

Ricci flow and Perelman's proof of the Poincaré conjecture

Siddhartha Gadgil* and Harish Seshadri

The Poincaré conjecture was one of the most fundamental unsolved problems in mathematics for close to a century. This was solved in a series of highly original preprints by the Russian mathematician Grisha Perelman, for which he was awarded the Fields Medal (2006). Perelman's proof, building on the work of Hamilton, was based on the Ricci flow, which resembles a nonlinear heat equation. Many of Perelman's and Hamilton's fundamental ideas may be of considerable significance in other settings. This article gives an exposition of the work, starting with some basic concepts.

Keywords: Poincaré conjecture, Ricci flow, topology.

THE field of topology was born out of the realization that in some fundamental sense, a sphere and an ellipsoid resemble each other but differ from a torus – the surface of a rubber tube (or a doughnut). A striking instance of this can be seen by imagining water flowing smoothly on them. On the surface of a sphere or an ellipsoid (or an egg), the water must (at any given instant of time) be stationary somewhere. This is not so in the case of the torus.

In topology, we regard the sphere and the ellipsoid as having the same topological type, which we make precise later. Topology is the study of properties that are shared by objects of the same topological type. These are generally the global properties. Understanding the different topological types of spaces, the so-called classification problem is thus a fundamental question in topology.

In the case of surfaces (more precisely closed surfaces), there are two infinite sequences of topological types. The first sequence, consisting of the so-called orientable surfaces, is composed of the sphere, the torus, the two-holed torus, the three-holed torus and so on (see Figure 1). One would like to have a similar classification in all dimensions. However, due to fundamental algorithmic issues, it is impossible to have such a list in dimensions four and above.

There is a simple way to characterize the sphere among surfaces. If we take any curve on the sphere, we can shrink it to a point while remaining on the sphere. A space with this property is called simply-connected. A torus is not simply-connected as a curve that goes around the torus cannot be shrunk to a point while remaining on the torus. In fact, the sphere is the only simply-connected surface.

In 1904, Poincaré raised the question as to whether a similar characterization of the (three-dimensional) sphere

holds in dimension 3. That this is so has come to be known as the Poincaré conjecture. As topology exploded in the twentieth century, several attempts were made to prove this (and some to disprove it). However, at the turn of the millennium, this remained unsolved. Surprisingly, the higher dimensional analogue of this statement turned out to be easier and has been solved.

In 2002–03, three preprints^{1–3} rich in ideas but frugal with details, were posted by the Russian mathematician Grisha Perelman, who had been working on this in solitude for seven years at the Steklov Institute. These were based on the Ricci flow, which was introduced by Richard Hamilton in 1982. Hamilton had developed the theory of Ricci flow through the 1980s and 1990s, proving many important results and developing a program⁴ which, if completed, would lead to the Poincaré conjecture and much more. Perelman introduced a series of highly original ideas and powerful techniques to complete Hamilton's program.

It has taken two years for the mathematical community to assimilate Perelman's ideas and expand his preprints into complete proofs. Recently, a book⁵ containing complete and mostly self-contained proofs of the Poincaré conjecture has been posted. An earlier set of notes had filled in many details in Perelman's papers⁶.

In this article we attempt to give an exposition of Perelman's work and the mathematics that went into it.

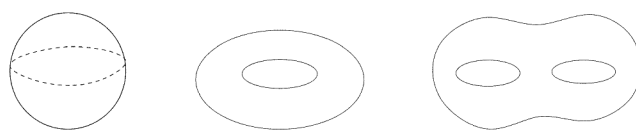


Figure 1. The first three orientable surfaces.

The authors are in the Department of Mathematics, Indian Institute of Science, Bangalore 560 012, India.

*For correspondence. (e-mail: gadgil@math.iisc.ernet.in)

An expanded version of this article will be submitted to the *Mathematical Intelligencer*.

Some notions of topology

In this section, we informally formulate the Poincaré conjecture. To do this, we first need to introduce the higher-dimensional analogues of surfaces, namely smooth manifolds. For those in the know, we consider throughout diffeomorphism types of smooth manifolds as this suffices in dimension 3.

We first take a closer look at surfaces. A surface in \mathbf{R}^3 is the set of zeroes of a smooth function $f(x, y, z)$ which is non-singular, i.e. for each point on the surface the gradient $\nabla f(x, y, z)$ of f is non-zero. Basic examples of this are the plane $z = 0$ and the sphere $x^2 + y^2 + z^2 - 1 = 0$.

In analogy with this, we can consider a subset $M \subset \mathbf{R}^n$ which is the set of zeroes of $n - k$ smooth functions f_1, \dots, f_{n-k} whose gradients ∇f_i are linearly independent for all points in M . Such a subset of \mathbf{R}^n is a k -dimensional manifold or a k -manifold.

More generally, a set M given as above may have several components. We consider each component of M to be a k -manifold. For the rest of this article, by a k -manifold M we mean a component of the subset $M \subset \mathbf{R}^n$ which is the set of zeroes of $n - k$ smooth functions f_1, \dots, f_{n-k} , whose gradients ∇f_i are linearly independent for all points in M . (This is equivalent to the usual definition by a theorem of Nash.)

