

REMARKS ON PERELMAN'S PAPERS

MICHAEL T. ANDERSON

This note is informal commentary, (very informal in comparison with the Kleiner-Lott notes [KL]), on Grisha Perelman's work [I], [II] on Ricci flow and geometrization of 3-manifolds. The comments concern issues or questions that either arise from my own thoughts, or in response to those raised by others. They are also influenced by Grisha's lectures at Stony Brook in April, 03.

By and large, these comments do not address the details of the proofs in [I] or [II]; for [I], this has already been carried out wonderfully by Kleiner-Lott. Instead, these notes basically just record some of my thoughts and views on the papers at this time. It is hoped that this is of some use to other non-experts on the Ricci flow, partly as a guide in understanding some of the main issues. I hope to add further remarks and discussion as time goes on. Comments and criticism are welcome.

§1. Comments on \mathcal{F} .

The stationary points of the Ricci flow on the space of metrics \mathbb{M} on a given manifold are the Ricci-flat metrics, or more generally Einstein metrics for the volume-normalized Ricci flow. If the (volume normalized) Ricci flow is the gradient flow of some functional, then the functional must have critical points exactly the class of Einstein metrics. The only known functional on \mathbb{M} with this property is the Einstein-Hilbert action $\mathcal{S}_{EH} = \int (R - 2\Lambda)dV$ or total scalar curvature functional $\mathcal{S}_{tot} = v^{-(n-2)/n} \int R dV$. However, the Ricci flow is not the gradient flow of (any such) \mathcal{S} ; the gradient flow of \mathcal{S} does not exist, since it implies a backward heat equation for the evolution of the scalar curvature R .

Since early work of Hamilton, it has been recognized that Ricci solitons are important in the study of the Ricci flow, in particular regarding issues related to singularity formation. (When a sequence of metrics is rescaled by a divergent sequence of factors, the sequence must be pulled back by large, local diffeomorphisms to obtain any limit). These solutions evolve by a flow by diffeomorphisms: $g(t) = \phi_t^* g(0)$, so that $\frac{d}{dt}g(t) = -2Ric = \delta^* X$, where X is the infinitesimal generator of ϕ_t . Gradient solitons, where $X = \nabla f$, then satisfy the equation $Ric + D^2 f = 0$, for some function f .

The Ricci flow is invariant under the action of the diffeomorphism group \mathcal{D} on \mathbb{M} and so descends to a flow on the moduli space $\mathcal{M} = \mathbb{M}/\mathcal{D}$. The stationary points of the Ricci flow on \mathcal{M} are then the (equivalence classes of) Ricci-flat metrics and Ricci solitons. On the other hand, it is well-known, cf. [I, 2.4,2.5], that there are no non-trivial, i.e. non-Einstein, Ricci solitons on *closed* n -manifolds; soliton here means steady soliton.

One can then ask if there is a functional on \mathbb{M} whose critical points are Ricci-flat metrics together with Ricci solitons. This now involves extra data, namely metrics and vector fields, or metrics and functions in the case of gradient solitons. This is exactly what \mathcal{F}^m does, cf. [I, (1.2)].

The functional \mathcal{F} is initially defined as a functional on the product $\mathbb{M} \times C^\infty(M, \mathbb{R})$. However, the idea is to couple these two factors. $C^\infty(M, \mathbb{R})$ is identified with the space of volume forms by $f \rightarrow e^{-f} dV$ where dV is any fixed volume form. Under the Ricci flow, the volume form dV_{g_t} changes. (Modifying it to leave the volume form fixed leads to a backwards heat equation for R). Given then any smooth measure dm , Perelman considers the functional \mathcal{F}^m , where $dm = e^{-f} dV_g$ couples f and g . \mathcal{F}^m is then a functional on all of \mathbb{M} ; given any g , there is a unique f such that

Partially supported by NSF Grants DMS 0305865.

$e^{-f}dV_g = dm$. The L^2 gradient flow for \mathcal{F}^m is $\tilde{g}_t = -2(\text{Ric} + D^2f)$, differing from the Ricci flow just by diffeomorphisms.

The functional \mathcal{F} is invariant under \mathcal{D} , acting diagonally by pullback on (g, f) . Each \mathcal{F}^m is invariant under the diffeomorphisms \mathcal{D}_0 that preserve the measure dm . The remaining part of the diffeomorphism group $\mathcal{D}/\mathcal{D}_0$ is essentially the gradient-like diffeomorphisms (generated by gradient vector fields), thus corresponding to $C^\infty(M, \mathbb{R})$, i.e. the different smooth measures dm . Thus one has a family of functionals \mathcal{F}^m on \mathbb{M} , parametrized by the functions $C^\infty(M, \mathbb{R})$. The symmetry group of \mathcal{F} has been reduced from \mathcal{D} to \mathcal{D}_0 and the remaining part of the action of \mathcal{D} on \mathbb{M} , $\mathcal{D}/\mathcal{D}_0$, has been decoupled into a space of parameters.

The equation for the evolution for f , $f_t = -R - \Delta f$, is a backward heat equation for f . This is natural: f must balance the forward evolution of the volume form of $g(t)$. Thus, this flow will not exist for most f , starting at $t = 0$ and going forward in time. However, one of the basic points of view is to let the (pure) Ricci flow flow for a time $t_0 > 0$. At t_0 , one may then take an arbitrary f , and flow this f backward in time, (forward in $\tau = t_0 - t$) to obtain an initial value $f(0)$ for f . The choice of f at time t_0 determines, together with the volume form of $g(0)$, or $g(t_0)$, the measure dm and so choice of \mathcal{F}^m .

The evolution equation for f may be recast in the form $\square^*u = 0$, where $u = e^{-f}$ and \square^* is the conjugate or adjoint to \square . This is a very natural scalar evolution equation, balancing the evolution of $g(t)$. It is hard to imagine anything more natural.

Compare this situation with that of either scalar curvature functional \mathcal{S} . The gradient flow of \mathcal{S} does not exist, since it gives a backwards evolution equation for the scalar curvature R - the trace part of the curvature evolution. However, the trace-free part of the evolution of curvature, i.e. the evolution of Ric_0 , is a forward evolution equation; see the Appendix. For this flow, since the trace and trace-free parts are coupled together, the flow does not exist. The Perelman flow thus uncouples this dependence.

It is quite remarkable what this buys. The freedom in the choice of f (at later times) allows one to use f to “probe the geometry” of $(M, g(t))$; one can use suitable choices of f to detect local geometric behavior of $g(t)$. The first main example of this is [I, Thm 4.1] - the first non-collapse theorem where e^{-f} is a suitable approximation to a delta function in a collapsed region (under the assumption that such exists). At a given $g(t)$, all the data in \mathcal{F}^m (or \mathcal{W}) are fixed except for the choice of f , (or measure). Since this is freely chosen, \mathcal{F} is basically the Dirichlet energy of f , (on small scales) and its very simple to relate this to the local collapse/non-collapse behavior of $g(t)$.

Each functional \mathcal{F}^m is neither bounded above nor below (as with \mathcal{S}), for example since its not scale invariant. Flow lines of the gradient flow will in general only exist for a finite time - at most as long the time existence of the Ricci flow. Thus, most flow lines (all flow lines on a complicated manifold) escape to infinity in \mathbb{M} in finite time. On such flow lines \mathcal{F}^m may or may not remain bounded.

Observe that the proof of [I, Prop. 1.2] is basically identical to the time estimate for positive scalar curvature for the Ricci flow: $R(0) \geq a > 0 \Rightarrow t \leq n/2a$. Whereas this estimate uses the maximum principle, the time estimate for \mathcal{F}^m involves integration.

§2. Comments on λ .

The proof that λ increases under the Ricci flow is simple, given the right perspective. At time t , so at $g(t)$, choose $f = f(t)$ so that $e^{-f/2}$ is the lowest eigenfunction of \mathcal{F} ; thus the measure is no longer fixed, just its total mass (or volume) is normalized to 1. For such an $f(t)$, $\mathcal{F}(g(t), f(t)) = \lambda(t)$. Now evolve $f(t)$ back along Ricci flow to $t - \varepsilon$. \mathcal{F} or \mathcal{F}^m , for the induced dm , decreases to the past, so $\mathcal{F}(g(t - \varepsilon), f(t - \varepsilon)) \leq \mathcal{F}(g(t), f(t))$. Then $\lambda(t - \varepsilon) = \inf_f \mathcal{F}(g(t - \varepsilon), f(t - \varepsilon)) \leq \mathcal{F}(g(t - \varepsilon), f(t - \varepsilon))$, which proves the result.

The Ricci flow mod diffeos can be considered as the L^2 gradient flow of \mathcal{F}^m , for any dm . Since λ is invariant under the action of \mathcal{D} on \mathbb{M} , the Ricci flow mod diffeos can also be considered as the gradient flow for $\lambda : \mathbb{M} \rightarrow \mathbb{R}$. Note that λ is global invariant attached to g .

It is interesting to compare the situation here with the usual treatment of the total scalar curvature \mathcal{S}_{tot} ; there are several formal similarities, but also major differences. \mathcal{S} is invariant under all diffeomorphisms, (as is \mathcal{F} without the constraint). In seeking Einstein metrics via the well-known minimax procedure, the diffeomorphism invariance is completely broken by first restricting to conformal classes, and solving the Yamabe problem. Similarly, the Perelman coupling of \mathcal{F} to \mathcal{F}^m breaks the symmetry from \mathcal{D} to \mathcal{D}_0 . In the \mathcal{S} context, the function f conformally changes the metric, instead of changing the choice of measure. Minimizing \mathcal{F} for a given g corresponds to minimizing \mathcal{S} in a given conformal class $[g]$, so that the eigenvalue λ formally corresponds to the scalar curvature of a Yamabe metric. In fact, observe that if g is a metric of constant scalar curvature R_g , then $\lambda = R_g$, (and $f = 0$).

The variational problem for λ closely resembles the variational problem for \mathcal{S} in a given conformal class $[g]$ - i.e. the Yamabe problem. The latter is minimizing

$$\int Rh^2 + 4\frac{n-1}{n-2}|\nabla h|^2 dV,$$

subject to the constraint that the $L^{2n/n-2}$ norm of h equals 1. This has the same form as the variational problem for (\mathcal{F}, λ) with $h = e^{-f/2}$ but with constraint that the L^2 norm of h equals 1. The difference in the constraints is of course important, but the L^2 eigenvalue problem associated to the Yamabe problem

$$-4\frac{n-1}{n-2}\Delta h + Rh = \lambda h,$$

has been studied for a long time, cf. [B, Ch.4]. The difference in the factors $4\frac{n-1}{n-2}$ and 4 before $|\nabla h|^2$ turns out to be significant. Only the coefficient 4 before $|\nabla h|^2$ in \mathcal{F} gives a simple or natural evolution equation - the modified Ricci flow. There is a cancellation in the derivation of the Euler-Lagrange equations between the R and $|\nabla f|^2$ factors that occurs only when their coefficients are equal. For unequal coefficients, the EL equation is more complicated and less natural.

From a certain perspective, the Yamabe problem is not natural, since it is not a diffeomorphism invariant problem - a conformal class $[g]$ is not invariant under diffeomorphisms. The space of minima, i.e. the space of Yamabe metrics \mathcal{Y} , is invariant under diffeos, and all the work takes place on \mathcal{Y} , (with regard to finding Einstein metrics). The condition that g is Yamabe is a global condition on g . In a certain sense, \mathcal{Y} corresponds to one “preferred measure”, and all the other “measures” are ignored.

Modulo finite dimensional subspaces, the functional \mathcal{S} has negative Hessian on \mathcal{Y} , transverse to the diffeomorphisms, but positive Hessian on $[g]$. This min/max mixture causes a lot of headaches. In a certain sense (that I don’t see how to formulate exactly in a correct way), the evolution equation for the L^2 gradient of \mathcal{S} is backwards on $[g]$, and forwards on the transverse space \mathcal{C} of conformal classes - corresponding to the backward evolution of f and forward evolution of g for the gradient flow of \mathcal{F}^m . On the other hand, each functional \mathcal{F}^m has essentially negative Hessian on the part of $T\mathbb{M}$ orthogonal to the diffeomorphisms \mathcal{D} ; see the Appendix.

§3. Comments on Section 3.

The same remarks as above motivating \mathcal{F} also motivate \mathcal{W} , a functional whose critical points are homothetic gradient solitons, and so in particular Einstein metrics. The exact form is motivated by scale invariance and the form of the heat kernel. Further motivation is given in §5 regarding entropy. It would be very interesting to explore this further. The “explicit insertion of the scale parameter τ ” is very interesting and natural, relating the time evolution with the scale of the metric. Note

that physicists never work with the scale-invariant total scalar curvature \mathcal{S}_{tot} ; physically and in other respects, this is an unnatural “Lagrangian”, due to the global volume factor.

If T is the (first) singular time of the Ricci flow on $[0, T)$, the functional \mathcal{W} may or may not remain bounded on $g(t)$ as $t \rightarrow T$. If it remains bounded, then by scale-invariance, it remains bounded on any rescaling or blow-up limit modelling the singularity, e.g. an ancient solution. At the end of [I, §, 5.1], Perelman raises the question whether the limit is a gradient shrinking soliton if this is the case.

As far as I know, it is an open question whether blow-up limits of the Ricci flow are necessarily Ricci solitons.

§4. Comments on Section 4.

The simplest notion of collapse or non-collapse is w.r.t. volume. A ball $B_x(r)$ is ν -collapsed if $\text{vol } B_x(r) \leq \nu r^n$, and ν -noncollapsed if $\text{vol } B_x(r) \geq \nu r^n$. Thus \mathbb{R}^n with flat metric is ω_n -noncollapsed globally, (for all balls of all radii), while the flat product $\mathbb{R}^{n-1} \times S^1$ is ω_n -noncollapsed for small balls, but highly collapsed for large balls; $\omega_n = \text{volume of unit ball in } \mathbb{R}^n$.

Definitions [I, 4.1, 4.2 and 8.1] are crucial and differ subtly from the definition above. Given a Riemannian manifold (M, g) and point $x \in M$, define the (L^∞) curvature radius ρ to be the radius of the largest ball centered at x such that

$$|Rm|(y) \leq \rho^{-2}, \quad (1)$$

for all $y \in B_x(\rho)$. It is important to note that this radius scales as a distance function. Thus, when one rescales the metric by ρ^{-2} so that $B_x(\rho)$ becomes $B_x(1)$, the curvature of the rescaled metric has L^∞ norm at most 1 in $B_x(1)$. One then says that the metric g is κ -noncollapsed at x on the scale of the curvature radius if $\text{vol } B_x(1) \geq \kappa$. The volume comparison theorem then implies that all smaller balls in $B_x(1)$ are κ' -noncollapsed, $\kappa' = \kappa'(\kappa)$.

The same notions of collapse/noncollapse on the scale of ρ appear in my work on geometrization, cf. [A1,2] for instance.

For flat metrics on \mathbb{R}^n or its quotients, κ -collapsed or κ -noncollapsed is the same as the usual definition, since $\rho = \infty$. The spherical cylinder $S^{n-1} \times \mathbb{R}$ is κ -noncollapsed everywhere, (i.e. on scales no larger than the curvature radius), but is highly collapsed on large scales.

Suppose a complete, noncompact, non-flat manifold (N, g) has curvature satisfying

$$|Rm|(x) \leq cr^{-2}(x), \text{ and } \text{vol } B_{x_0}(r) \leq Kr^n, \quad (2)$$

for some c and K ; here $r(x) = \text{dist to some point } x_0$. It follows that $\rho \sim r$, i.e. $cr(x) \leq \rho(x) \leq Cr(x)$, for some constants c, C provided $r(x) \geq 1$. If x_i is any divergent sequence of base points, the pointed manifolds (N, g_i, x_i) , $g_i = \rho(x_i)^{-2}g$ have uniformly bounded curvature outside the ε -ball about x_0 w.r.t. g_i , for any fixed $\varepsilon > 0$. (N, g) has Euclidean volume growth, i.e. $\text{vol } B_{x_0}(r) \geq kr^n$, for some $k > 0$, if and only if it is everywhere noncollapsed on the scale of its curvature radius, i.e. κ -noncollapsed. In this case (N, g_i, x_i) has a subsequence converging smoothly ($C^{1,\alpha}$) to a limit. The limit is called a tangent cone at infinity. If (N, g) has sub-Euclidean volume growth, so $\text{vol } B_{x_0}(r)/r^n \rightarrow 0$, then balls $B_{x_i}(\frac{1}{2}r_i)$ are κ -collapsed, for i large and any fixed κ . Hence, in this situation, the notions of collapse/noncollapse in the “usual” sense and κ -collapse/noncollapse are essentially the same.

However, the situation is different when

$$|Rm|(x) \gg r^{-2}(x). \quad (3)$$

Here, $\rho \ll r$ and when one rescales as above on a divergent sequence to make $\rho = 1$, $\text{dist}_{g_i}(x_i, x_0) \rightarrow \infty$. Thus, one may have κ -noncollapse, but “usual” collapse on much larger scales about x_i . The simplest example is of course the round product $S^2 \times \mathbb{R}$ or a warped product with the radius of the S^2 increasing slowly in the \mathbb{R} factor.

Consider for instance the cigar soliton; this is the metric on $\mathbb{R}^2 \times \mathbb{R}$ given by

$$g = (1 + r^2)^{-1} g_{Eucl} + dt^2,$$

where g_{Eucl} is the flat metric on \mathbb{R}^2 . A simple calculation shows that this metric is of positive sectional curvature, and the sectional curvature decays exponentially fast to 0 in the geodesic distance to some base point. Thus (2) holds. Since the volume growth satisfies $vol_g(B(s)) \sim s$, this metric collapses on the scale of its curvature, i.e. is κ -collapsed, (for any small κ), for s sufficiently large.

As stated in [I, Cor 4.2], [I, Thm 4.1] thus rules out both collapse of blow-ups and cigar solution limits (and limits with similar geometry) in finite time, at points of maximal, or near maximal curvature. These two issues were previously two major stumbling blocks in the Hamilton program.

Note that [I, Cor 4.2] pertains only to blow-ups at points of near maximal curvature. However, it is not nearly sufficient to analyse blow-up limits only in such regions; one must examine blow-up limits at arbitrary base points where the curvature diverges to infinity - the same issues arise my approach to geometrization, cf. [A2].

As seen above, a metric may be κ -noncollapsed, (i.e. non-collapsed on the scale of its curvature radius), but be collapsed in the usual sense at (much) larger scales. In fact, this is always the case for non-negatively curved blow-up limits of the Ricci flow, cf. [I, Prop. 11.4]. The distinction or tension between κ -noncollapse and large scale collapse plays a crucial role throughout [I, §11-12] and in [II]. It is perhaps the single most important underlying theme throughout the later part of the work.

While Sections [I, §1-4] introduce some of the main ideas and of course already prove several major results, this part of the paper is essentially a warm-up for the deeper and more important work to follow. Thus, the rest of [I] and [II] focus on local behavior of the Ricci flow $g(t)$, (in space and time), and are no longer concerned with global behavior (as with λ , μ , etc). The work begins in earnest in [I, §7], after a remarkable motivation of this work in [I, §6].

§5. Comment on §7.

The function u satisfies the equation $\square^* u = -u_t - \Delta u + R = 0$. For a fixed metric, this is the backward heat equation with fixed potential R . Replace t by $\tau = -t$, and view it as a forward heat equation in τ . In this context, the study of the associated functional \mathcal{L} , (without the τ factor), is classical and used frequently in physics. It arises in the derivation of the Feynman-Kac formula for the heat kernel, the Wiener measure, partition function, etc; see [F, §3] or other classic texts for example. The Li-Yau paper referred to in [I, 7.4] also makes an attribution to classical work, well-known to physicists. The variational formulas for \mathcal{L} to various orders are also studied in the perturbation expansion of the partition function.

What is novel here is of course that the background metric is now time-dependent, and evolves with the equation. Novel here also is the introduction of classical comparison geometry in this setting.

Note that the factor of $\sqrt{\tau}$ gives \mathcal{L} the dimensions of length, (and not energy), since τ has the dimension of length². Finally, one very minor point. I had a few difficulties (purely of my own making) verifying a few formulas in [I, 7.1]. These all come from the fact that time derivatives $\partial/\partial\tau$, e.g. $(\partial/\partial\tau)L$, are full time derivatives, while L_τ denotes partial derivatives in the direction of τ . This comes up in the equation preceding [I, (7.3)], in [I, (7.5), (7.15)], etc. All the formulas are fine in [I, §7] when one observes this distinction, but without it one is led astray.

§6. Comment on §8.

[I, Thm 8.2] is an important result for later work. A main point is that the estimate is local and time-independent. In contrast to [I, Thm.4.1], it can be applied at arbitrarily large times, independent of the initial metric.

The result is a natural analogue of the volume comparison theorem in the space-time (Ricci flow) setting - under the presence of a uniform local curvature bound. Thus, if the curvature radius at (x_0, t) is at least r_0 , for $0 \leq t \leq r_0^2$ and the initial ball $B_{x_0}(r_0)$ is A^{-1} non-collapsed, then at times r_0^2 , balls $B_x(r_0)$ are κ -noncollapsed, at all points (x, r_0^2) with $\text{dist}_{r_0^2}(x, x_0) \leq Ar_0$; here A may be chosen to be arbitrary, with then $\kappa = \kappa(A)$. This result would be trivial, by the volume comparison theorem, if one had a lower curvature bound on all of $(B_{x_0}(Ar_0), g_0^2)$. Its striking that only a curvature bound on $(B_{x_0}(r_0), t)$, $t \leq r_0^2$ is needed.

Using this result, [I, Thm 12.2] strengthens the conclusion to give uniform control of the curvature in the large ball $B_{x_0}(Ar_0)$ at time r_0^2 , under essentially the same hypotheses.

[I, §9] and [I, §10] are needed only in very minor ways regarding the geometrization conjecture itself. The pseudo locality theorem [I, Thm 10.1] is remarkable in its own right and leads to very significant consequences, (e.g. [I, 10.5]). [I, Cor. 9.5] relating the reduced distance with the potential $u = e^{-f}$ starting at a delta function is also very interesting.

§7. Comments on §11.

This chapter describes the basic models for singularity formation in finite time. In passing to a blow-up limit of a singularity, the hypothesis of non-negative curvature follows from the Hamilton-Ivey pinching estimate, (in dimension 3), while the κ -noncollapsed assumption, (and in particular the existence of a limit), follows from [I, Thm 4.1 or §7.3]. The assumption of bounded curvature (at time 0), corresponds formally to taking base points of (near) maximal curvature, as in [I, Thm 4.2].

Taking blow-up limits at general base points may well not give limits with uniformly bounded curvature. However, a good enough understanding of the bounded curvature limits as in [I, Thms. 11.7, 11.8] allows one to obtain a sufficient understanding of the general limits. Note also that the Hamilton distinction between Type I, II limits, although natural, does not arise here at all.

To start, the topological structure complete manifolds (M, g) of non-negative curvature is well understood in dimension 3. By the Gromoll-Meyer-Cheeger theorem, if M is non-compact then M is diffeomorphic to \mathbb{R}^3 - or a quotient of it - or isometric to $S^2 \times \mathbb{R}$ or its \mathbb{Z}_2 quotient. If M is compact, then M is diffeomorphic to a spherical space form, or isometric to $S^2 \times S^1$, by Hamilton's theorem.

Although it seems to be unknown whether all (spatially complete) blow-up limits of the Ricci flow are (homothetic) Ricci solitons. [I, Prop 11.2] shows that this is the case "asymptotically", i.e. any κ -solution is asymptotic to the past to a non-flat gradient shrinking soliton, with potential f given by the limiting reduced distance l !

This striking result is only used in a minor way in [I, §11]. Only the result in 2-dimensions is used, to obtain [I, Cor 11.3] via Hamilton's results. This latter result is used in turn to prove the important [I, Prop 11.4] (in all dimensions) by an induction argument on dimension! The remaining elements in the proof of [I, 11.4] are all quite standard and could have been proved long ago. The idea of dimension reduction is well-known and goes at least back to the study of singularities of minimal varieties several decades ago.

[I, Thm. 11.7 and Cor. 11.8] give an essentially complete understanding of the asymptotics of non-compact κ -solutions. Outside a sufficiently large compact set, the solution is (arbitrarily) close to the standard neck $S^2 \times \mathbb{R}$, on (arbitrarily) large regions. The size of the neck, i.e. the radius of S^2 , may grow or oscillate as one goes to infinity, but it can grow at most *sublinearly* in the distance r to a fixed base point, by [I, 11.4]. Hence, [I, Thm 11.7] implies that the scalar curvature satisfies

$$R(x) \gg r^{-2}(x), \text{ when } r(x) \gg 1. \quad (4)$$

If $R(x_i) \rightarrow 0$ as $x_i \rightarrow \infty$, and one rescales the metric at x_i to make $R(x_i) = 1$, then a subsequence converges to a complete κ -solution; in particular, the curvature stays uniformly bounded.

Observe that [I, Cor 11.6] turns this argument around. If one assumes (usual) non-collapse on scales much larger than the curvature radius, then the large scale collapse of κ -solutions in [I, Prop 11.4 and Cor 11.5] imply that one must have a uniform curvature bound. This allows one to control curvature just by local control of the volume!

[I, Cor 11.8], and the related [I, Thm 12.1], are of central importance. The ends of such κ -solutions far out have arbitrarily large regions arbitrarily close (after scaling to $R = 1$) to arbitrarily large regions in the standard product $S^2 \times \mathbb{R}$. In particular, no ends are of the form $S^1 \times \mathbb{R}^2$.

This means that one has 2-spheres embedded in a canonical geometry on which one can do surgery to try to simplify the Ricci flow. This has been the main unresolved conceptual issue in the Ricci flow program.

§8. Comments on §12.

This is the most important section of the paper. We'll basically just paraphrase the results, and give names to them, since all of them are important.

[I, Theorem 12.1] - Canonical Neighborhood Theorem.

In finite time, at any point (x_0, t_0) where the curvature $R(x_0, t_0) = Q$ is sufficiently large, a Q -scaled space-time neighborhood at and to the past of t_0 is (arbitrarily) close to an (arbitrarily) large space-time domain in a κ -solution from §11; here κ is freely specifiable, with $Q = Q(\kappa)$. Arbitrarily close, or arbitrarily large size of the space-time domain just depend on Q .

This applies to arbitrary base points - not just those of near maximal curvature on a spatial slice $(M, g(t_0))$ or of near maximal curvature in time t , $t \leq t_0$. It implies that blow-up limits at any sequence of points where the curvature diverges to infinity are necessarily complete spatially and ancient in time, and of locally bounded curvature; the curvature of the limit may not be bounded globally.

This result is effectively a version of the “general differential Harnack inequality” long sought for in the Hamilton program. Given an arbitrary base point (x, t_0) , it allows one to control the curvature in large space-time neighborhoods “centered” at (x, t_0) , in the right scale, in terms of control of the curvature at (x, t_0) . It prevents the curvature (or the metric) from fluctuating (wildly) both in space and time as one approaches a singularity.

The time interval $[0, T]$ on which this result holds is determined, by the local non-collapse results, just by the choice of κ .

[I, Thm 12.1] suffices to carry out the finite time surgery arguments in [II]. The rest of [I, §12] is needed to understand the long term behavior $t \geq T$, where one no longer has the κ -noncollapsed property, for sufficiently large $T = T(\kappa)$.

[I, Theorem 12.2] (see [II, 6.2]) - Large Scale Curvature Bound.

(After rescaling), if the curvature is bounded in $(B_{x_0}(1), t)$, $t \in [0, 1]$, then it stays bounded in arbitrarily large balls about x_0 at time 1. The assumptions needed are just a non-collapse constant for $vol B_{x_0}(1)$ at time 0 and sufficiently small value of the pinching function ϕ , for example if $tR(x_0, t)$ is sufficiently large, for the original (unscaled) flow.

This serves as a partial replacement for [I, Thm 12.1] in studying the long-time behavior in non-collapsed regions, provided one has control over the curvature in balls of unit size. The next result addresses this local hypothesis on the curvature bound.

[I, Theorem 12.3] - Apriori Local Curvature Bound.

This result gives an apriori local *upper* bound on the curvature, given just a *lower* bound and non-collapse in somewhat larger balls. Moreover, the estimate is uniform for a definite amount of time to the past of t_0 . A lower bound on the curvature is usually easy to obtain by the pinching theorem. The main hypothesis is non-collapse on the ball. The estimates are time-independent.

Thus, if the curvature is bounded below by $-r_0^2$ and at time t_0 the ball $B_{x_0}(r_0)$ is non-collapsed (with bound ω), then the curvature is bounded above, $R(x, t) \leq Kr_0^{-2}$, for $x \in B_{x_0}(\frac{1}{4}r_0)$, and on

a definite time interval to the past, depending only on ω . This for all r_0 such that $\phi(r_0^{-2})$ is small and $r_0 \leq \bar{r}t_0^{1/2}$, \bar{r} depending only on ω .

[I, Corollary 12.4] - Large negative curvature implies collapse.

If the curvature is sufficiently negative somewhere on some ball, then the ball must be (highly) collapsed. This holds provided the ball is of small radius r_0 , with r_0 satisfying the bounds above in [I, Thm 12.3].

As explained in [KL], the idea, by contradiction, is that if such a ball is non-collapsed, one can get an apriori upper curvature bound to the past on smaller ball, then move forward by [I, Thm 12.2 and Thm 8.2] to get an upper curvature bound on the original ball. Via the pinching estimate, an upper bound on the curvature implies a (small) lower bound on the curvature.

These results basically lead to the thick/thin decomposition, where one has uniform two-sided curvature control on the thick part, and a local lower curvature bound on the collapsed part. Essentially all the estimates in [I] are scale-invariant. In particular, the results after [I, Thm. 12.1] apply also to large-scale behavior, where one blows-down $g(t)$ by t far out in time.

§9. Comments on II, §3.

It is not asserted that the limit domain Ω at time T has only finitely many components. There are only finitely many components to Ω_ρ , for any given $\rho > 0$ small. Each component C of Ω containing a component C_ρ of Ω_ρ has finite topological type, with a finite number of ends, and C is obtained from C_ρ by attaching ε -tubes, ε -caps or ε -horns to the ends of C_ρ . However, there could exist infinitely many components of Ω , none of them in Ω_ρ for a fixed ρ ; each of these is either a double ε -horn or a capped ε -horn. There is no apriori lower volume bound on such components.

Thus, the singular set $M \setminus \Omega$ of the Ricci flow could be rather complicated and Ω may contain infinite strings of double ε -horns.

Suppose for example that $\Omega \neq \emptyset$ but $\Omega_\rho = \emptyset$ at some singularity time T , for some fixed ρ sufficiently small. Thus Ω consists of a collection of very thin, high curvature, double ε -horns and capped ε -horns. Slightly before time T , the canonical neighborhood theorem implies that M is obtained from Ω by adding in very thin ε -tubes or ε -caps to produce a closed manifold. Hence, as stated, M is topologically S^3 , $S^2 \times S^1$, \mathbb{RP}^3 or $\mathbb{RP}^3 \# \mathbb{RP}^3$. Moreover, the metric has large positive scalar curvature, and so becomes extinct in very short time. This discussion also applies to each component of M after any number of surgeries has been performed. It also applies when $\Omega = \emptyset$ if S^3/Γ is added to the list. Note that the process above involves no surgeries on M .

