariant.

Our goal is to indicate the proof of the following theorem, which is proved in Perelman [26] 7.1:

Theorem 5.12 (Monotone Reduced Volume)**.** *The reduced volume $\tilde{V}$ is non-increasing in τ, and is strictly decreasing unless*

$$\mathrm{Ric}(\tau) + \mathrm{Hess}(l(\tau)) = \frac{g(\tau)}{2\tau}$$

so that (in particular) $M, g(t)$ is a shrinking Ricci soliton.

With the heuristic $l \sim d(p,q)^2/4\tau$ the formula for reduced volume very closely resembles Huisken's S_t functional for mean curvature flow. However, S_t is non-increasing with t, whereas $\tilde{V}(\tau)$ is non-increasing in $\tau = -t$ so that the monotonicity is in the opposite sense!

Theorem 5.12 gives another proof of κ-noncollapsing. Consider Ricci flow for a manifold M on a time interval $[-\bar{\tau}, 0]$ Roughly speaking, if we are κ-collapsed near the point $(p,0)$ then $\tilde{V}$ is very small for small τ, and therefore by monotonicity, also for $\bar{\tau}$ (i.e. for the initial metric). We shall show that there is some point q so that $l(q,\bar{\tau}) \le n/2$. A calculation (Lemma 8.3 in [26]) shows that l can't grow too quickly on a ball of constant size around q in the metric at time $\bar{\tau}$; thus $\tilde{V}(\bar{\tau})$ is bounded from below by a constant depending only on M, and therefore we are κ-noncollapsed at $(p,0)$.

5.4.1. *Sketch of the proof.* Here is the idea of the proof. The function L acts rather like a distance function function from $(p,0)$ to $(q,\bar{\tau})$, and it makes sense to talk about $\mathcal{L}$-geodesics, $\mathcal{L}$-Jacobi fields, the $\mathcal{L}$-Index form and so on. Computing the first and second variation of $\mathcal{L}$-length gives formulae for these objects. Along an $\mathcal{L}$-minimizing geodesic, the index form is minimized by Jacobi fields, amongst all variations with the same end values. Thus if we compute the index form along a suitable family of 'comparison' vector fields V_i along γ which have length proportional to $\sqrt{\tau}$ and are orthonormal at $(q,\bar{\tau})$ we get an upper bound on Δl. Using this we can show $\tilde{V}$ is non-increasing, with equality if and only if the comparison vector fields V_i are $\mathcal{L}$-Jacobi fields, in which case $M, g(t)$ is a gradient shrinking soliton.

This comparison is valid radially along every $\mathcal{L}$-geodesic up to the 'cut locus'; at or beyond it each geodesic makes no further contribution to reduced volume, so the inequality still holds, just as in the proof of the Bishop–Gromov Theorem 4.1.

Here are some of the details of the computation. Compare with the formulae in § 4.1.1:

(1) A path γ is critical for $\mathcal{L}$-length (i.e. it is an $\mathcal{L}$-*geodesic*) iff $X := d\gamma/d\tau$ satisfies $\mathrm{Geo}(X) = 0$ where

$$\mathrm{Geo}(X) := \nabla_X X - \frac{1}{2}\nabla R + \frac{1}{2\tau}X + 2\mathrm{Ric}(X, \cdot)^\sharp$$

Here ∇R means $\mathrm{grad}R$; i.e. $(dR)^\sharp$.

(2) A vector field Y along an $\mathcal{L}$-geodesic γ with $X = d\gamma/d\tau$ is tangent to a variation through $\mathcal{L}$-geodesics (i.e. it is an $\mathcal{L}$-*Jacobi field*) iff $\mathrm{Jac}(Y) = 0$ where

$$\begin{aligned}\mathrm{Jac}(Y) := \nabla_X\nabla_X Y + \mathcal{R}(Y, X)X - \frac{1}{2}\nabla_Y\nabla R \\ + \frac{1}{2\tau}\nabla_X Y + 2(\nabla_Y \mathrm{Ric})(X, \cdot)^\sharp + 2\mathrm{Ric}(\nabla_X Y, \cdot)^\sharp\end{aligned}$$

(3) The Hessian of $\mathcal{L}$-length on the space of smooth variations of an $\mathcal{L}$-geodesic γ from $(p, 0)$ to $(q, \bar{\tau})$ has the form

$$I(Y, Y) = 2\sqrt{\tau}\langle Y', Y\rangle|_0^{\bar{\tau}} - 2\int_0^{\bar{\tau}} \sqrt{\tau}\langle Y, \mathrm{Jac}(Y)\rangle d\tau$$

An $\mathcal{L}$-length minimizing γ has $\mathcal{L}$-length equal to L by definition. Thus

$$\mathrm{Hess}(L)(v, v) \le I(Y, Y)$$

for any vector field Y along γ with $Y(0) = 0$ and $Y(\bar{\tau}) = v$ with equality iff Y is an $\mathcal{L}$-Jacobi field.

(4) By computing $\partial_\tau\langle\nabla_X Y, Y\rangle$ and integrating by parts, the formula for $I(Y, Y)$ can be rewritten as

$$\begin{aligned}I(Y, Y) = \int_0^{\bar{\tau}} \sqrt{\tau}\Big(\mathrm{Hess}(R)(Y, Y) + 2\langle\mathcal{R}(Y, X)Y, X\rangle \\ - 4(\nabla_Y \mathrm{Ric})(X, Y) + 2(\nabla_X \mathrm{Ric})(Y, Y) + 2|\nabla_X Y|^2\Big)d\tau\end{aligned}$$

(5) A vector field Y along an $\mathcal{L}$-geodesic γ from $(p, 0)$ to $(q, \bar{\tau})$ is said to be *adapted* if it is of the form $Y = \sqrt{\tau/\bar{\tau}}v$ where v satisfies $\nabla_X v = -\mathrm{Ric}(v, \cdot)^\sharp$. Note that the length of $v(\gamma(\tau))$ in the $g(\tau)$ metric is constant.