We say that two smooth k -dimensional manifolds M and N are diffeomorphic if there is a smooth one-to-one correspondence $f: M \rightarrow N$ between the points of the manifolds with a smooth inverse. The function f is called a diffeomorphism.

We say that a manifold (defined as above) is closed or compact if it is contained in a bounded subset of \mathbf{R}^n .

In this language, the Poincaré conjecture can be stated as follows.

Conjecture (Poincaré). Any closed, simply-connected three-manifold is diffeomorphic to the three-dimensional sphere S^3 .

For a brief history of the Poincaré conjecture, see Milnor⁷.

A small region around any point in a surface can be given a pair of local coordinates. For example, away from the poles, the latitude and longitude form coordinates for any small region on the sphere. Local coordinates correspond to making a map of a region of the surface on a piece of paper in such a way that objects that are close to each other on the surface remain close on the map. One cannot make a single such map of the whole surface, but it is easy to see that one can construct an atlas of such maps. Each map is usually called a chart.

Similarly, a small region around any point in a k -manifold M can be given a system of k local coordinates

x_1, \dots, x_k . It is frequently convenient to study local properties of a manifold using these coordinates. These allow one to treat small regions of the manifold as subsets of the Euclidean space, using a chart as in the case of surfaces. Using an atlas of such charts, one can study the whole manifold.

Why the Poincaré conjecture is difficult

Both the plane and three-dimensional space are simply-connected but with an important difference. If we take a closed, embedded curve in the plane (i.e. a curve which does not cross itself), it is the boundary of an embedded disc. However, an embedded curve in three-dimensional space may be knotted (see Figure 2). This means that as we deform a knotted curve to a point, along the way it must cross itself.

Thus, an embedded curve in a simply-connected three-manifold M may not bound an embedded disc. Furthermore, such a curve may not be contained in a ball B in M . While embedded disks are useful in topology, immersed disks (i.e. disks that cross themselves) are not. It is this which makes the Poincaré conjecture difficult (in dimension 3).

The analogue of the Poincaré conjecture in dimensions 5 and above is easier than in dimension 3 for a related reason. Namely, any (two-dimensional) disc in a manifold of dimension at least 5 can be perturbed to an embedded disc, just as a curve in three-dimensional space can be perturbed so that it does not cross itself.

What made Perelman's proof and Hamilton's program, possible was the work of Thurston in the 70s, where he proposed a kind of classification of three-manifolds, the so-called geometrization conjecture⁸. Thurston's geometrization conjecture had as a special case the Poincaré conjecture, but being a statement about all three-manifolds could be approached without using the hypothesis of simple-connectivity.

However, most of the work on geometrization in the 1980s and 1990s was done by splitting into cases, so to

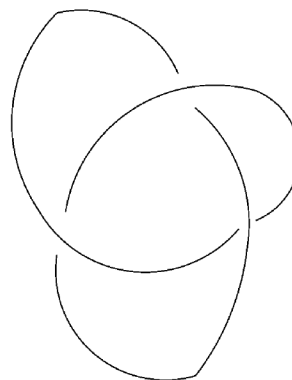


Figure 2. A knotted curve.

prove the Poincaré conjecture one was still stuck with trying to use the simple-connectivity hypothesis. An exception to this was Hamilton's program. Interestingly, Perelman found a nice way to use simple-connectivity within Hamilton's program, which simplified his proof of the Poincaré conjecture (but not of the full geometrization conjecture).

To introduce Hamilton's approach we need to reformulate the Poincaré conjecture as a statement relating topology to Riemannian geometry, namely that a compact, simply-connected three-manifold has an Einstein metric. To make sense of this we need some Riemannian geometry.

Some Riemannian geometry

Intrinsic differential geometry and curvature

In intrinsic differential geometry, we study the geometry of a space M in terms of measurements made on it. This began with the work of Gauss, who was involved in surveying large areas of land where one had to take into account the curvature of the earth. Even though the earth is embedded in three-dimensional space, the measurements we make cannot take advantage of this.

Concretely, one has to consider the question of whether we can make a map of a region of the earth on a flat surface (a piece of paper) without distorting distances (allowing all distances to be scaled by the same amount). This is impossible, as can be seen by considering the area of the region consisting of points with distance at most r from a fixed point P on the surface M . The area in case M is a sphere can be seen to be less than πr^2 , which would be the area if we did have a map that did not distort distances. In fact, for small r , the area of the corresponding region on a surface is of the form $\pi r^2(1 - \frac{K}{12}r^2 + \dots)$, and K is called the Gaussian curvature.

Intrinsic differential geometry gained new importance because of the general theory of relativity, where one studies curved space-time. Thus, we have manifolds with distances on them that do not arise from an embedding in some \mathbf{R}^n . This depended on the higher-dimensional and more sophisticated version of intrinsic differential geometry developed by Riemann. Today, intrinsic differential geometry is generally referred to as Riemannian geometry.

To study Riemannian geometry, we need to understand the analogues of the usual geometric concepts from Euclidean geometry as well as the new subtleties encountered in the more general setting. Most of the new subtleties are captured by the curvature.