If either M , or if some component of M after surgery has this form, it is not analysed any further and “thrown away”. One sees here the importance of the Kneser finiteness theorem (there are only finitely many (non-isotopic) essential 2-spheres in any closed 3-manifold), so that topologically one knows exactly what is being thrown away and how it was glued to the original M . It might be complicated to analyse the Ricci flow further on these pieces. One has no apriori lower volume bound, so the set of surgery times may not be locally finite, etc. On the other hand, the kinds of singularities that can occur are just repetitions of those above: ε -caps and horns. (I have not checked if the surgery can be performed well in this context).

Suppose instead that, at some singular time T , (and component of M), $\Omega_\rho \neq \emptyset$ at T . Then $\Omega \setminus \Omega_\rho$ may have the infinite complexity discussed above. However, now the surgery is performed near the boundary components of Ω_ρ , (within the ε -horns of Ω and at a small but definite scale compared with ρ); thus 3-balls of the form of a standard ε -cap are glued onto each side of the 2-sphere boundary. The geometry, and in particular volume, is controlled by ρ and so there are only finitely many such surgeries on Ω_ρ , each one producing a closed 3-manifold with “canonical” metric $g(T)$.

In this way, one obtains a solution to the Ricci flow with surgery, defined for infinite time $[0, \infty)$, and defined possibly with more and more components, but with all but a finite number (depending on ρ) thrown away just after they appear in surgery.

§10. Comments on II, §7.

Suppose the Ricci flow with surgery exists for a sufficiently large time T , assuming normalized initial data. Thus, parts of M may have been cut off in earlier time, but these parts have already been identified to be S^3/Γ or $S^2 \times S^1$ components, attached to M by connected sum. Thus, at time T , the Ricci flow takes place on closed 3-manifolds \hat{M}_i , with

$$M = (\hat{M}_1 \# \dots \# \hat{M}_k) \# (\# S^3/\Gamma_i) \# (\# S^2 \times S^1). \quad (5)$$

Here S^3/Γ_i is a spherical space form - possibly S^3 , and the sum above is finite, depending on T . In the following, we work on some \hat{M}_i and set $\hat{M}_i = \hat{M}$.

The main point is then at time T , \hat{M} is topologically a union of a hyperbolic manifold H , (complete, finite volume), not necessarily connected, and a graph manifold G , again not necessarily connected;

$$\hat{M} = H \cup G. \quad (6)$$

It is possible that either G or H is empty, in which case \hat{M} is a closed hyperbolic manifold or graph manifold respectively. If both G and H are non-empty, their boundaries are a collection of tori, each incompressible in \hat{M} . The decomposition of \hat{M} into H and G corresponds to the thick/thin decomposition of the metric $g(T)$.

This is a topological statement. It does not imply that the smooth Ricci flow exists for all time $t > T$. Further singularities may form, and further surgeries may be necessary to continue the Ricci flow past T . However, the main point is that all singularities form in G , i.e. in the collapsing region. On compact subsets of $H \subset \hat{M}$, the smooth flow exists for all time, and converges, when rescaled by t^{-1} , to the complete hyperbolic metric of curvature $-\frac{1}{4}$.

Note that graph manifolds need not be irreducible; in fact the class of graph manifolds is closed under connected sum. Thus, the Ricci flow arguments do not (yet) prove analytically or geometrically the Kneser sphere theorem, that there are only finitely many essential 2-spheres in M . Analytically, G may a priori possess infinitely many ‘‘essential’’ 2-spheres along which surgery takes place. (Of course Ricci flow may also surger inessential 2-spheres, giving connected sums with S^3).

The decomposition (6) is unique up to isotopy; thus, given \hat{M} , the topological types of H and G are unique, and the tori dividing M into the H and G regions are unique up to isotopy in \hat{M} . This of course rests crucially on the fact that the tori are incompressible. For proofs, see [A3] or [JS].

At least if $H \neq \emptyset$, it follows that *all* solutions of the Ricci flow with surgery exist for infinite time, and all converge to the same decomposition (6). All solutions converge to the hyperbolic metric on H , unique by Mostow rigidity. No claim regarding uniqueness of the behavior on G is made. If $H = \emptyset$, so $\hat{M} = G$, it is possible that the solution could become extinct in some later time. This situation is now much better understood by Perelman’s most recent paper [III].

We describe the thick-thin decomposition in more detail, although everything is already in [II]. This is done in a two-step process. Let \hat{M} be as above, and consider the flow at time T and times $t \geq T$. First, as in [II, §7.3], let now ρ be the curvature radius w.r.t. a *lower* curvature bound: $\rho(x, t) = \sup r$ such that

$$Rm(y, t) \geq -r^{-2}, \text{ for all } y \in B_x(r, t). \quad (7)$$

From the work in [II, §6] (compare with [I, Thm 12.4], it follows that for any choice of $\omega > 0$ (small), there exists $\bar{\rho} = \bar{\rho}(\omega)$ so that if $\rho(x, t) \leq \bar{\rho}t^{1/2}$, then $B_x(\rho(x, t))$ is ω -collapsed. Let $\hat{M}^-(\omega, t)$ be the collection of such ω -collapsed points in \hat{M} , the ω -thin part of \hat{M} . In some places ρ may be very small,

but in other places, it may not be small (on the scale $t^{-1}g(t)$). It is possible that $\hat{M}^-(\omega, t) = \emptyset$, so that no part of $(M, g(t))$ is highly collapsed. Its equally possible that $\hat{M}^-(\omega, t) = \hat{M}$, so that $g(t)$ is locally collapsed everywhere with a local lower curvature bound; there is no apriori uniform lower bound for ρ . Singularities may (or may not) form in $\hat{M}^-(\omega, t)$ for some finite t . Near singularities, balls will be κ -noncollapsed, i.e. noncollapsed on the scale of the curvature, for some (maybe small) κ . The full curvature scale will then be much less than ρ near such points.

In any case, fix $\omega > 0$ sufficiently small from now on, and let $\hat{M}^+(\omega, t) = \hat{M} \setminus \hat{M}^-(\omega, t)$, the ω -thick part of \hat{M} . Then [II, 7.2(a)] implies that all sufficiently small balls in the metric $t^{-1}g(t)$ centered at points in $\hat{M}^+(\omega, t)$ are ξ -close to the hyperbolic metric of curvature $-1/4$. This holds for any $t \geq T$, with ξ and the size of the ball depending only on T . [II, 7.2(b),(c)] show the same holds for (arbitrarily) large balls, and (arbitrarily) large times to the future of $t \geq T$, again provided T is chosen correspondingly sufficiently large again.

It follows that for $t \geq T$ sufficiently large, $\hat{M}^+(\omega, t)$ is diffeomorphic to a compact domain in a complete hyperbolic manifold H of finite volume, (possibly disconnected). If $\partial\hat{M}^+(\omega, t) = \emptyset$, then $\hat{M} = \hat{M}^+(\omega, t)$ is topologically a closed hyperbolic manifold. It follows that $g(t)$ can't be ω -collapsed anywhere, for some absolute choice of ω , for any t sufficiently large. Thus, [II, 7.2] implies that the normalized smooth flow exists for all time and converges to the hyperbolic metric of curvature $-1/4$. If $\partial\hat{M}^+(\omega, t) \neq \emptyset$ and ω is chosen sufficiently small, then $\partial\hat{M}^+(\omega, t)$ appears far down a standard hyperbolic cusp $T^2 \times \mathbb{R}^+$. One may cut off any part of $\hat{M}^+(\omega, t)$ sufficiently far down any cusp to give a new boundary consisting of nearly standard tori in the hyperbolic cusp. It follows that $\hat{M}^+(\omega, t)$ is diffeomorphic to H itself. Now redefine $\hat{M}^-(\omega, t)$ to be \hat{M} with the modified $\hat{M}^+(\omega, t)$ removed. Then [II, Thm 7.4] implies that $\hat{M}^-(\omega, t)$ is a graph manifold, with almost flat, almost umbilic toral boundary. This gives the decomposition (6). Hamilton's minimal surface argument [H] gives the incompressibility of the tori; see the end of this section for more on this. The Ricci flow with surgery continues on $\hat{M}^-(\omega, t)$ for infinite time, unless it becomes extinct.

Application to the Sigma Constant.

The estimates in [II, §7.1] also lead easily to the determination of the Sigma constant $\sigma(M)$ of M , when $\sigma(M) \leq 0$. Here

$$\sigma(M) = \sup_{g \in \mathcal{Y}} R_g V^{2/3}, \quad (8)$$

where \mathcal{Y} is the space of Yamabe metrics on M and R_g is the scalar curvature (constant) of g . The claim is then that $\sigma(M)$ is realized by the hyperbolic metric on H , i.e.

$$|\sigma(M)| = 6(\text{vol}_{-1}H)^{2/3} = \frac{3}{2}(\text{vol}_{-1/4}H)^{2/3}. \quad (9)$$

In particular, the graph manifold part G , and the positive parts S^3/Γ_i and $S^2 \times S^1$, if any, are invisible to $\sigma(M)$. To see this, suppose first that M is irreducible, so that the connected sum decomposition (5) is trivial, and $M = \hat{M}$. Let

$$\bar{R}(M) = \sup R_{\min} V^{2/3}, \quad (10)$$

where the sup is taken over \mathbb{M} and R_{\min} is the minimum of the scalar curvature of the metric. Since Yamabe metrics are of constant scalar curvature,

$$\bar{R}(M) \geq \sigma(M).$$

On the other hand, given any g , if $\tilde{g} = u^4g$ is the Yamabe metric of the same volume as g in $[g]$, then

$$u^5 \tilde{R} = -8\Delta u + Ru.$$

The maximum principle then implies that $\tilde{R} \geq R_{min}$, when $\tilde{R} \leq 0$. It follows that

$$\bar{R}(M) = \sigma(M). \quad (11)$$

To prove (9), use (11). Given the fact that \hat{M} has the decomposition (6), it is easy to construct a metric g on \hat{M} such that $R_{min}V^{2/3}(g) \geq -\frac{3}{2}(vol_{-1/4}H)^{2/3} - \varepsilon$, for any given $\varepsilon > 0$, so that $\bar{R}(M) \geq -\frac{3}{2}(vol_{-1/4}H)^{2/3}$. Suppose then there is a metric g_0 such that $\hat{R}(g_0) = R_{min}V^{2/3}(g_0) > -\frac{3}{2}(vol_{-1/4}H)^{2/3}$. Use this as an initial metric for the Ricci flow. The quantity $\hat{R} = R_{min}V^{2/3}$ is monotone non-decreasing in t , and converges to a limit \tilde{R} . When rescaling by t^{-1} , $R_{min} \rightarrow -3/2$. However, the decomposition (6) is unique, (assuming incompressible tori) and so one must have that the limit volume $\tilde{V} \geq vol_{-1/4}H$. Hence, in the limit, $\tilde{R} \leq -\frac{3}{2}(vol_{-1/4}H)^{2/3}$, a contradiction.

One could base a similar argument on the behavior of λ instead, as in [II,§8].

If M is not irreducible, then M is a connected sum of positive factors S^3/Γ , $S^2 \times S^1$, and non-positive factors \hat{M}_i . The equality (9) holds for each i . One can perform the connected sum surgery to increase the scalar curvature pointwise, and with an arbitrarily small change to the volume, so that

$$\bar{R}(M) \geq -\frac{3}{2}(vol_{-1/4}H)^{2/3},$$

where H is now the union of the hyperbolic manifolds in each \hat{M}_i ; this argument is basically due to O. Kobayashi [K].

Suppose then there is a metric g_0 on M such that $\hat{R}(g_0) > -\frac{3}{2}(vol_{-1/4}H)^{2/3}$. Then the same argument as before running the Ricci flow on g_0 as initial metric gives a contradiction.

Remark. It is not asserted that the Ricci flow, (rescaled by t^{-1}) volume collapses the graph manifold part G , i.e. that

$$vol_{t^{-1}g(t)}G \rightarrow 0. \quad (12)$$

It is reasonable to expect that (12) does hold. However, it is not needed for the determination of $\sigma(M)$. While the asymptotic behavior of the Ricci flow on the hyperbolic part H is canonical - the same for all initial data - the flow on G may be very different for different initial data.

The proof of the incompressibility of the tori dividing H and G in (6) in [A3] implicitly uses the fact that $\sigma(\hat{M})$ is given by (9). However, it is easy to see that the proof holds if one starts the Ricci flow at a metric g_0 on \hat{M} such that $\hat{R}(g_0) > \bar{R}(\hat{M}) - \varepsilon$, where ε is a fixed constant depending only on the topology of \hat{M} and $\bar{R}(\hat{M}) < 0$. Thus, start the Ricci flow at g_0 and repeat the argument above establishing (9) to obtain a limit configuration (6). Any compressible torus in this configuration will increase $\hat{R}(g(T))$, for T large, by a definite amount more than ε . This contradiction then shows that the decomposition (6) is along incompressible tori.

If it can be proved that (12) holds, for any solution to the Ricci flow with surgery, then the argument above can be iterated a finite number of times - depending on the initial metric - to again prove the incompressibility of the tori (and the corresponding uniqueness of (6)). Otherwise, without either of these assumptions, one must use Hamilton's argument [H] to prove incompressibility.

§11. Appendix.

The trace-free part of $\nabla \mathcal{S}$ is $-z = -(Ric - \frac{R}{3}g)$. Consider first the variation of the trace-free curvature in the direction $+z$; $z' = (Ric - \frac{R}{3}g)'$. From standard formulas, ([B, Ch 1K]),

$$Ric' = \frac{1}{2}D^*Dz - \delta^*\delta z - \frac{1}{2}D^2trz + O,$$

where O denotes lower order. Also $R' = -\Delta trz + \delta\delta z - |z|^2$. Hence, under the ‘‘gradient flow’’, this gives the evolution equation

$$\frac{dz}{dt} = -(z') = -\frac{1}{2}D^*Dz + \delta^*\delta z + \frac{1}{3}(\delta\delta z)g + O.$$

The first (main) term is basically the Laplacian, so this corresponds to a forward evolution equation for the trace-free curvature. I have not checked however if the right side above is elliptic in z . (This is not hard - just lazy to compute the symbol).

Regarding the 2nd variational formula for \mathcal{F}^m , we have

$$\mathcal{F}^m = \int (R + |\nabla f|^2)e^{-f} dV,$$

where the measure $dm = e^{-f} dV$ is fixed. Let g_s be a curve of metrics through $g = g_0$, and similarly for f_s , so that $e^{-f_s} dV_{g_s}$ is fixed. Taking the derivative of this gives

$$\frac{1}{2}tr_{g_s} \frac{dg_s}{ds} = \frac{df_s}{ds}, \quad (13)$$

along the curve. Suppose $h = dg/ds$ is orthogonal to the orbit of \mathcal{D} , so that

$$\delta h = 0.$$

The 1st variation of \mathcal{F}^m is given by

$$\frac{d\mathcal{F}^m}{ds}(g_s, f_s)|_{s=0} = (\mathcal{F}^m)'(h, \eta) = - \int \langle Ric + D^2 f, h \rangle dm.$$

For the 2nd variation,

$$\frac{d^2\mathcal{F}^m}{ds^2}(g_s, f_s)|_{s=0} = - \int \langle \frac{d}{ds} Ric_{g_s} + \frac{d}{ds}(D_{g_s}^2 f_s), h \rangle dm + O,$$

where O denotes other terms which are lower order. Again, modulo lower order terms,

$$\frac{d}{ds} Ric_{g_s} = \frac{1}{2}D^*Dh - \delta^*\delta h - \frac{1}{2}D^2 tr h + O$$

Also, to leading order,

$$\frac{d}{ds}(D_{g_s}^2 f_s) = D^2 \frac{df}{ds} = \frac{1}{2}D^2 tr h + O.$$

Since $\delta h = 0$, it follows that to leading order

$$\frac{d^2\mathcal{F}}{ds^2}(g_s, f_s)|_{s=0} = -\frac{1}{2} \int \langle D^*Dh, h \rangle dm + O.$$

The operator $-D^*D$ is negative definite, so it follows that the full 2nd order variation has at most a finite dimensional space of positive directions orthogonal to \mathcal{D} .

If h is tangent to \mathcal{D} , so that $h = \delta^*X$ for some vector field X , then a computation using (13) shows that $\frac{1}{2}D^*Dh - \delta^*\delta h$ is lower order, (i.e. 1st order in h). I’ve not carried out the further computations to see form the 2nd variation takes in such directions.

REFERENCES

- [I] G. Perelman, The entropy formula for the Ricci flow and its geometric applications, (preprint), math.DG/0211159.
- [II] G. Perelman, Ricci flow with surgery on three-manifolds, (preprint), math.DG/0303109.
- [III] G. Perelman, Finite extinction time for the solutions to the Ricci flow on certain three-manifolds, (preprint), math.DG/0307245.
- [A1] M. Anderson, Extrema of curvature functionals on the space of metrics on 3-manifolds, *Calc. Var. and PDE*, **5**, (1997), 199-269.
- [A2] M. Anderson, Scalar curvature, metric degenerations and the static vacuum Einstein equations, I, II, *GAF*, **9**, (1999), 855-967, and **11**, (2001), 273-381.
- [A3] M. Anderson, Scalar curvature and geometrization conjectures for 3-manifolds, in *Comparison Geometry*, (Berkeley 1993-94), MSRI Publications, **30**, (1997), 49-82, and *Scalar curvature and the existence of geometric structures on 3-manifolds, I*, *Jour. Reine Angew. Math.*, **553**, (2002), 125-182.
- [B] A. Besse, *Einstein Manifolds*, Springer Verlag, 1987.
- [F] R. Feynman, *Statistical Mechanics*, Addison-Wesley, 1972.
- [H] R. Hamilton, Nonsingular solutions of the Ricci flow on three-manifolds, *Comm. Anal. Geom.*, **7**, (1999), 695-729.
- [JS] W. Jaco and P. Shalen, Seifert fibered spaces in 3-manifolds, *Memoirs Amer. Math. Soc.*, **21**, No. 220, (1979).
- [KL] B. Kleiner and J. Lott, Notes on Perelman's Paper, (homepages of Kleiner or Lott).
- [K] O. Kobayashi, Scalar curvature of a metric of unit volume, *Math. Annalen*, **279**, (1987), 253-265.

July 30, 2003

E-mail: andersonmath.sunysb.edu