J. DIFFERENTIAL GEOMETRY
33 (1991) 325-334

THE RICCI FLOW ON THE 2-SPHERE

BENNETT CHOW

1. Introduction

The classical uniformization theorem, interpreted differential geometrically, states that any Riemannian metric on a 2-dimensional surface is pointwise conformal to a constant curvature metric. Thus one can consider the question of whether there is a natural evolution equation which conformally deforms any metric on a surface to a constant curvature metric. The primary interest in this question is not so much to give a new proof of the uniformization theorem, but rather to understand nonlinear parabolic equations better, especially those arising in differential geometry. A sufficiently deep understanding of parabolic equations should yield important new results in Riemannian geometry.

The question in the preceding paragraph has been studied by Richard Hamilton [3] and Brad Osgood, Ralph Phillips and Peter Sarnak [6]. In [3], Hamilton studied the following equation, which we refer to as *Hamilton's Ricci flow*

$$(*) \qquad \dot{g}(x, t) = (r - R(x, t))g(x, t), \qquad x \in M, \ t > 0,$$

where g is the metric, R is the scalar curvature of g (= twice the Gaussian curvature K), r is the average of R, and $\dot{} = \partial/\partial t$. The r in the equation above is inserted simply to preserve the area of M. He proved:

Theorem 1.1 (Hamilton). *Let (M, g) be a compact oriented Riemannian surface.*

(1) *If M is not diffeomorphic to the 2-sphere S^2, then any metric g converges to a constant curvature metric under equation $(*)$.*

(2) *If M is diffeomorphic to S^2, then any metric g with positive Gaussian curvature on S^2 converges to a metric of constant curvature under $(*)$.*

Osgood, Phillips and Sarnak have given a different proof of part (1). The object of this paper is to remove the assumption in Hamilton's theorem

Received October 11, 1988 and, in revised form, March 14, 1989. The author was supported in part by a National Science Foundation Postdoctoral Fellowship.

that a metric on S^2 have positive Gaussian curvature. We prove:

Theorem 1.2. *If g is any metric on S^2, then under Hamilton's Ricci flow, the Gaussian curvature becomes positive in finite time.*

Combining the two theorems above yields:

Corollary 1.3. *If g is any metric on a Riemann surface, then under Hamilton's Ricci flow, g converges to a metric of constant curvature.*

The proof of Hamilton's Theorem 1.1 is based in part on two remarkable estimates: a Harnack-type inequality for the scalar curvature and a decay estimate for what he calls the entropy. His proof of the Harnack inequality was inspired by the work of Peter Li and Shing-Tung Yau [4]. However, Hamilton's Harnack inequality is more delicate due to the fact that the scalar curvature satisfies a *nonlinear* equation.

The importance of studying the Ricci flow on surfaces is, as remarked by Hamilton in [3], that it may help in understanding the Ricci flow on 3-manifolds with positive scalar curvature, especially in analyzing the singularities that develop under the flow. The classification of compact 3-manifolds with positive scalar curvature is part of one possible differential geometric approach to the Poincaré conjecture.

The structure of this paper is as follows. In §2 we extend Hamilton's proof of the Harnack inequality to the case where R may change sign on M. However, our proof of the entropy estimate differs significantly from Hamilton's. We observe that an upper bound for the entropy follows from the fact that the energy associated to Hamilton's Ricci flow is bounded from below. This is explained in §§3 and 4. Finally, in §5, we state a theorem of Hamilton giving a lower bound for the injectivity radius of metrics evolving under ($*$). The proofs in this section are due to him.

We remark that Corollary 1.3 extends, in a rather straightforward manner, to the case where M has geodesic boundary when g satisfies Hamilton's Ricci flow with Neumann boundary conditions.

I am particularly grateful to Richard Hamilton for his generous help and encouragement. I would like to thank Lang-Fang Wu for helpful discussions, and also to thank my parents Yutze and Wanlin Chow for providing a wonderful environment during the course of the research this past summer.

2. The Harnack inequality

For a positive solution to a linear parabolic equation, the Harnack inequality gives an upper bound for the ratio of the solution at two different points in space-time. Under Hamilton's Ricci flow, the scalar curvature

R satisfies a nonlinear parabolic equation. Despite this fact, Hamilton [3] derived a Harnack inequality for R assuming that it is initially positive.