Tangent spaces

Let M be a k -dimensional manifold in \mathbf{R}^n and let $p \in M$ be a point. Consider all smooth curves $\gamma: (-1, 1) \rightarrow M$ with $\gamma(0) = p$. The set of vectors $v = \gamma'(0)$ for such

curves γ gives the tangent space $T_p M$. This is a vector space of dimension k contained in \mathbf{R}^n . For example, the tangent space of a sphere with centre as the origin at a point p on the sphere consists of vectors perpendicular to p .

If a particle moves smoothly in M along the curve $\alpha(t)$, its velocity $V(t) = \alpha'(t)$ is a vector tangent to M at the point $\alpha(t)$, i.e. $V(t) \in T_{\alpha(t)} M$.

Riemannian metrics

If $\alpha: (a, b) \rightarrow \mathbf{R}^n$ is a smooth curve, then its length is given by $l(\alpha) = \int_a^b \|\alpha'(t)\| dt$. In Riemannian geometry we consider manifolds with distances that are given in a similar fashion in terms of inner products on tangent spaces.

A Riemannian metric g on M is an inner product specified on $T_p M$ for each $p \in M$. Thus, g refers to a collection of inner products, one for each $T_p M$. We further require that g varies smoothly in M . For a point $p \in M$ and vectors $V, W \in T_p M$, the inner product of V and W corresponding to the Riemannian metric g is denoted $g(V, W)$.

A Riemannian manifold (M, g) is a manifold M with a Riemannian metric g on it. Recall that near any point in M , a small region $U \subset M$ can be given a system of local coordinates x_1, \dots, x_k . If we denote the corresponding coordinate vectors by $\hat{e}_1, \dots, \hat{e}_k$, then for any point p in U the inner product on $T_p U$ is determined by the matrix $g_{ij} = g(\hat{e}_i, \hat{e}_j)$. This is a symmetric matrix.

The first examples of Riemannian manifolds are $M \subset \mathbf{R}^n$, with the inner product on $T_p M$, the restriction of the usual inner product on \mathbf{R}^n . This metric is called the metric induced from \mathbf{R}^n .

A second important class of examples are product metrics. If (M, g) and (N, h) are Riemannian manifolds, we can define their product $(M \times N, g \oplus h)$. The points of $M \times N$ consist of pairs (x, y) , with $x \in M$ and $y \in N$. The tangent space $T_{(x, y)} M \times N$ of the product consists of pairs of vectors (U, V) with $U \in T_x M$ and $V \in T_y N$. The inner product $(g \oplus h)$ is given by

$$(g \oplus h)((U, V), (U', V')) = g(U, U') + h(V, V').$$

We can identify the space of vectors of the form $(U, 0)$ (respectively $(0, V)$) with $T_x M$ (respectively $T_y N$).

Distances and isometries

Given a pair of points $p, q \in M$ in a Riemannian manifold (M, g) , the distance $d(p, q)$ between the points p and q is the minimum (more precisely the infimum) of the lengths of curves in M joining p to q .

For $p \in M$ and $b > 0$, the ball of radius r in M around p is the set of points $q \in M$ such that $d(p, q) < r$. Note that this is not in general diffeomorphic to a ball in Euclidean space.

Two Riemannian manifolds (M, g) and (N, h) are said to be isometric if there is a diffeomorphism from M to N , so that that between any pair of points in M is the same as that between their images in N . In Riemannian geometry, we regard two isometric manifolds as the same.

Geodesics and the exponential map

Geodesics are the analogues of straight lines. A straight line segment is the shortest path between its end points. A curve with constant speed that minimizes the distance between its end points is called a minimal geodesic.

More generally, a geodesic is a smooth curve with constant speed that locally minimizes distances, i.e. it is a smooth function $\gamma: (a, b) \rightarrow M$ such that $\|\gamma'(t)\|$ is constant and has the following property: for any $p = \gamma(t_0)$, there is an $\varepsilon > 0$ so that the segment of the curve γ from time $t_0 - \varepsilon$ to $t_0 + \varepsilon$ has minimal length among all curves joining $\gamma(t_0 - \varepsilon)$ to $\gamma(t_0 + \varepsilon)$.

Let $p \in M$ be a fixed point. Then we can find $r > 0$ such that if $d(p, q) < r$, then there is a unique minimal geodesic γ joining p to q . We can parametrize γ (i.e. choose the speed along γ) so that $\gamma(0) = p$ and $\gamma(1) = q$. Then the initial velocity $\gamma'(0)$ gives a vector in $T_p M$ with norm less than r . This gives a one-to-one correspondence between points q in M with $d(p, q) < r$ and vectors $V \in T_p M$ with norm less than r . The point that corresponds to the vector V is denoted by $\exp_p(V)$ and this correspondence is called the exponential map.

As an example, consider the exponential map at the north-pole of the two-sphere p . This map is one-to-one on $B_0(\pi)$ and it maps $B_0(\pi)$ to the sphere minus the south pole.