Thus for any vector $u \in T_q M$ we can form an adapted vector field Y where $v(\bar{\tau}) = u$ is as above. For such a Y, we can compute the Index Form:

$$I(Y, Y) = \frac{1}{\sqrt{\bar{\tau}}} - 2\sqrt{\bar{\tau}}\,\mathrm{Ric}(Y, Y) - \int_0^{\bar{\tau}} \sqrt{\tau}H(X, Y)d\tau$$

where $H(X, Y)$ is the term arising in Hamilton's Harnack inequality, and is given by equation 3.3.

Summing over an orthonormal basis for T_qM at time $\bar{\tau}$ and applying the index inequality, we obtain the inequality

$$\Delta l \le \frac{n}{2\bar{\tau}} - R - \frac{1}{2\bar{\tau}^{3/2}} \int_0^{\bar{\tau}} \tau^{3/2} H(X) d\tau$$

(6) For an $\mathcal{L}$-geodesic γ, define

$$K^{\bar{\tau}}(\gamma) := \int_0^{\bar{\tau}} \tau^{3/2} H(X) d\tau$$

Then we can write our inequality as

$$\Delta l \le \frac{n}{2\bar{\tau}} - R - \frac{K^{\bar{\tau}}(\gamma)}{2\bar{\tau}^{3/2}}$$

(7) Using the equation for an $\mathcal{L}$-geodesic one can directly compute

$$\partial_\tau l = R - \frac{1}{\tau} l + \frac{1}{2\tau^{3/2}} K \text{ and } |\nabla l|^2 = -R + \frac{1}{\tau} l - \frac{1}{\tau^{3/2}} K$$

so that we deduce the pointwise inequality

$$\partial_\tau l - \Delta l + |\nabla l|^2 - R + \frac{n}{2\tau} \ge 0 \tag{5.2}$$

This is true along each $\mathcal{L}$-geodesic up to the cut locus, and in the barrier sense thereafter. From this Theorem 5.12 follows.

Compare the inequality in equation 5.2 with the formula for the evolution of f in Equation 5.1, remembering $\tau = -t$.

5.4.2. *Asymptotic solitons.* Further geometric inequalities follow from the estimates in § 5.4.1. If we denote $\bar{L} := 2\sqrt{\tau} L$ then from the formula for l_τ and the index inequality for Δl we obtain the inequality

$$\bar{L}_\tau + \Delta \bar{L} \le 2n$$

Since $l \sim d(p,q)^2/4\tau$ for q far away from p, it follows that for each fixed time τ slice l and therefore $\bar{L}$ is proper. In particular, the spacewise minimum $\bar{L}_{\min}$ is defined for all τ, and the quantity $\bar{L}_{\min} - 2n\tau$ is nonincreasing with τ. In particular, $l_{\min} \le n/2$ for all τ.

Now let's suppose $M, g(t)$ is a κ-solution (see Definition 4.5), perhaps obtained as a blow-up limit of Ricci flow on a compact manifold near a finite time singularity. Since $\mathrm{Rm} \ge 0$ it follows that l is strictly positive everywhere. Since the solution is ancient, Hamilton's Harnack inequality and the formula for $|\nabla l|^2$ imply an inequality $|\nabla l|^2 + R \le Cl/\tau$ for a suitable constant C.

Take a sequence of times $\tau_i \to \infty$ and for each i choose q_i with $l(q_i, \tau_i) \le n/2$. Then from the estimate on $|\nabla l|^2$ for any ϵ we can find a δ so that for $d^2_{\tau_i}(q, q_i) \le \epsilon \tau_i$ and $\tau \in [\tau_i/2, \tau_i]$ we have

$$l(q,\tau) \le \delta^{-1} \text{ and } R(q,\tau) \le \delta^{-1}\tau^{-1}$$

It follows that if we parabolically rescale the flows centered at (q_i, τ_i) by a factor of τ_i^{-1} some subsequence converges on compact subsets to a limit $N, h(t)$ on the time interval $t \in [-1, -1/2]$. One can show that the reduced volume $\tilde{V}(\tau)$ for N is finite, and equal to the limit of the reduced volumes $\tilde{V}(\tau_i \tau)$ for M. Monotonicity of these reduced volumes for

M imply that $\tilde{V}(\tau)$ is *constant*. But this implies equality in equation 5.2 applied to N, so that N is a gradient shrinking soliton.

The reduced volume of N (at any time) is bounded from above by the reduced volume of M (at any time). Now, since M has $\mathrm{Rm} \geq 0$ but is not flat, an easy estimate gives $\tilde{V}(\tau) < (4\pi)^{n/2}$ for M for any τ. The same inequality therefore holds for the reduced volume of N, so that N is not flat either. In short, we have sketched the proof of the following, which is Perelman [26] Prop. 11.2:

Theorem 5.13 (Asymptotic Soliton)**.** *Let $M, g(t)$ be a κ-solution. Then there are (q_i, τ_i) with $\tau_i \to \infty$ so that the sequence of parabolic rescalings of M by factors τ_i^{-1} at the points (q_i, τ_i) converge to a non-flat gradient shrinking soliton.*

5.4.3. *The classification of finite time singularities.* By analyzing the possible asymptotic solitons, Perelman was able to give a complete structure theorem for κ-solutions.

Theorem 5.14. *Let $M, g(t)$ be a gradient shrinking soliton that arises as the asymptotic soliton of a κ-solution. If the dimension $n = 2$ then $M, g(t)$ is a finite quotient of a round shrinking S^2. If the dimension $n = 3$ then $M, g(t)$ is a finite quotient of a round shrinking S^3 or a round shrinking $S^2 \times \mathbb{R}$.*

Note that although a κ-solution has bounded curvature on compact time intervals, we cannot assume this *a priori* for an asymptotic soliton.

We give the outline of a proof.

Proof. The case of dimension 2 was proved by Hamilton. If M is 3-dimensional and does not have strictly positive Ricci curvature, then it splits locally as a product of a line and a 2-dimensional gradient shrinking soliton, so that by the 2-dimensional case $M, g(t)$ is a finite quotient of a shrinking $S^2 \times \mathbb{R}$. That leaves the case that M is noncompact with strictly positive Ricci curvature.