The object of this section is to prove a Harnack inequality for the quantity $R - s$ which is a modification of Theorem 6.3 of [3]. The function $s = s(t)$ is defined to satisfy the ordinary differential equation: $\dot{s} = s(s-r)$, with $s(0) < \min_{x \in M} R(x, 0)$. Note that $s(t) = r/(1 - ce^{rt}) < 0$, where $c > 1$. The motivation for this definition is as follows. The partial differential equation satisfied by the scalar curvature R is given in [3] as

$$\dot{R} = \Delta R + R(R - r).$$

The ordinary differential equation for s is obtained from the partial differential equation for R simply by dropping the Laplacian term. Following [3, Definition 6.2], we let

$$\Delta = \Delta(x_1, t_1, x_2, t_2) = \inf_\gamma \int_{t_1}^{t_2} |\dot{\gamma}(t)|^2 \, dt,$$

where the infimum is taken over all paths γ in M whose graphs $(\gamma(t), t)$ join (x_1, t_1) and (x_2, t_2). The rest of this section consists of proving the following Harnack inequality:

Lemma 2.1. *There are constants $t_0 > 0$ and C such that for any $x_1, x_2 \in M$ and $t_2 > t_1 \geq t_0$*

$$R(x_2, t_2) - s(t_2) \geq e^{-\Delta/4 - C(t_2 - t_1)}(R(x_1, t_1) - s(t_1)).$$

We first recall that Hamilton [3] proved that a solution to $(*)$ exists for all time. The difficulty is showing that the metric converges as t approaches infinity and that the limiting metric has constant curvature. From the equations for R and s we compute that

$$\tfrac{\partial}{\partial t}(R - s) = \Delta(R - s) + (R - s)(R - r + s).$$

By the maximum principle, it follows that since $R - s$ is positive initially, it stays positive for all time. Therefore the quantity $L = \log(R - s)$ is well equally defined. We compute that

$$\dot{L} = \Delta L + |\nabla L|^2 + R - r + s.$$

Let $Q = \dot{L} - |\nabla L|^2 - s = \Delta L + R - r$. We then have the following crucial estimate for Q.

Sublemma 2.2. *There exist constants t_0 and C such that for $t > t_0$ and $x \in M$*

$$Q(x, t) \geq -C.$$

Proof of Lemma 2.1. Let $x_1, x_2 \in M$, $t_2 > t_1$, and γ be any curve whose graph joins (x_1, t_1) and (x_2, t_2). Then

$$L(x_2, t_2) - L(x_1, t_1) = \int_{t_1}^{t_2} \frac{d}{dt} L(\gamma(t), t)\, dt = \int_{t_1}^{t_2} \left(\frac{dL}{dt} + \langle \nabla L, \dot{\gamma} \rangle \right) dt.$$

From Sublemma 2.2, we have $\dot{L} - |\nabla L|^2 \geq -C + s \geq -C - 1$. Therefore, as in [3, 6.3], we conclude that

$$L(x_2, t_2) - L(x_1, t_1) \geq -C(t_2 - t_1) - \Delta/4,$$

which completes the proof of Lemma 2.1.

Proof of Sublemma 2.2. Following [3, 6.3], we compute the evolution equation for Q:

$$\begin{aligned}
\dot{Q} &= \Delta \dot{L} + \dot{\Delta} L + \dot{R} \\
&= \Delta(\Delta L + |\nabla L|^2 + R) + (R - r)\Delta L + \Delta R + R(R - r) \\
&= \Delta Q + 2|\nabla^2 L|^2 + 2\langle \nabla L, \nabla \Delta L \rangle + R|\nabla L|^2 + (R - r)\Delta L \\
&\quad + (R - s)(\Delta L + |\nabla L|^2) + R(R - r) \\
&= \Delta Q + 2\langle \nabla L, \nabla Q \rangle + 2|\nabla^2 L|^2 + 2(R - r)\Delta L + (R - r)^2 \\
&\quad + s|\nabla L|^2 + (r - s)\Delta L + r(R - r).
\end{aligned}$$

Therefore we obtain the following formula for $\dot{Q}$:

$$\begin{aligned}
\dot{Q} &= \Delta Q + 2\langle \nabla L, \nabla Q \rangle + 2|\nabla\nabla L + \tfrac{1}{2}(R - r)g|^2 + (r - s)Q \\
&\quad + s|\nabla L|^2 + s(R - r).
\end{aligned}$$

Recall that by [3, 4.6], we have the upper bound $R \leq Ce^{rt}$ so that $sR \geq -C$. Thus

$$\dot{Q} \geq \Delta Q + 2\langle \nabla L, \nabla Q \rangle + Q^2 + (r - s)Q + s|\nabla L|^2 - C,$$

where we have used the inequality

$$|\nabla\nabla L + \tfrac{1}{2}(R - r)g|^2 \geq \tfrac{1}{2}(\Delta L + R - r)^2 = \tfrac{1}{2}Q^2.$$