Sectional, Ricci and scalar curvatures

Let $p \in M$ be a point and let $\xi \subset T_p M$ be a two-dimensional subspace. Choose an orthonormal basis $\{U, V\}$ of ξ and consider the following family of closed curves in M :

$$C_r(\theta) = \exp(r \cos(\theta) U + r \sin(\theta) V), \theta \in [0, 2\pi].$$

It can be proved that the length of C_r has the following expansion:

$$l(C_r) = 2\pi r \left(1 - \frac{K(p, \xi)}{6} r^2 + O(r^3) \right).$$

We define the sectional curvature of (M, g) along ξ to be the number $K(p, \xi)$ above. Other notations for sectional curvature include $K_g(p, \xi)$ to clarify what metric we consider and $K(p, U, V)$ to indicate that ξ is the linear span of U and V . In the latter notation, we put $K(p, U, V) = 0$, if U and V are linearly dependent. We often omit the point p in the notation if it is clear from the context.

Averaging all the sectional curvatures at a point gives the scalar curvature $R(p)$. More precisely, let $\{E_1, \dots, E_n\}$ be an orthonormal basis of $T_p M$. Then we define

$$R(p) = \sum_{i,j} K(E_i, E_j).$$

There is an intermediate quantity, called the Ricci tensor which is fundamental in our situation. The Ricci tensor $R(U, V)$ at a point $p \in M$ depends on a pair of vectors U and V in $T_p M$. Further, it is linear in U and V and is symmetric (i.e. $\text{Ric}(U, V) = \text{Ric}(V, U)$).

If U is any unit vector in $T_p M$, then we define

$$\text{Ric}(U, U) = K(E_1, U) + \dots + K(E_n, U).$$

By linearity, for a general vector aU , with U a unit vector, $\text{Ric}(aU, aU) = a^2 \text{Ric}(U, U)$. Further, by linearity and symmetry, if U and V are any two arbitrary vectors in $T_p M$, then we put $\text{Ric}(U, V) = \frac{1}{4} (\text{Ric}(U+V, U+V) - \text{Ric}(U-V, U-V))$ (by the analogue of the formula $(a+b)^2 - (a-b)^2 = 4ab$).

Remark. It is important to note that in local coordinates these curvature quantities can be expressed in terms of g_{ij} and its first and second derivatives.

We consider some examples.

- (1) Euclidean space. This is just \mathbf{R}^n with the usual inner product. In this case, all the sectional curvatures are zero. Hence so is the Ricci tensor and the scalar curvature.
- (2) Sphere $S^n(r)$ of radius r with the metric induced from \mathbf{R}^{n+1} . In this case, all sectional curvatures are equal to r^{-2} , $\text{Ric}(U, V) = (n-1)r^{-2}g(U, V)$ and $R(p) = n(n-1)r^{-2}$ for any point p . Here $g(\cdot, \cdot)$ is (the restriction of) the standard inner product in \mathbf{R}^n .
- (3) There is an analogue of example 2, called hyperbolic space, for which the sectional curvature is $-r^{-2}$. The underlying manifold can be taken to be \mathbf{R}^n . We will not describe the metric since we will not need it.

- We have the following important converse of the above examples: Let (M, g) be a simply-connected, complete, Riemannian manifold of constant sectional curvature k . Then M is isometric to the Euclidean space, the sphere of radius $\sqrt{1/k}$ or hyperbolic space according as $k = 0$, $k > 0$ or $k < 0$ respectively.
- (4) A product Riemannian manifold $(M \times N, g = g_1 \oplus g_2)$: If ξ is a plane in $T_p(M \times N)$ that is tangent to M (respectively N), then $K(p, \xi) = K_1(\xi)$ (respectively $K_2(\xi)$). Here K_1 and K_2 denote the sectional curvatures with respect to g_1 and g_2 . On the other hand, if ξ is the span of a vector tangent to M and one tangent to N , then $K(\xi) = 0$.

- (5) As a special case of the above, consider a surface M which is the product of two circles, possibly of different radii, with the product metric. Then the tangent plane at any point is spanned by a vector tangent to the first circle and one tangent to the second circle. Hence the sectional curvature of M at any point is zero.
- (6) Another example of a product metric that we need is that on $M = S^2 \times \mathbf{R}$. In this case, the sectional curvature $K(x, \xi)$ is 1 if ξ is the tangent plane of S^2 and 0 if ξ contains the tangent space of \mathbf{R} .

Manifolds with non-negative sectional curvature

We have defined sectional curvature in terms of the growth of lengths of circles under the exponential map. In other words, sectional curvature measures the divergence of radial geodesics.

In particular, if a Riemannian manifold has non-negative curvature, geodesics do not diverge faster than in Euclidean space. This has strong consequences for the geometry and topology of these manifolds. In fact, if a simply-connected three-manifold (M, g) has non-negative sectional curvature, it has to be diffeomorphic to one of \mathbf{R}^3 , S^3 and $S^2 \times \mathbf{R}$.

Scaling and curvature

Suppose (M, g) is a Riemannian manifold and $c > 0$ is a constant. Then the sectional curvature K' of the Riemannian manifold (M, cg) is related to the sectional curvature K of (M, g) by

$$K'(p, \xi) = c^{-1}K(p, \xi)$$

for every point $p \in M$ and every tangent plane $\xi \subset T_p M$ at that point.

Note that if c is large, then K' is small. Hence, given a compact Riemannian manifold (M, g) we can always choose c large enough so that (M, cg) has sectional curvatures lying between -1 and 1 .

Einstein metrics and the Poincaré conjecture

An Einstein metric is a metric of constant Ricci curvature. More precisely, an Einstein metric with constant curvature a is one that satisfies, for all $p \in M$ and $U, V \in T_p M$, the equation

$$\text{Ric}(U, V) = ag(U, V).$$

In general relativity, one studies an action functional on the space of Riemannian metrics called the Einstein–Hilbert action, which is the integral of the scalar curvature of a metric. Einstein metrics are the critical points of this

functional among Riemannian metrics on a manifold with fixed volume.

To relate Einstein metrics to the Poincaré conjecture, one notes that an Einstein metric g on a three-manifold necessarily has constant sectional curvature (in all dimensions metrics of constant sectional curvature are Einstein metrics). Hence one concludes that if (M, g) is closed, simply-connected and Einstein, then (M, g) is isometric to S^3 with a round metric. Note that we can rule out Euclidean and hyperbolic space since they are not closed. In particular, M is diffeomorphic to S^3 .

Hence the Poincaré conjecture can be formulated as saying that any closed, simply-connected three-manifold has an Einstein metric. More generally, Thurston's geometrization conjecture says that every closed three-manifold can be decomposed into pieces in some specified way so that each piece admits a locally homogeneous metric, a concept more general than that of a metric with constant sectional curvature.

Hamilton's Ricci flow

In the 1980s and 1990s, Hamilton developed a program to prove geometrization, beginning with a paper⁹, where he showed that if a 3-manifold has a metric with positive Ricci curvature then it has an Einstein metric. By positive Ricci curvature we mean that if $p \in M$ and $U \in T_p M$ is non-zero, then $\text{Ric}(U, U) > 0$.

Hamilton's approach was to start with a given metric g and consider the one-parameter family of Riemannian metrics $g(t)$ satisfying the Ricci flow equation

$$\frac{\partial g}{\partial t} = -2\text{Ric}(t), \quad g(0) = g, \quad (1)$$

where $\text{Ric}(t)$ is the Ricci curvature of the metric $g(t)$.

To get a feeling for the analytical properties of this equation, we first consider the simpler case of the heat equation which governs the diffusion of heat in an isolated body. The heat equation is

$$\frac{\partial u}{\partial t} = \Delta u.$$

The temperature in an isolated body becomes uniform as time progresses. Further, the minimum temperature of the isolated body increases (and the maximum temperature decreases) with time. This latter property is called a maximum principle.

To see the relation of the Ricci flow with the heat equation, we use special local coordinates called harmonic coordinates (i.e. coordinates $\{x_i\}$ such that the functions x_i are harmonic). We can find such coordinates around any point in a Riemannian manifold M . In these coordinates we have

$$\text{Ric}_{ij} = -\frac{1}{2}\Delta g_{ij} + Q(g, \partial g),$$

where Q is an expression involving g and the first partial derivatives of g and $R_{ij} = \text{Ric}(\hat{e}_i, \hat{e}_j)$.

Hence the Ricci flow resembles the heat flow $\frac{\partial g_{ij}}{\partial t} = \Delta g_{ij}$ leading to the hope that the metric becomes symmetric (more precisely, the Ricci curvature becomes constant) as time progresses. However, there is an extra term $Q(g, \partial g)$ of lower order. Such a term is called the reaction term and equations of this form are known as reaction–diffusion equation. In order to understand such an equation, one needs to understand both the nature of the reaction term and conditions that govern whether the reaction or the diffusion terms dominate.

Let us consider some examples: If g is the induced metric on the sphere S^3 of radius 1, then $g(t) = (1 - 4t)g$ is the solution to eq. (1). Note that the radius of $(S^3, g(t))$ is $\sqrt{1-4t}$ and the sectional curvatures are $\frac{1}{1-4t}$. As $t \rightarrow \frac{1}{4}$, these curvatures blow-up.

More generally, if $g(t)$ is an Einstein metric the Ricci flow simply rescales the metric. In fact, if $\text{Ric} = ag$, then $g(t) = (1 - 2at)g$ satisfies eq. (1). Note that $(M, g(t))$ shrinks, expands or remains stationary depending on whether $a > 0$, $a < 0$ or $a = 0$.

On the other hand, if the metric is fixed up to rescaling by the Ricci flow, then it is an Einstein metric.

Let $(M_1 \times M_2, g_1 \oplus g_2)$ be a product Riemannian manifold. Then the Ricci flow beginning at $g_1 \oplus g_2$ is of the form $g(t) = g_1(t) \oplus g_2(t)$, where $g_1(t)$ and $g_2(t)$ are the flows on M_1 and M_2 beginning with g_1 and g_2 . Ricci flow preserves product structure. In particular, the flow beginning with the standard product metric $g_0 \oplus g_1$ on $S^2 \times \mathbf{R}$ is $g(t) = (1 - 2t)g_0 \oplus g_1$, i.e. S^2 shrinks while the \mathbf{R} direction does not change. This example is crucial for understanding regions of high curvature along the Ricci flow.

We now consider some analytical properties of the Ricci flow. One of the first results proved by Hamilton was that, given any initial metric $g(0)$ on a smooth manifold M , the Ricci flow equation has a solution on some time interval $[0, \epsilon)$. Furthermore, this solution is unique. It follows that a solution to the equation with initial metric $g(0)$ exists on some maximal interval $[0, T)$, with T either finite or infinite and is unique on this interval. Further, if T is finite then the maximum of the absolute value of the sectional curvatures becomes large as we approach T .

The main idea of Hamilton's program is to evolve an arbitrary initial metric on a closed simply-connected three-manifold along the Ricci flow and hope that the resulting metric converges, up to rescaling, to an Einstein metric. Hamilton showed that this does happen when g has positive Ricci curvature.

It is convenient to analyse separately the cases where the maximal interval of existence $[0, T)$ is finite and infinite. It turns out, as we explain later, that if the manifold

is simply-connected, then this time-interval is finite. In particular, the curvature blows-up in finite time on certain parts of the manifold.

The central issue in Hamilton's program was to understand, topologically and geometrically, parts of the manifold where curvature blows-up along the Ricci flow.

Curvature pinching

The first major steps in understanding the geometry near points of large sectional curvature were due to Hamilton and Ivey, using maximum principles.

In the simple case of a heat equation we have the maximum principle which implies that if the temperature is initially greater than a constant α at all points in the manifold, then this continues to hold for all subsequent times. In the case of the Ricci flow, we have a similar maximum principle for the scalar curvature. This is because the scalar curvature also satisfies a reaction–diffusion equation with the reaction term positive. As a consequence, the scalar curvature evolving along the Ricci flow is larger than the solution to the heat equation with the same initial conditions. In particular, we obtain the important conclusion that scalar curvature R is bounded below along the Ricci flow.

Hamilton also developed a maximum principle for tensors. Using this, Hamilton and Ivey independently obtained an inequality for the curvature using this maximum principle, which we mention and use in the next section. A consequence of the Hamilton–Ivey inequality is that if for a point $p \in M$, the maximum of the absolute values $|K(p, \xi)|$ goes to infinity, then $R(p) \rightarrow \infty$.

All these maximum principles amount to showing and using positivity properties of the reaction term.

Blow-up and convergence of Riemannian manifolds

To study points of high curvature, we use a version of a classical technique in PDEs called blow-up analysis. Namely, given a closed Riemannian manifold (M, g) , let $k_{\max} = |K(x, \xi)|$ be the maximum of the absolute values of sectional curvatures. We rescale g to $k_{\max}g$ to get a manifold with bounded sectional curvature, which is necessary for considering limiting manifolds as below.

We rescale the manifolds $(M, g(t))$ as $t \rightarrow T$ as above. This gives a sequence of manifolds with curvature uniformly bounded. One can study such a sequence by considering limiting manifolds, i.e. limits (in the sense of the next paragraph) of subsequences of the given sequence of manifolds, provided that such limiting manifolds exist.

Let (M_i, g_i) be a sequence of Riemannian n -manifolds and (N, h) be another Riemannian n -manifold. Let $x_i \in M_i$ and $x \in N$. We say that (M_i, g_i, x_i) converges to (N, h, x) if for any $\epsilon > 0$, we can find k large enough and

a diffeomorphism f from the ball B_k of radius $1/k$ in M_k to the ball B of radius $1/k$ in N with $f(x) = y$, so that for $p, q \in B_k$, $1 - \varepsilon < \frac{d(f(p), f(q))}{d(p, q)} < 1 + \varepsilon$. We shall call such a map an almost isometry. Note that our notion of limits depends on basepoints $x_i \in M_i$.

As shown by the example in Figure 3, it is necessary that the curvatures of (M_i, g_i) are bounded.

However, even a sequence of manifolds with bounded curvature need not have limiting manifolds (of the same dimension), as the manifolds may collapse to lower dimensions. For example, let $M_i = S^1 \times S^1$ be the two-torus, $g_i = i^{-1}g_0 \oplus g_0$ and $p_i = (p, q)$, where g_0 is the usual metric on the circle. Observe that (M_i, g_i) is the torus with the product metric obtained by viewing the torus as a product of a circle of radius $1/i$ with a circle of radius 1. In this case the sectional curvature of (M_i, g_i) is zero for any i . On the other hand, the limit of this sequence of metrics is the degenerate metric $0 \oplus g_0$. Hence the limit of the Riemannian manifolds (in the appropriate sense) is a circle (see Figure 4).

If no collapsing takes place, it is known that we do have limiting manifolds (which are Riemannian manifolds of the same dimension). One of the major results of Perelman was that in the situation of the blow-up limit of the previous section, i.e. when the metrics g_i arise as rescalings at certain times along the Ricci flow, collapsing does not occur.

Further, the Hamilton–Ivey pinching estimate implies that the limiting manifold is non-negatively curved. This, together with Perelman’s non-collapsing result shows that the Ricci flow for the limiting manifold is a so-called κ -solution. Perelman proved that points in a κ -solution have canonical neighbourhoods (which we explain later in the article). Furthermore, he proved a technical result giving a bound on the derivative of curvature for κ -solutions, which was crucial in understanding the behaviour near points of high (but not necessarily maximum) curvature.

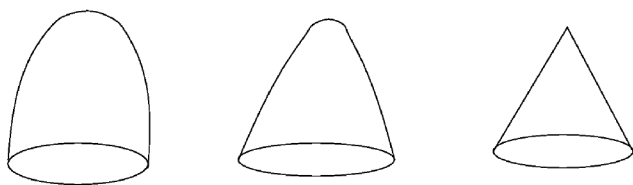


Figure 3. A sequence without bounded curvature with the limit singular.

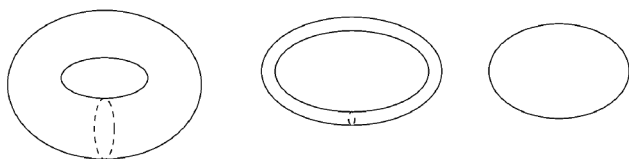


Figure 4. An example of collapsing.

Perelman’s canonical neighbourhoods

By considering limiting manifolds as above, it follows that small neighbourhoods of the points of maximum curvature are close to being ‘standard’. However, this procedure does not work if we want to understand points with high curvature which are not the maximal curvature points. The problem is that rescaling with respect to these points does not give metrics with curvature bounded independent of i .

A surprising and remarkable result of Perelman, which overcomes this difficulty and can be considered to be one of the central results in his proofs, is the canonical neighbourhood theorem. This says that if M is simply-connected, either M is diffeomorphic to S^3 or every point of high scalar curvature has a canonical neighbourhood which is an ε -neck or an ε -cap. An ε -neck is a Riemannian manifold almost isometric to the product of a sphere of radius ε and an interval of length $1/\varepsilon$. An ε -cap is diffeomorphic to a ball and satisfies certain other geometric conditions.

This result is surprising in many ways. Normally, by the kind of rescaling argument sketched above, we can study a neighbourhood of a point of maximal curvature. However, one expects that near points of high (but not maximal) curvature, there are nearby points where the curvature is much higher. This means that the curvature can be fractal-like, and the resulting system has behaviour at many scales (as happens with complex systems).

To study a neighbourhood of a point of high scalar curvature, Perelman used the bounds on the derivative of the curvature of standard solutions in an ingenious inductive argument (which proceeds by contradiction) to show that the curvature of the appropriate rescaled metric is bounded near the point. After refining this using geometric arguments (based on so-called Alexandrov spaces), Perelman showed that one can construct blow-up limits at points of high (but not necessarily maximum) curvature. Hence the results mentioned in the previous section can be used to construct canonical neighbourhoods for all points of high curvature.

Ricci flow with surgery

The canonical neighbourhood theorem allows one to understand regions where the curvature becomes large. However, if the curvature remains bounded on some region of the manifold, we cannot deduce much about the topology of the manifold. One would like to continue the Ricci flow in regions with bounded curvature, while using the canonical neighbourhood theorem to study regions with high curvature. This is accomplished by a process known as Ricci flow with surgery. This process involves modifying the manifold, geometrically and topologically, at regions of high curvature at a time close to T .

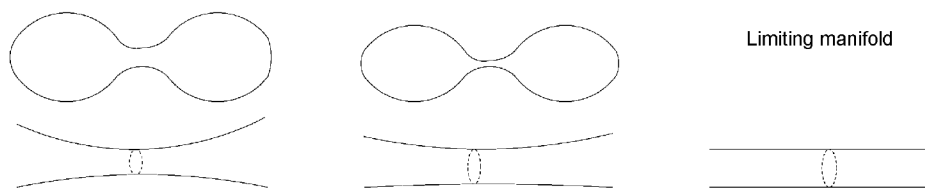


Figure 5. Sequences of manifolds and their blow-up limits.

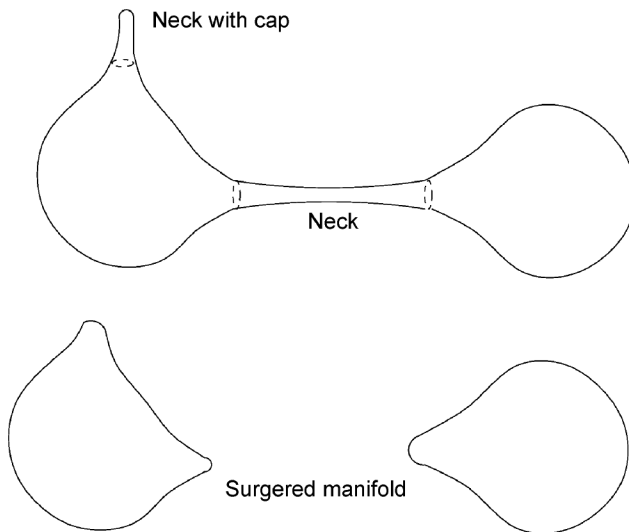


Figure 6. Surgery.

The resulting manifold has bounds on curvature that allow the process to continue beyond time T .

In case $T = \infty$, there is no need to perform surgery. Hence it suffices to consider the case where $T < \infty$.

Consider the subset Ω_ρ of M where the scalar curvature is bounded by a large number ρ for all $t \in [0, T)$, i.e. let $\Omega_\rho = \{x \in M \mid R(x, t) \leq \rho \text{ for all } t\}$. We choose ρ large enough that points of scalar curvature greater than ρ have a canonical neighbourhood.

For a time t close to T , the canonical neighbourhood theorem holds for the complement N of the interior of Ω_ρ . Thus, every point in this complement has a neighbourhood that is a neck, cap or diffeomorphic to a sphere (if the initial manifold is simply-connected). Putting these neighbourhoods together, we get either a sphere or a manifold diffeomorphic to $S^2 \times [-1, 1]$ (which is a union of several necks), which may have a cap attached at one or both ends. Topologically in each of these cases we obtain a sphere, a ball, or $S^2 \times [-1, 1]$. It follows in particular that the boundary of Ω_ρ consists of two spheres.

If Ω_ρ is empty, then however large we choose ρ , in other words if the curvature blows up on the entire manifold M , then the above implies that M is diffeomorphic to S^3 . We then replace M by the empty manifold.

Otherwise, we remove the interior of the set $N = M - \Omega_\rho$ and we attach balls to each of the boundary spheres of Ω_ρ to get a Riemannian manifold. This operation is called surgery.

Now we continue to evolve the manifold, which in general has several components by the Ricci flow. Repeating the above procedure for each of the components, we can inductively define Ricci flow with surgery (see Figure 6).

Note that if the curvature becomes high at all points in all components of M , then the manifold after surgery is empty. In this case, we say that the manifold has become extinct. One can deduce from the canonical neighbourhood theorem that in this case, the manifold just before surgery was a collection of three spheres.

We need technical results which confirm that all the properties that we have for the ordinary Ricci flow hold for Ricci flow with surgery. We also need a result indicating that in any finite time interval only finitely many surgeries are required to show that Ricci flow with surgery can be defined for all positive times. To achieve these results one needs to choose the parameter ρ carefully, in general depending on the time T .

Outline of the proof

We are now in a position to outline the proof of the Poincaré conjecture. Consider a simply-connected three-manifold M with a Riemannian metric on it. We evolve this using the Ricci flow with surgery.

A result of Perelman (for which a simpler and more elegant proof was provided by Colding and Minicozzi¹⁰) says that if the manifold M is simply-connected, then the Ricci flow with surgery becomes extinct in finite time. This is proved by considering a geometric quantity called the waist and showing that it goes to zero in finite time.

Consider Ricci flow with surgery up to the time when it becomes extinct. If we view the process backwards from the extinction time, we see that either spheres are created (the opposite of extinction) or two components are connected by a tube (the opposite of surgery). Note that when two spheres are connected by a tube, the result is still a sphere. As a result, when each surgery is viewed backwards, we see spheres either being created or merged

with other spheres. Thus at each time the manifold we see is a collection of spheres. In particular, as manifold M we started with is connected, it must have been a sphere.

Concluding remarks

The value of a mathematical theorem in science and engineering often lies not just in its statement but in the ideas that are developed in the course of proving the theorem. In this respect, Perelman's (and Hamilton's) work is rich in ideas which, when digested, may have consequences in a wide range of subjects outside mathematics. Further, techniques and ideas applicable to Ricci flow in all dimensions may be widely applicable to complex systems, while those special to dimension 3 may help us understand when a complex system is well behaved.

Remark. An article¹¹ concerning the proof of the Poincaré and geometrization conjectures has appeared recently.

1. Perelman, G., *The Entropy Formula for the Ricci Flow and its Geometric Application*, math.DG/0211159.
2. Perelman, G., *Ricci Flow with Surgery on Three-manifolds*, math.DG/0211159.
3. Perelman, G., *Finite Extinction Time for the Solutions to the Ricci Flow on Certain Three-manifolds*, math.DG/0211159.

4. Hamilton, R. S., The formation of singularities in the Ricci flow. In *Surveys in Differential Geometry*, Cambridge, MA, 1993, vol. II, pp. 7–136.
5. Morgan, J. and Tian, G., *Ricci Flow and the Poincaré Conjecture*, math.DG/0605667.
6. Kleiner, B. and Lott, J., *Notes on Perelman's Papers*, math.DG/0605667.
7. Milnor, J., Towards the Poincaré conjecture and the classification of 3-manifolds. *Not. Am. Math. Soc.*, 2003, **50**, 1226–1233.
8. Thurston, W. P., Three-dimensional manifolds, Kleinian groups and hyperbolic geometry. *Bull. Am. Math. Soc. (N.S.)*, 1982, **6**, 357–381.
9. Hamilton, R. S., Three-manifolds with positive Ricci curvature. *J. Differ. Geom.*, 1982, **17**, 255–306.
10. Colding, T. H. and Minicozzi II W. P., Estimates for the extinction time for the Ricci flow on certain 3-manifolds and a question of Perelman. *J. Am. Math. Soc.*, 2005, **18**, 561–569.
11. Cao, H.-D. and Zhu, X.-P., A complete proof of the Poincaré and geometrization conjectures – Application of the Hamilton–Perelman theory of the Ricci flow. *Asian J. Math.*, 2006, **10**, 185–492.

ACKNOWLEDGEMENTS. We thank Kalyan Mukherjee for several helpful comments that have considerably improved the exposition, and Gerard Besson for inspiring lectures on Perelman's work. We also thank Basudeb Datta, Gautham Bharali and Joseph Samuel for helpful comments.

Received 15 August 2006; accepted 25 October 2006