Fix a time slice t and a basepoint p. First let's suppose R is unbounded. Then there are a sequence of points $p_i \to \infty$ so that $R(p_i) \to \infty$. Then certainly also $d_t(p, p_i)^2 R(p_i) \to \infty$. By adjusting the points p_i if necessary we can further insist that $R(z) \leq 4R(p_i)$ whenever $d(p_i, z) \leq CR(p_i)^{-1/2}$ for any constant C. Then we can take a limit of a subsequence $(M, R(p_i)^{-1/2} g(t), p_i) \to (M_\infty, g_\infty, p_\infty)$.

Pass to a subsequence so that the distance-minimizing geodesics from p to p_i converge, and make an angle arbitrarily close to π with the geodesics from p_i to p_{i+1}. In the limit we obtain a *line* in M_∞ — i.e. an isometrically embedded copy of $\mathbb{R}$. By the Splitting Theorem 4.9, M_∞ splits as an isometric product of a surface Σ with $\mathbb{R}$.

Since $M, g(t)$ is a gradient shrinking soliton, and $R(p_i) \to \infty$, the limit M_∞, g_∞ is a non-flat, non-negatively curved gradient *steady* soliton, and therefore so is Σ. The only 2-dimensional gradient-steady Ricci soliton with positive curvature is Hamilton's cigar soliton [14]; however, the cigar soliton is asymptotically flat with bounded injectivity radius near infinity, and is therefore κ-collapsed. Thus we obtain a contradiction.

The last possibility is that M is noncompact with strictly positive Ricci curvature, and R is bounded. As before, we can obtain a limit M_∞, g_∞ which splits as a product and is a gradient-shrinking soliton, and is therefore a round shrinking $S^2 \times \mathbb{R}$, up to a finite quotient. A comparison of the geometry of the level sets of the gradient functions for M and for M_∞ leads to a contradiction in this case. □

Example 5.15 (Bryant soliton). Theorem 5.14 implies that every asymptotic soliton of a non-compact κ-solution is a shrinking round cylinder or a $\mathbb{Z}/2\mathbb{Z}$ quotient of it. The Bryant soliton is a κ-solution; this is a gradient steady soliton, so the time slices are all isometric. On the face of it, this seems to contradict the classification theorem — how can the asymptotic limit of a 'stationary' sequence be different?

The answer of course has to do with basepoints. Near the cap of the Bryant soliton the curvature R stays bounded away from zero, independent of τ. This means that for any sequence (p_i, τ_i) centered in the caps we have $l(p_i, \tau_i) \sim C\tau_i$. To form an asymptotic soliton we need to take points (q_i, τ_i) for which $l(q_i, \tau_i)$ is bounded. Then necessarily $R(q_i)$ has order $1/\tau_i$, and the q_i must exit the end of the soliton. The end of the Bryant soliton is parabolic, and bigger and bigger subsets based at points further and further away converge (after rescaling) to a cylinder.

5.4.4. *Classification of κ-solutions and the structure of high curvature regions.* We now present Perelman's classification of 3-dimensional κ-solutions. First we make some definitions.

Definition 5.16 (ϵ-round component)**.** An *ϵ-round component* is a compact manifold M diffeomorphic to a spherical space-form so that the scaled pullback metric is within ϵ of the round metric in the $C^{1/\epsilon}$ topology.

Definition 5.17 (ϵ-neck)**.** An *ϵ-neck* centered at a point $p \in (M, g)$ is an injective diffeomorphism $\phi : S^2 \times (-\epsilon^{-1}, \epsilon^{-1}) \to M$ so that the scaled pullback metric $R(p)\phi^* g$ is within ϵ of the standard round metric on the cylinder, in the $C^{1/\epsilon}$ topology.

Definition 5.18 (C-component)**.** A *C-component* is a compact manifold with $\mathrm{Rm} > 0$ diffeomorphic to S^3 or $\mathbb{RP}^3$ whose diameter, curvature and volume are bounded (after rescaling) between C^{-1} and C.

Definition 5.19 (C, ϵ-cap)**.** A *C, ϵ-cap* is a noncompact 3-manifold with $\mathrm{Rm} > 0$ which is the union of an ϵ-neck, and a compact core glued along one of the S^2 boundaries of the neck. The core is diffeomorphic to $\mathbb{R}^3$ or $\mathbb{RP}^3$. Its diameter, curvature and volume are bounded (after rescaling) between C^{-1} and C.

An *ϵ-tube* is a product $S^2 \times \mathbb{R}$ in which every point is the center of an ϵ-neck, in such a way that the S^2 foliations of the necks match up with the global product structure. An ϵ-tube can be *capped* or *doubly capped* if one or both of the ends is replaced with a C, ϵ-cap.

Theorem 5.20 (Classification of 3-dimensional κ-solutions)**.** *Every connected oriented 3-dimensional κ-solution is one of the following possibilities:*

(1) A shrinking round spherical space form;
(2) A shrinking round cylinder or finite quotient;
(3) A C-component;
(4) A C-capped ϵ-tube: or
(5) A doubly C-capped ϵ-tube.

We give the sketch of a proof.

Proof. To prove Theorem 5.20 one first shows that the space of pointed 3-dimensional curvature-normalized κ-solutions is compact; this can be proved by the geometric estimates in § 5.4.1. Since the only asymptotic solitons are finite quotients of shrinking round spheres or cylinders, every κ-solution of sufficiently large normalized diameter is made up of ϵ-tubes and regions of uniformly bounded diameter.

A κ-solution which is not already a shrinking cylinder or quotient has strict $\mathrm{Rm} > 0$, so if it is non-compact, the Cheeger–Gromoll Soul Theorem implies it is diffeomorphic to $\mathbb{R}^3$. Thus a noncompact κ-solution with $\mathrm{Rm} > 0$ is a C-capped ϵ-tube where the existence of such a C comes implicitly from the compactness of the space of κ-solutions.

A similar argument shows that a compact κ-solution either has uniformly bounded diameter, or it is a doubly-capped ϵ-tube. In every case $\mathrm{Rm} > 0$ so Hamilton's theorem implies that the manifold is diffeomorphic to a spherical space form. The uniformly bounded diameter components are either round or C-components, where again C comes from the compactness of the space of κ-solutions. □

Example 5.21. There are infinitely many diffeomorphism types of lens spaces. These are all spherical space forms. However most are very highly collapsed: only finitely many diffeomorphism types can arise as a κ-solution, for any fixed κ.