Since we are trying to show that Q is increasing in time, the bad (i.e., negative) term in the line above is $s|\nabla L|^2$. However, we have

$$\frac{\partial}{\partial t}(sL) = \Delta(sL) + s|\nabla L|^2 + s(R - r + s) + s(s - r)L.$$

Noting that $L \geq -C - Ct$, we obtain

$$\frac{\partial}{\partial t}(sL) \geq \Delta(sL) + 2\langle \nabla L, \nabla(sL) \rangle - s|\nabla L|^2 - C.$$

This leads us to define $P = Q + sL$. Clearly

$$\dot{P} \geq \Delta P + 2\langle \nabla L, \nabla P \rangle + Q^2 + (r - s)Q - C.$$

Since sL is bounded, there exists a constant $C > 0$ such that for t large enough

$$\dot{P} \geq \Delta P + 2\langle \nabla L, \nabla P \rangle + \tfrac{1}{2}(P^2 - C^2).$$

Applying the maximum principle yields

$$P \geq C \frac{1 + ce^{Ct}}{1 - ce^{Ct}},$$

where $c > 1$. Hence, for t large enough, $P \geq -2C$; likewise $Q \geq -3C$. This completes the proof of Sublemma 2.2.

3. The energy functional and convexity

Let h be a fixed metric on M conformal to, and with the same volume as, the family of metrics $g(t)$. For any metric g with the same volume as h, we define the energy of g (relative to h) to be

$$E(g) = \int_M \log(g/h)(R_g dA_g + R_h dA_h),$$

where dA denotes the area form. We remark that the energy is, up to a negative constant, the logarithm of the *determinant* of the Laplacian of g (see [6]). The term $\log(g/h)$ is actually the first Bott-Chern form, and the energy above has a generalization to unitary connections on holomorphic vector bundles over Kähler manifolds [2]. The gradient of E at the metric g, computed with respect to the induced metric on the space of symmetric 2-tensors, is given by

$$\nabla E = (R_g - r)g.$$

If we write $g = e^u h$, where u is a function on M, then E takes the form

$$E(u) = \int_M (|\nabla u|_h^2 + 2R_h u)\, dA_h,$$

where we have used the formula

$$R_g = e^{-u}(-\Delta_h u + R_h).$$

In the space of *all* metrics conformal to h, the second variation of E is given by

$$\frac{d^2}{dt^2} E(u + tv) = \int_M |\nabla v|_h^2\, dA_h \geq 0,$$

where $v \in C^{\infty}(M)$. However, on the space of metrics with fixed volume, E is not a convex functional. This is because the volume constraint is nonlinear. In spite of this, E. Onofri proved that the energy functional is bounded from below on S^2 [5], [6].

Lemma 3.1. *On the space of all metrics in a conformal class on S^2, the energy functional E is bounded from below. Moreover, the minimum of E is obtained by the 6-dimensional family of constant curvature metrics in the conformal class.*

Now let g be a solution to Hamilton's Ricci flow. We then compute

$$\dot{E} = -2\int_M (R_g - r)^2 \, dA_g.$$

Thus Lemma 3.1 implies the following:

Corollary 3.2. *There exists a constant C independent of time such that, under Hamilton's Ricci flow,*

$$\int_0^{\infty} \left\{ \int_{M_t} (R_g - r)^2 \, dA_g \right\} dt \le C.$$

4. The entropy estimate

In [3, 7.2], Hamilton proved that $\int_M R \log R \, dA$ is decreasing under $(*)$ provided $R > 0$. His argument involved showing that its time derivative $Z = \frac{d}{dt}\int_M R \log R \, dA$ satisfies the following ordinary differential inequality: $\dot{Z} \ge Z^2 + rZ$. He then reasoned that if Z ever became positive, it would blow up in a finite amount of time, a contradiction to the fact that the solution to $(*)$ exists for all time. Unfortunately his arguments do not seem to extend to the case where R is not always positive on M. Instead, in this section, we observe that the results of the last section imply an upper bound for the entropy.