Example 5.22. The Bryant soliton is an example of a C-capped ϵ-tube. A C-capped ϵ-tube can occur as the parabolic blow up near a degenerate neckpinch singularity.

Example 5.23. A C-component is a kind of degenerate doubly C-capped ϵ-tube in which the 'tube' part has normalized diameter less than $1/\epsilon$. In backwards time the length of this neck increases indefinitely.

We now state the main theorem on the classification of high curvature regions in finite time. For applications to surgery, it's important to control the various constants involved.

We say that a Riemannian manifold M of dimension n is *normalized* if

(1) $|\mathrm{Rm}| \le 1$ everywhere; and
(2) for every $p \in M$ we have $\mathrm{vol}(B_1(p)) \ge \omega/2$ where ω is the volume of the unit ball in $\mathbb{R}^3$.

Any metric on a compact manifold may be normalized by rescaling it suitably.

Theorem 5.24 (Finite time high curvature region)**.** *Let $M, g(t)$ be Ricci flow on a compact orientable 3-manifold defined on some time interval $[0, T)$. Suppose that the metric at time 0 is normalized. Then for every $\epsilon > 0$ there exists $C > 0$ depending on ϵ, and K depending on ϵ and T so that for every $t \in [0, T)$, the subset of $(M, g(t))$ with $R \ge K$ is partitioned into four subsets:*

(1) ϵ-round components;
(2) C-components;
(3) centers of ϵ-necks; and
(4) cores of C, ϵ-caps.

Morally speaking, this theorem is proved by taking parabolic blow-ups to produce κ-solutions, and analyzing the possibilities for their associated asymptotic gradient-shrinking solitons.

In the regions covered by ϵ-necks the foliations by spheres can be matched up topologically by a small perturbation. Thus a compact component of the part of $M, g(t)$ where $R \ge K$ is diffeomorphic to a spherical space form or a finite quotient of $S^2 \times S^1$.

6. The Geometrization Conjecture

In the remainder of this chapter we give the barest outline of Perelman's argument to prove the Geometrization Conjecture. For details see Perelman [27, 28], Kleiner–Lott [23] or Morgan–Tian [24, 25].

6.1. **Surgery.**

We have seen that when high curvature regions develop in finite time, they are either compact components (which are topologically finite quotients of S^3 or $S^2 \times S^1$) or they are ϵ-tubes, possibly capped at one end.

Surgery takes the manifold at some fixed time, and modifies it in the high curvature regions. The compact components are discarded, and the ϵ-tubes are truncated near their uncapped ends and replaced with round three-balls with a specially chosen metric, called the *standard solution*.

This changes the underlying diffeomorphism type of the manifold. Let M be the manifold before surgery, and N the manifold after. Then topologically M is obtained from N in three steps:

(1) certain distinct components of N are connect summed together;
(2) some components of the result are self-connect summed — equivalently, they are connect summed with copies of $S^2 \times S^1$; and
(3) finally, we add the disjoint union of finitely many S^3 or $S^2 \times S^1$ quotients.

After surgery, Ricci flow is restarted on N. After a further finite amount of time, more high curvature regions will develop, and we can perform another surgery operation, and repeat the process.

6.1.1. *The standard solution.* The *standard solution* is a Ricci flow $M, g(t)$ where M is the 3-ball, and $g(0)$ is a complete metric obtained (roughly speaking) by gluing a round hemispherical cap to a round $S^2 \times \mathbb{R}^+$ with $R = 1$. The key properties this should satisfy are:

(1) SO(3) symmetry;
(2) positive curvature $\mathrm{Rm} > 0$ throughout, for all t;
(3) the solution exists for $t \in [0, 1)$;
(4) the end of the solution asymptotically matches the shrinking round unit cylinder $S^2 \times \mathbb{R}$; and
(5) there is some r and κ so that the solution is κ-noncollapsed on scales less than r for all $t \in [0, 1)$.

The existence of a standard solution is proved by direct construction. Since M is noncompact, one needs to check existence and uniqueness of Ricci flow. This can be done by adapting the DeTurck trick to the noncompact setting, with a careful analysis of the geometry of the ends.

6.1.2. *Estimates beyond surgery.* Before surgery we have a substantial amount of geometric control on the metric including Hamilton-Ivey pinching, κ-noncollapsing, and a structure theorem for the high-curvature regions. This control is expressed in terms of several constants which in turn depend on the initial metric, and the time elapsed. It's important to be able to perform surgery in such a way that this geometric control persists.

The properties of the standard solution ensure that Hamilton-Ivey pinching persists after surgery. Likewise, κ-noncollapsing holds within the region where a neck has been replaced by the standard solution. Away from it, one needs a *localized* version of κ-noncollapsing. As explained in § 5.4, κ-noncollapsing can be derived from monotonicity of reduced volume. Furthermore, the argument proving monotonicity of reduced volume is *localized* — it follows from the pointwise inequality equation 5.2, true along each $\mathcal{L}$-geodesic up to the cut locus. Performing surgery produces 'holes' in these $\mathcal{L}$-geodesics, making a comparison argument a priori difficult. However, it turns out that the holes only occur in $\mathcal{L}$-geodesics for which the l-length is big, and therefore the argument sketched in § 5.4 deriving κ-noncollapsing from monotonicity of $\tilde{V}$ still goes through.

6.1.3. *Surgeries do not accumulate.* The key point here is to show that if we perform a surgery on a neck of width h, the volume goes down by a definite constant $c(h)$. If we are careful, we can ensure that surgeries at time t are all performed on necks of width at least $h(t)$, in such a way that other quantities (κ, pinching constants etc.) can be controlled in terms of the original metric and the time t. At non-surgery times, volume satisfies $d\mathrm{vol}(t)/dt \leq -R_{\min}(t)\mathrm{vol}(t)$, which is to say it grows at most exponentially in time. Thus, volume can grow by only a bounded amount in any finite time interval, so that there can be only finitely many surgeries in that time interval.

6.2. **Finite time extinction.** We have seen that a 3-manifold which becomes extinct in finite time under Ricci flow with surgery satisfies the geometrization conjecture. More precisely, such a manifold is necessarily a connect sum of spherical space forms and finite quotients of $S^2 \times S^1$, and therefore its fundamental group is a free product of finite groups and infinite cyclic groups.

In fact, it turns out that the converse is true:

Theorem 6.1 (Finite time extinction; Perelman [28])**.** *The following conditions are equivalent for a compact* 3*-manifold* M*:*

(1) M *becomes extinct in finite time under Ricci flow with surgery;*
(2) M *is a finite connect sum of spherical space forms and finite quotients of* $S^2 \times S^1$*;*
(3) $\pi_1(M)$ *is a finite free product of finite and infinite cyclic groups.*

That (1) implies (2) follows from our analysis of surgery. That (2) implies (3) is obvious. In the rest of this section we shall prove (3) implies (1), following Perelman. The most significant corollary is:

Corollary 6.2 (Poincaré Conjecture)**.** *Let* M *be a compact* 3*-manifold with* $\pi_1(M)$ *trivial. Then* M *is homeomorphic to* S^3*.*

Proof. By Theorem 6.1, M is a finite connect sum of spherical space forms and finite quotients of $S^2 \times S^1$. All of these have nontrivial π_1 except for S^3. Thus M is a finite connect sum of S^3s and is therefore homeomorphic to S^3. □

6.2.1. *Nontrivial* π_2. A nontrivial free decomposition of $\pi_1(M)$ gives rise to a nontrivial action of $\pi_1(M)$ on a tree with trivial edge stabilizers, which correspond to homotopically essential S^2s. The first step in the proof of Theorem 6.1 is to show that after a finite amount of time, the result of Ricci flow with surgery is a finite collection of components all with trivial π_2.

Let's call a *homotopically essential* surgery one of the following three operations:

(1) cut along a non-separating S^2
(2) cut along a homotopically nontrivial separating S^2
(3) a π_1-nontrivial spherical space form or quotient of $S^2 \times S^1$ is discarded

The number of homotopically essential surgeries is bounded by the number of terms in a free decomposition of $\pi_1(M)$, which (by Grushko's theorem) is bounded by the number of generators in any generating set. Thus for any initial compact M there can only be finitely many homotopically essential surgeries. We now show:

Proposition 6.3. *Suppose M is a compact manifold with $\pi_2(M)$ nontrivial. Then under Ricci flow there must be some essential surgery in finite time. Consequently, for any compact manifold M, there is some finite time T so that under Ricci flow with surgery, every component has trivial π_2 for all $t > T$.*

To prove Proposition 6.3 we let W_2 denote the least area of a map of a sphere in M representing some nontrivial class in $\pi_2(M)$. Note that this minimum is achieved, and realized by some (possibly branched) minimal surface $F : S^2 \to M$, by the theorem of Sacks-Uhlenbeck. The function W_2 is continuous in time as we evolve M under Ricci flow until we come to a surgery. If the surgery is homotopically essential we are done. If the surgery is inessential it splits off finitely many homotopy S^3 summands and a unique component M' homotopy equivalent to M. Thus we can consider nontrivial classes in $\pi_2(M')$ and the least area of a sphere representing some such class, so that W_2 continues to be well-defined until the first homotopically essential surgery.

We claim that under Ricci flow with only inessential surgeries, W_2 satisfies the following differential inequality:

$$\frac{dW_2}{dt} \le -4\pi - \frac{1}{2} R_{\min} W_2$$

in the sense of forward difference quotients. Let's see how Proposition 6.3 follows from this.

If we normalize the metric so that $|\mathrm{Rm}|_{\max} = 1$ at time 0 then $R_{\min}(0) \ge -6$ so by Proposition 3.3 we have an inequality $R_{\min}(t) \ge -6/(4t+1)$ under Ricci flow. At a surgery, regions with $R \gg 0$ are cut out and replaced with standard solutions; thus, the spatial minimum of R is not affected by surgery, and this differential inequality for $R_{\min}(t)$ is true under Ricci flow with surgery.

Consequently, if there are no essential surgeries, W_2 is bounded above by a solution to the differential equation

$$\frac{dw_2}{dt} = -4\pi + \frac{3w_2}{(4t+1)}$$

This can be solved explicitly; the general solution is

$$w_2(t) = C(4t+1)^{3/4} - 4\pi(4t+1)$$

so that w_2 (and hence W_2) must become negative in finite time. This is absurd, and the claim is proved, modulo the differential inequality for W_2.

We now prove the differential inequality. At every time t there is a minimal immersed 2-sphere $F : S^2 \to M$ so that $W_2 = \text{area}(F)$. First let's consider the effect on W_2 of an inessential surgery. There is a finite system of inessential 2-spheres $\Sigma = \cup_i \Sigma_i$ which decompose M into $M_0 \cup B_i$ where the B_i are all 3-balls. The effect of surgery is to cut out all the B_i and insert a collection of rounded 3-balls B_i' in their place, modelled on the standard solution. Each B_i consists of a neck neighborhood of Σ_i, together with a manifold A_i which happens to be diffeomorphic to a ball. Evidently there is a 1-Lipschitz map $\psi_i : B_i \to B_i'$ which crushes A_i to a point, and pinches the neck neighborhood down to a rounded ball. Performing ψ_i on each B_i produces a 1-Lipschitz map from M to the result of surgery, and carries F to a new map F' representing the same homotopy class (in the sense we have discussed), but $\text{area}(F') \le \text{area}(F)$. In other words, W_2 can only go down when we perform an inessential surgery.

It remains to prove the differential inequality when M evolves by ordinary Ricci flow. For any surface F, the area of F evolves under Ricci flow as

$$\frac{d}{dt}\text{area}(F) = \int_F \frac{1}{2}\text{tr}|_F\left(\frac{\partial g}{\partial t}\right) d\text{area} = \int_F \sum_i -\text{Ric}(e_i) d\text{area} = -\int_F (R - \text{Ric}(n)) d\text{area}$$

where e_i is an orthonormal basis for TF, and n is the unit normal vector to F.

Now, $R = 2\text{Ric}(n) + 2K_M$ where K_M is the sectional curvature of the tangent plane to F, as measured in M. For a minimal surface $K_M = K + |A|^2/2$ where A is the second fundamental form and K is the sectional curvature of the tangent plane to F *as measured in* F. Thus we can rewrite this formula as

$$\frac{d}{dt}\text{area}(F) = -\int_F K d\text{area} - \frac{1}{2}\int_F (|A|^2 + R) d\text{area}$$

By Gauss-Bonnet the first term is 4π, while of course $|A|^2 + R \ge R_{\min}$ pointwise. From this the differential inequality follows, completing the proof of Proposition 6.3.

6.2.2. *Nontrivial* π_3. We now continue the proof of Theorem 6.1. From Proposition 6.3 we conclude that for any compact M, after finite time Ricci flow with surgery results in a manifold N consisting of finitely many components, each of which has π_2 trivial. If our original manifold had $\pi_1(M)$ equal to the free product of finitely many $\mathbb{Z}$s and finite groups, then each component of N has finite π_1, and is therefore finitely covered by a homotopy 3-sphere. In other words, each component of N has $\pi_3 = \mathbb{Z}$. Thus the proof of the theorem will follow from:

Proposition 6.4. *Suppose N is a compact 3-manifold with $\pi_2(M)$ trivial and $\pi_3(N) = \mathbb{Z}$. Then under Ricci flow with surgery N vanishes in finite time.*

For such an N every future surgery either consists of cutting along a neck whose core is a homotopically trivial (hence separating) S^2, or throwing away a spherical space form. Let γ be the generator of $\pi_3(N)$. If we decompose N by connect sum into a collection of pieces, then γ restricts to the generator of π_3 of each piece.

Now, if $\pi_2(N) = 1$ then $\pi_3(M)$ is equal to $\pi_2(\Lambda M)$ where ΛM denotes the space of homotopically trivial loops in M. To see this, observe that a map from S^2 to ΛM is the same thing as a map from $S^2 \times S^1$ to M. If the base point in S^2 maps to the trivial loop, then this map factors through $S^2 \times S^1/\text{point} \times S^1$ which is homotopic to $S^3 \vee S^2$ so if π_2

is trivial, this is the same (up to homotopy) as a map from S^3 to M. We may therefore think of γ as an element of $\pi_2(\Lambda M)$.

A representative of γ is a family of contractible loops in M. For each such loop we may compute the minimum area of a spanning disk, and take the maximum area over all loops in the family, and then define W_γ to be the minimum of this quantity over all families representing γ.

We claim that under Ricci flow W_γ satisfies the following differential inequality:

$$\frac{dW_\gamma}{dt} \le -2\pi - \frac{1}{2} R_{\min} W_\gamma$$

in the sense of forward difference quotients. We mean this in the following sense. As we pass through a surgery, our manifold decomposes into a finite collection of summands, and the class γ can be 'continued' to a class in π_3 of each summand. Then we claim that the differential inequality holds for W_γ as computed in any summand as we pass through the surgery. As we argued previously, this implies that unless the component we are following vanishes altogether (which is what we want), W_γ would become zero in finite time (which is absurd). In other words, there is an a priori estimate of a time after which every component obtained from N by Ricci flow with surgery becomes extinct. This completes the proof of the proposition and therefore the theorem, modulo the proof of the differential inequality.

At surgery times the monotonicity of W_γ follows for the same reasons as for the corresponding statement for W_2; if N is the manifold before surgery and N' a component afterwards, there is a 1-Lipschitz map $\psi : N \to N'$ taking the class of γ in N to the class of γ in N'. A minimax family realizing W_γ in N is taken to a new family in N'. Since area can only go down under ψ, the same is true for W_γ.

Thus it remains to verify the differential inequality under ordinary Ricci flow. For each loop c in a minimax family γ we can span c by a minimal area disk D, and simultaneously evolve c by curve-shortening. That is, if H is the geodesic curvature vector of c in N (in the metric at time t) we let $dc/dt = H$. Note that for *any* disk D spanned by c, the inner product $H \cdot \nu$ of H with the inner unit normal field ν to ∂D in D is equal to the geodesic curvature k of ∂D in D.

If c is sufficiently smooth, we can compute as before (since D is minimal)

$$\begin{aligned}
\frac{d}{dt}\text{area}(D) &= \int_D \frac{1}{2}\text{tr}|_D\left(\frac{\partial g}{\partial t}\right) d\text{area} - \int_{\partial D} k d\text{length} \\
&= -\int_D K d\text{area} - \int_{\partial D} k d\text{length} - \frac{1}{2}\int_D (|A|^2 + R) d\text{area} \\
&\le -2\pi - \frac{1}{2} R_{\min}\text{area}(D)
\end{aligned}$$

There is a serious technical problem, that under the curve-shortening flow c might become singular. This is a nontrivial issue, solved by Perelman using the method of ramps, but we do not go into it here.

6.3. **Infinite time.** Now consider a general compact orientable 3-manifold M and evolve it by Ricci Flow with surgery. By Proposition 6.3 and Proposition 6.4 together with the Poincaré Conjecture, after a finite amount of time every remaining component is an irreducible $K(\pi, 1)$, at which point every future surgery merely pinches off an inessential neck bounding a 3-ball.

Bamler has recently shown that if surgeries are performed correctly, only finitely many surgeries will occur, and after some time there is a constant C so that curvature is bounded by Ct^{-1}. This confirms a conjecture of Perelman from [27]. Bamler's work is carried out in a series of papers, summarized in [3]. Although very nice to know, this fact is technically unnecessary for applications to Geometrization.

6.3.1. *Thick-thin decomposition.* In finite time we have seen that there can be no local volume collapsing under Ricci flow with surgery. However, as $t \to \infty$ asymptotic collapsing is possible, and in fact it must necessarily occur whenever the manifold has a nontrivial JSJ decomposition.

It turns out that for $t \gg 0$ the manifold in the metric $g(t)$ decomposes neatly into two pieces — a thick part, where the volume is locally non-collapsed on the negative curvature scale, and a thin part, where it is.

Given $w > 0$ and a non-negative function ψ on M we say that M is *w-locally volume collapsed on scale ψ* if for all $x \in M$ we have an estimate

$$\mathrm{vol}(B(x, \psi(x))) \leq w\psi(x)^n$$

If M is complete and connected and has negative sectional curvature somewhere, we defined the *negative curvature scale* to be the non-negative function ρ on M such that at each point x, if we rescale the metric on the ball $B(x, \rho(x))$ to have radius 1, then the infimum of the (rescaled) sectional curvature on this ball is equal to -1. In other words, for each x the value of $\rho(x)$ is such that the infimum of the sectional curvature on the ball $B(x, \rho(x))$ is $-\rho(x)^{-2}$.

With this definition, for each time t and for each $w > 0$ we define the *w-thin part* $M^-(w, t)$ to be the subset of points x in M which are w-locally volume collapsed on the negative curvature scale in the metric $g(t)$; and we define the *w-thick part* $M^+(w, t)$ to be the complement.

6.3.2. *Incompressible tori.* Hamilton [18] § 11-12 already more or less analyzed the thick-thin decomposition, and showed that the thick part is asymptotically hyperbolic, and that the two pieces meet along a family of incompressible tori. We shall indicate Hamilton's proof that the thick part is asymptotically hyperbolic in § 6.3.3.

Incompressibility of the tori is proved by an argument very similar to the argument we already gave in § 6.2.2, namely: if one of these tori is compressible, one can define $A(t)$ to be the least area of a compressing disk with boundary on the torus, and show that for every $\delta > 0$, when $t \gg 0$ there is a differential inequality $dA/dt \leq -(2\pi - \delta)$ in the sense of forward difference quotients (an additional $-(1/2)R_{\min}A(t)$ term can be absorbed in δ when t is large, since $R_{\min} = O(1/t)$). This inequality is valid at nonsurgery times, but as we argued in § 6.2.2 the area of a minimizing disk can only go down under surgery, so the inequality is valid more generally. Since A must be positive for all t this gives a contradiction.

6.3.3. *Hamilton's Endgame.* Recall that if we normalize the metric so that $R_{\min}(0) \geq -6$, then $R_{\min}(t) \geq -6/(4t+1)$. Define normalizations

$$\hat{R}(t) := R_{\min}(t)\text{vol}(t)^{2/3} \text{ and } \hat{V}(t) := \text{vol}(t)/(t+1/4)^{3/2}$$

Since $d\text{vol}(t)/dt = -\int_M R dt \leq -R_{\min}\text{vol}(t)$ it follows that away from surgery times,

$$\frac{d\hat{V}(t)}{dt} \leq -\hat{V}(t)\left(R_{\min} + \frac{3}{2(t+1/4)}\right)$$

and

$$\frac{d\hat{R}(t)}{dt} \geq \frac{2}{3}\hat{R}(t)\text{vol}(t)^{-1}\int (R_{\min} - R)d\text{vol}$$

At a surgery time $R_{\min}$ does not change, whereas vol(t) goes down. Hence $\hat{V}$ is monotone nonincreasing, and $\hat{R}$ is monotone nondecreasing.

Define $\hat{V}(\infty)$ and $\hat{R}(\infty)$ to be the limits of these two quantities. The analysis falls into two cases.

If $\hat{V}(\infty) > 0$ then the thick part $M^+(w,t)$ is nonempty for all t. After we rescale the metric by $1/t$, for any fixed positive w the geometry of the thick part $M^+(w,t)$ becomes bounded, in the sense that the sectional curvatures and the injectivity radius are bounded below while the volume is bounded above. In particular, the diameter of the rescaled metric is bounded above.

Theorem 6.5 (Thick part hyperbolic)**.** *Suppose $\hat{V}(\infty) > 0$. Then for any $w > 0$, throughout the thick part $M^+(w,t)$ the rescaled metrics converge to a metric of constant sectional curvature $-1/4$.*

Leaving aside the issue of convergence, we indicate the proof of this theorem, following Hamilton [18] § 7.

Proof. First we show that $R_{\min}(t)$ is asymptotic to $-3/2t$. From the formulae above, we obtain the estimate for the logarithmic derivative of $\hat{V}$:

$$\frac{d(\log(\hat{V}))}{dt} \leq -\left(R_{\min}(t) + \frac{3}{2(t+1/4)}\right)$$

Since we are in the case $\hat{V}(\infty) > 0$ it follows that

$$\int_0^\infty R_{\min}(t) + \frac{3}{2(t+1/4)}dt < \infty$$

Now,

$$\hat{R}(\infty)/\hat{V}(\infty)^{2/3} = \lim_{t\to\infty} R_{\min}(t)(t+1/4)$$

so this limit exists, and by the inequality above it is equal to $-3/2$.

From the formula for $d\hat{R}(t)/dt$ and the positivity of $\hat{V}(\infty)$ it follows that when t is very large, $|R_{\min} - R|$ must be small throughout most of the thick part. Since for any w the rescaled diameter of the thick part is uniformly bounded, and we can control ∇Rm in terms of $|\text{Rm}|$, it follows that $tR(t)$ must converge to $-3/2$ throughout $M^+(w,t)$. From the evolution equation for normalized R it follows then that the trace-free part of normalized

Ric must converge to zero, so that the normalized metric has constant sectional curvature $-1/4$ as claimed. □

If the thin part is empty, we are done. Otherwise the frontier of the thick part consists of a nontrivial union of incompressible tori. Since (by Theorem 6.1) M is irreducible, it is Haken, and therefore already known to be geometric by Thurston. This completes the proof when $\hat{V}(\infty) > 0$.

If $\hat{V}(\infty) = 0$ then for any positive w we eventually have $M^+(w,t)$ empty, and the entire manifold is in the thin part. In this case Perelman simply asserts [28] Thm. 7.4 that M is either flat or has the structure of a graph manifold, and that the proof will be given in a separate paper. In fact no such paper ever appeared, but a proof of the assertion can be deduced in this case from the theory of collapsing with lower curvature bounds as developed by Shioya–Yamaguchi [29, 30].