Lemma 4.1. *There exists a constant C independent of time such that, under Hamilton's Ricci flow,*

$$\int_M (R - s) \log(R - s) \, dA \le C.$$

Proof. First, we recall that $\dot{s} = s(s - r)$. Let $v = 1/(r - s)$ so that $\dot{v} = -vs$. We compute that $\frac{\partial}{\partial t}(R \, dA) = \Delta R \, dA$ and $\frac{\partial}{\partial t}(s \, dA) = s(s - r) \, dA$. Therefore

$$\frac{\partial}{\partial t}\{v(R - s) \, dA\} = v \Delta E \, dA.$$

Thus

$$\begin{aligned}\frac{\partial}{\partial t}&\left\{v\int_M (R-s)\log(R-s)\,dA\right\}\\&=-v\int_M \frac{|\nabla R|^2}{R-s}\,dA+v\int_M (R-s)(R-r+s)\,dA\\&\le v\int_M (R-r)^2\,dA.\end{aligned}$$

Integrating the above formula, we obtain

$$\int_M (R-s)\log(R-s)\,dA\le C\int_0^\infty\left\{\int_{M_t}(R-r)^2\,dA\right\}dt+C.$$

The lemma now follows from Corollary 3.2.

5. A lower bound for the injectivity radius

As stated in the introduction, the proofs and results of this section are due to Richard Hamilton. The object of this section is to get a lower bound for the injectivity radius of M under Hamilton's Ricci flow. We need this in order to combine the Harnack inequality with the entropy estimate to get a time-independent supremum bound for the scalar curvature R. The injectivity radius lower bound implies a bound for the diameter of M. This is needed to obtain a positive lower bound for $R-s$. As a direct consequence, R eventually becomes positive under Hamilton's Ricci flow. The rest of this section consists of proving the following.

Theorem 5.1. (*Hamilton*). *If the metric g on M is flowing under* ($*$), *then for any $t>0$*

$$\imath(M_t)\ge\min\left\{\imath(M_0),\ \min_{s\in[0,t]}\frac{\pi}{\sqrt{K_{\max}(s)}}\right\},$$

where $\imath(M)$ denotes the injectivity radius of M.

Recall that a geodesic is stable if the second variation of arc length is nonnegative. We have that short geodesics are stable:

Lemma 5.2. *Let γ be the shortest closed geodesic on M. If γ has length $l(\gamma)<2\pi/\sqrt{K_{\max}}$, then γ is stable.*

Proof. If γ is unstable, then there exists a curve γ^* near γ with $l(\gamma^*)<l(\gamma)$. Pick two points p and q on γ that divide γ into two segments γ_1 and γ_2 of equal length. Choose two points p^* and q^* on γ^* near p and q, respectively, which divide γ^* into γ_1^* and γ_2^*. Since $l(\gamma_i)<\pi/\sqrt{K_{\max}}$, for $i=1,2$, there exist unique geodesics β_1^* and

β_2^*, connecting p^* and q^*, which are close to γ_1^* and γ_2^*, respectively. Moreover, as a consequence of the Gauss lemma [1, p. 8], $l(\beta_i^*) \le l(\gamma_i^*)$ for $i = 1, 2$.

Claim 5.3. There exists a smooth closed geodesic μ on M with

$$l(\mu) \le l(\beta_1^*) + l(\beta_2^*).$$

The claim entails the lemma since then: $l(\mu) \le l(\gamma^*) < l(\gamma)$, which is a contradiction.

Proof of Claim 5.3. Let Γ be the space of all nondegenerate geodesic 2-gons in M with side lengths less than $\pi/\sqrt{K_{\max}}$; that is,

$$\Gamma = \{\alpha = (\alpha_1, \alpha_2) | \alpha_1 \text{ and } \alpha_2 \text{ are distinct smooth geodesic segments joining the same endpoints, with } l(\alpha_i) < \pi/\sqrt{K_{\max}}, \text{ for } i = 1, 2\}.$$

We remark that Γ is an open 4-manifold locally parametrized by points in $M \times M$. Define $m = \min_{\alpha\in\Gamma}\{l(\alpha_1) + l(\alpha_2)\}$. Since the exponential map of M is a diffeomorphism when restricted to any ball of radius strictly less then $\iota(M)$, we have $m \ge \iota(M) > 0$. Now let μ be a geodesic 2-gon with $l(\mu) = m$. If μ is not smooth, then by the first variation formula [1, p. 5], we could shorten μ inside Γ. Therefore μ is a smooth geodesic loop. This proves the claim and hence also Lemma 5.2.