The main theorem of the latter paper, i.e. [30] Thm. 1.1 is as follows:

Theorem 6.6 (Shioya–Yamaguchi Volume collapse)**.** *There exist $\epsilon, \delta > 0$ so that if M is an orientable Riemannian 3-manifold with* $\text{Rm} \ge -1$ *and* $\text{vol}(M) < \epsilon$ *then either M is homeomorphic to a graph manifold, or diam$(M) < \delta$ and $\pi_1(M)$ is finite.*

Now, if $\hat{V}(\infty) = 0$ then for all sufficiently large times we can rescale the metric on M to satisfy the hypotheses of this theorem. We have already seen that a prime summand with finite fundamental group will become extinct in finite time under Ricci flow with surgery. Thus we conclude that M is a graph manifold in this case, and the proof of geometrization is complete.

7. Acknowledgments

Most of the little I know about Ricci flow — and geometric flows on 3-manifolds in general — I learned at one time or another from Ian Agol, Klaus Ecker, Tom Ilmanen, Bruce Kleiner, Bill Minicozzi, Marty Ross and Terry Tao. These notes borrow heavily from the books of Chow–Knopf [7], Morgan–Tian [24, 25], Topping [32], and the online notes of Bamler [2], Kleiner–Lott [23] and Tao [31]. Danny Calegari was supported by NSF grant DMS 1405466.

References

[1] S. Angenent, *Shrinking doughnuts*, Nonlinear diffusion equations and their equilibrium states, 3 (Gregynog, 1989), 21–38, Progr. Nonlinear Diff. Eq. Appl. 7, Birkhäuser Boston, Boston, Ma, 1992

[2] R. Bamler, *Ricci flow*, Lecture notes by Otis Chodosh and Christos Mantoulidis, http://www.mit.edu/~cmad/Papers/RicciFlowNotes.pdf

[3] R. Bamler, *Long-time behavior of 3 dimensional Ricci flow — Introduction*, Geom. Topol. **22** (2018), 757–774

[4] J. Cheeger and D. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland Publ. Co., Amsterdam 1975

[5] J. Cheeger and D. Gromoll, *On the Structure of Complete Manifolds of Nonnegative Curvature*, Ann. Math (2) **96** (1972), 413–443

[6] J. Cheeger, M. Gromov and M. Taylor, *Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds*, J. Diff. Geom. **17** (1982), 15–53

[7] B. Chow and D. Knopf, *The Ricci Flow: an Introduction*, AMS MSM **110**, Providence, RI, 2004

[8] B. Chow and P. Lu, *The time-dependent maximum principle for systems of parabolic equations subject to an avoidance set*, Pac. J. Math. **214** (2004), no. 2, 201–222

[9] T. Colding and W. Minicozzi, *Generic mean curvature flow I; generic singularities*, Ann. Math **175** (2012), 755–833

[10] D. DeTurck, *Deforming metrics in the direction of their Ricci tensors*, J. Diff. Geom. **18** (1983), no. 1, 157–162

[11] M. Gromov, *Sign and geometric meaning of curvature*, Rend. Sem Mat e Fis di Milano, **61** (1991), no. 1, 9–123

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

[13] R. Hamilton, *Four-manifolds with positive curvature operator*, J. Diff. Geom. **24** (1986), 153–179

[14] R. Hamilton, *The Ricci flow on surfaces*, in 'Mathematics and general relativity' (Santa Cruz, CA, 1986), Contemp. Math. **71**, 237–262, AMS, Providence, RI, 1988

[15] R. Hamilton, *The Harnack estimate for the Ricci flow*, J. Diff. Geom. **37** (1993), 225–243

[16] R. Hamilton, *Harnack estimate for the mean curvature flow*, J. Diff. Geom. **41** (1995), 215–226

[17] R. Hamilton, *Four-manifolds with positive isotropic curvature*, Comm. Anal. Geom. **5** (1997), no. 1, 1–92

[18] R. Hamilton, *Non-singular solutions of the Ricci flow on three-manifolds*, Comm. Anal. Geom. **7** (1999), 695–729

[19] G. Huisken, *Flow by mean curvature of convex surfaces into spheres*, J. Diff. Geom. **20** (1984), 237–266

[20] G. Huisken, *Asymptotic behavior for singularities of the mean curvature flow*, J. Diff. Geom. **31** (1991), 285–299

[21] G. Huisken and C. Sinestrari, *Mean curvature flow with surgeries of two-convex hypersurfaces*, Invent. Math. **175** (1) (2009), 137–221

[22] T. Ilmanen, *Singularites of mean curvature flow of surfaces*, preprint, available from the author's website

[23] B. Kleiner and J. Lott, *Notes on Perelman's Papers*, preprint, https://arxiv.org/abs/math/0605667

[24] J. Morgan and G. Tian, *Ricci flow and the Poincaré Conjecture*, AMS Clay Math. Monog. **3** 2007

[25] J. Morgan and G. Tian, *The Geometrization Conjecture*, AMS Clay Math. Monog. **5** 2014

[26] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, preprint, https://arxiv.org/abs/math/0211159

[27] G. Perelman, *Ricci flow with surgery on three-manifolds*, preprint, https://arxiv.org/abs/math/0303109

[28] G. Perelman, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, preprint, https://arxiv.org/abs/math/0307245

[29] T. Shioya and T. Yamaguchi, *Collapsing three-manifolds under a lower curvature bound*, J. Diff. Geom. **56** (2000), 1–66

[30] T. Shioya and T. Yamaguchi, *Volume collapsed three-manifolds with a lower curvature bound*, Math. Ann. **333** (2005), no. 1, 131–155

[31] T. Tao, *Perelman's proof of the Poincaré Conjecture*, online lecture notes, https://terrytao.wordpress.com/category/teaching/285g-poincare-conjecture/page/1/

[32] P. Topping, *Lectures on the Ricci flow*, LMS Lect. Note Ser. **325**, Cambridge University Press, Cambridge, 2006

[33] B. White, *Evolution of Curves and Surfaces by Mean Curvature*, ICM 2002 Vol. **1** 525–538

University of Chicago, Chicago, Ill 60637 USA
E-mail address: dannyc@math.uchicago.edu