Lemma 5.4. *If γ is a geodesic loop in M_t as in Lemma* 5.2, *then under the Hamilton's Ricci flow, $\frac{d}{dt}\{l(\gamma)\} \ge r \cdot l(\gamma)$, at time t.*

Proof. Let T and N be unit tangent and normal vector fields to γ, respectively. Since γ is a geodesic, we have $\nabla_T N = 0$. Therefore, by the second variational formula [1, p. 21], we have

$$\frac{\partial^2}{\partial N^2}\{l(\gamma)\} = \int_\gamma \langle R(N, T)N, T\rangle = -\int_\gamma K,$$

where $K = R/2$ is the Gaussian curvature of M. Since γ is stable, we have $\int_\gamma K \le 0$. Therefore

$$\frac{d}{dt} l(\gamma) = \frac{1}{2}\int_\gamma (r - R) \ge \int_\gamma \frac{r}{2} = \frac{r}{2} \cdot l(\gamma).$$

Lemma 5.5. *Suppose γ_t is (one of) the shortest closed geodesic(s) on M_t. If $l_t(\gamma_t) < 2\pi/\sqrt{K_{\max}}$, then for $\varepsilon > 0$ small enough, there exists a geodesic $\gamma_{t-\varepsilon}$ on $M_{t-\varepsilon}$ with $l_{t-\varepsilon}(\gamma_{t-\varepsilon}) < l_t(\gamma_t)$.*

Proof. By Lemma 5.4, we have for ε small enough, $l_{t-\varepsilon}(\gamma_t) < l_t(\gamma_t)$. Choose points p and q on γ_t, which divide γ_t into two segments γ_1

and γ_2 of equal length with respect to the metric $g_{t-\varepsilon}$. Since $l_{t-\varepsilon}(\gamma_i) < \pi/\sqrt{K_{\max}(t-\varepsilon)}$ for ε small enough, there exist unique geodesics β_1 and β_2 joining p and q, with respect to the metric $g_{t-\varepsilon}$, which are near γ_1 and γ_2, respectively. Now $l_{t-\varepsilon}(\beta_i) \le l_{t-\varepsilon}(\gamma_i) < l_t(\gamma_i)$ so that $l_{t-\varepsilon}(\beta_1 \cup \beta_2) < l_t(\gamma_t)$. By Claim 5.3, there is a smooth closed geodesic μ on $M_{t-\varepsilon}$ with $l_{t-\varepsilon}(\mu) \le l_{t-\varepsilon}(\beta_1 \cup \beta_2) < l_t(\gamma_t)$. This completes the proof of Lemma 5.5.

To conclude this section we complete the proof of the theorem.

Proof of Theorem 5.1. Lemma 5.5 states that, under Hamilton's Ricci flow, the length of the shortest closed geodesic is increasing provided its length is less than $2\pi/\sqrt{K_{\max}}$. The theorem now follows from the fact that

$$\imath(M) \ge \min\{\pi/\sqrt{K_{\max}},\, l(\text{shortest closed geodesic})/2\}.$$

6. The scalar curvature becomes positive

In this section we apply the results of the preceding sections to show that, under Hamilton's Ricci flow on S^2, the scalar curvature R becomes positive in finite time.

Proof of Theorem 1.2. Clearly, for any $t > 0$,

$$K_{\max}(t) \le \max_{s\in[0,t]} K_{\max}(s).$$

Now choose $\tau \in [0, t]$ so that $K_{\max}(\tau) = \max_{s\in[0,t]} K_{\max}(s)$. By Theorem 5.1, $\imath(M_\tau) \ge \min\{\imath(M_0),\, \pi/\sqrt{K_{\max}(t)}\}$. As in §8 of [3], by combining the Harnack inequality for $R - s$, the entropy estimate, and the injectivity radius lower bound of the last section, we obtain $R_{\max}(\tau) \le C$, where C is a constant independent of t. Therefore $R_{\max}(t) \le C$ independent of t, and the injectivity radius of M is bounded from below independent of time. This implies that the diameter of M is also bounded. Thus, as in §8 of [3], the Harnack inequality also shows that $R - s \ge c > 0$. Since s approaches zero exponentially, we conclude that R becomes positive in finite time.

References

[1] J. Cheeger & D. G. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland, Amsterdam, 1975.

[2] S. K. Donaldson, *Anti self-dual Yang-Mills connections over complex algebraic surfaces and stable vector bundles*, Proc. London Math. Soc. **50** (1985) 1–26.

[3] R. S. Hamilton, *The Ricci flow on surfaces*, Math. and General Relativity, Contemporary Math. **71** (1988) 237–262.

[4] P. Li & S. T. Yau, *On the parabolic kernel of the Schrödinger operator*, Acta Math. **156** (1986) 153–201.

[5] E. Onofri, *On the positivity of the effective action in a theory of random surfaces*, Comm. Math. Phys. **86** (1982) 321–326.

[6] B. Osgood, R. Phillips & P. Sarnak, *Extremals of determinants of Laplacians*, J. Funct. Anal. **80** (1988) 148–211.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY