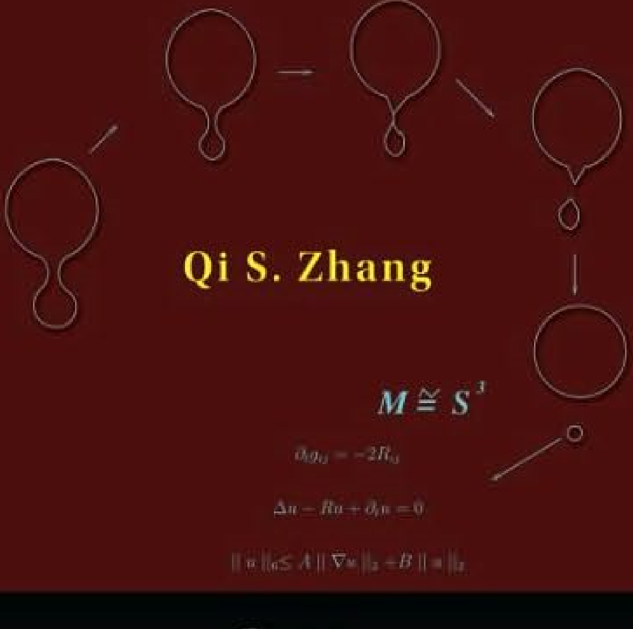
# Sobolev Inequalities, Heat Kernels under Ricci Flow, and the Poincaré Conjecture




## Qi S. Zhang

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

# Sobolev Inequalities, Heat Kernels under Ricci Flow, and the Poincaré Conjecture

# Sobolev Inequalities, Heat Kernels under Ricci Flow, and the Poincaré Conjecture

Qi S. Zhang

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742



Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4398-3459-6 (Hardback)



**Library of Congress Cataloging-in-Publication Data**

Zhang, Qi S.
Sobolev inequalities, heat kernels under Ricci flow, and the Poincaré conjecture / Qi S. Zhang.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4398-3459-6 (hardcover : alk. paper)
1. Inequalities (Mathematics) 2. Sobolev spaces. 3. Ricci flow. 4. Poincaré conjecture. I. Title.

QA295.Z43 2011
512.9'7--dc22 2010018868

**Visit the Taylor & Francis Web site at**
**http://www.taylorandfrancis.com**

**and the CRC Press Web site at**
**http://www.crcpress.com**

# Preface

First we provide a treatment of Sobolev inequalities in various settings: the Euclidean case, the Riemannian case and especially the Ricci flow case. Then, we discuss several applications and ramifications. These include heat kernel estimates, Perelman's $W$ entropies and Sobolev inequality with surgeries, and the proof of Hamilton's little loop conjecture with surgeries, i.e. strong noncollapsing property of 3 dimensional Ricci flow. Finally, using these tools, we present a unified approach to the Poincaré conjecture, which seems to clarify and simplify Perelman's original proof. The work is based on Perelman's papers [P1], [P2], [P3], and the works Chow etc. [Cetc], Chow, Lu and Ni [CLN], Cao and Zhu [CZ], Kleiner and Lott [KL], Morgan and Tian [MT], Tao [Tao] and earlier work of Hamilton's. The first half of the book is aimed at graduate students and the second half is intended for researchers.


## Acknowledgment

This writing is derived from the lecture notes for a special summer course in Peking University in 2008 and another summer course in Nanjing University in 2009. I am deeply grateful to Professor Gang Tian and Professor Meiyue Jiang for the invitation to the School of Mathematics at Peking University and to Professor Gang Tian again for the invitation to Nanjing University. I feel fortunate to have the opportunity to enjoy the hospitality, generosity and excellent working condition in both universities and the cities where ancient tradition and modernity are displayed in splendor. Thanks also go to Professors Xiao Dong Cao, Bo Dai, Yu Guang Shi, Xing Wang Xu, Hui Cheng Yin, and Xiao Hua Zhu for their interests in the course and discussions on related mathematical problems. I am also indebted to the students who come from many parts of China to attend the classes, and to

Ms Wu and Ms Yu for their technical assistance. Special thanks go to Professors Bennett Chow and Lei Ni who invited me to a summer workshop on geometric analysis in 2005, which introduced me to Ricci flow, to Professor Gang Tian who recommended to publish the Chinese version of the lecture notes as a book by Science Press Beijing, and to Dr. Sunil Nair for inviting me to submit the English version to CRC Press. During the preparation of the book, I have also received helpful suggestions or encouragement from Professors Huaidong Cao, Jianguo Cao, Xiu Xiong Chen, Bo Guan, Nicola Garofalo, Qing Han, Emmanuel Hebey, Zhen Lei, Junfang Li, John Lott, Peng Lu, Jie Qing, Yanir Rubinstein, Philippe Souplet, Bun Wong, Sumio Yamada, Zhong-Xin Zhao, Yu Zheng and Xiping Zhu. Materials from Chapter 2 and 4 were also used for a graduate course at University of California, Riverside. I thank Jennifer Burke and Shilong Kuang for taking notes, parts of which were incorporated in the book. I am also indebted to Professors Xiao Yong Fu and Murugiah Muraleetharan for reading and checking through the whole book. Professor Xiao Yong Fu with the assistance of Jun Bin Li also translated Chapters 2–6 into Chinese and made numerous corrections.

Finally I have benefited amply from studying the works Chow etc. [Cetc], Chow-Lu-Ni, [CLN], Cao-Zhu [CZ], Kleiner-Lott [KL], Morgan-Tian [MT], Tao [Tao] and Perelman [P1], [P2]. I wish to use the occasion to thank them all.

I dedicate this book to my family members: Wei, Ray, Weiwei, Misha, and to my parents.

# Contents

# Chapter 1

# Introduction

The book is centered around Sobolev inequalities and their applications to analysis on manifolds, and in particular to Ricci flow. There are two objectives. One is to serve as an introduction to the field of analysis on Riemann manifolds. The other is to use the tools of Sobolev imbedding and heat kernel estimates to study Ricci flows, especially in the case with surgeries, a research field that has attracted much attention. Rather than making a comprehensive presentation, the aim is to present key ideas, to explain the hard proofs and most important applications in a succinct, accessible and unified manner.

Roughly speaking, a Sobolev inequality states that if the derivative of a function is integrable in certain sense ($L^p$, etc.), then the function itself has better integrability. It lies in the foundation of modern analysis. For example, Sobolev imbedding is an essential tool in studying partial differential equations since the goal of solving a differential equation is to integrate out the derivatives to recover the unknown function. On the other hand, a Sobolev inequality will also yield interesting partial differential equations via minimizing the Sobolev constants. It can also reveal useful information on the underlying space or manifold. This last property is the focus of this book.

The book is divided into three parts. Chapter 2 is the first part, where we will present basic materials on Sobolev inequalities in the Euclidean case. These include the standard $W^{1,p}(\mathbf{R^n})$ imbedding into $L^{np/(n-p)}$, Orlicz and $C^\alpha$ spaces, when $p \in [1, n)$, $p = n$ and $p > n$ respectively. We will also present the Poincaré inequality and Log Sobolev inequality. All these materials can be found in standard books, such as [GT], [Maz], [Ad], [LL]. The prerequisite for this part is graduate Real Analysis.

The second part consists of Chapters 3–4. Here we discuss Sobolev imbedding on compact or noncompact Riemann manifolds with fixed metrics. The main theme is to prove several close relatives of the Sobolev imbedding. These include log Sobolev inequality, certain heat kernel estimates, Poincaré inequality and doubling condition, Harnack inequalities, etc. We will also show that the validity of certain Sobolev inequalities imply such geometric properties as volume noncollapsing, isoperimetric inequalities. Much of the material in this part is taken from [Heb2] and [Sal]. The reader needs some basic knowledge of Riemann geometry. In Chapter 3, a very brief summary of the most relevant results in basic Riemann geometry is provided.

The third part starts from Chapter 5, where we outline a few basic results of R. Hamilton's Ricci flow. Starting from Chapter 6, we turn to some recent research on the Poincaré conjecture.

From Perelman's original papers [P1], [P2], [P3] and the works by Cao and Zhu [CZ], Kleiner and Lott [KL] and Morgan and Tian [MT], and Tao [Tao2], [Tao], it is clear that the bulk of the proof of the Poincaré conjecture is consisted of two items. One is the proof of local noncollapsing with or without surgeries, and the other is the classification of backward limits of ancient $\kappa$ solutions. After these are done, by a clever blow-up argument, also due to Perelman, one can show that regions where the Ricci flow is close to forming singularity have simple topological structure, i.e. canonical neighborhoods. Then one proceeds to prove that the singular region can be removed by finite number of surgeries in finite time. When the initial manifold is simply connected, the Ricci flow becomes extinct in finite time [P3] (see also [CM]). Thus the manifold is diffeomorphic to $\mathbf{S}^3$, as conjectured by Poincaré.

Besides the results and techniques by R. Hamilton, the main new tools Perelman used in carrying out the proof are several monotone quantities along Ricci flow. These include the $W$ entropy, reduced volume and the associated reduced distance. In [P1], Perelman first used his $W$ entropy to prove local noncollapsing for smooth Ricci flows, i.e. the little loop conjecture by Hamilton. This result is a major breakthrough for the theory of Ricci flow. However he then turned completely to the reduced volume (distance) to prove the classification and a weaker form of noncollapsing with surgeries. The $W$ entropy is not used anymore. The reduced distance is a distance in space time, which is suitably weighted by the scalar curvature. Even though it is not smooth or positive in general, Perelman shows that the reduced distance miraculously satisfies certain differential equalities and inequalities in the weak sense. But a rigorous proof of these is lengthy and intensive.

The main aim for the third part of the book is to show that the W entropy and its monotonicity actually imply certain uniform Sobolev inequalities along Ricci flows, which have many ramifications. One of them is the aforementioned local noncollapsing result with or without surgeries. Another one is the classification of the backward limits of ancient $\kappa$ solutions. This enables one to give a proof of the Poincaré conjecture using techniques unified around W entropy, Sobolev imbedding and related heat kernel estimate. This method allows one to bypass reduced distance and volume which are central to Perelman's original proof. The reduced distance, being neither smooth nor positive in general, is one reason for the complexity of the original proof. The proof presented here, within Perelman's framework, seems more accessible to wider audience since the techniques involving Sobolev inequalities and heat kernel estimates are familiar to many mathematicians. Much of the highly intensive analysis involving reduced distance and volume is now replaced by the study of the $W$ entropy and the related uniform Sobolev inequalities and heat kernel estimates. These results have independent interest since they are instrumental to analysis on manifolds. Besides, due to the relative simplicity and versatility, we believe the current technique can lead to better understanding of other problems for Ricci flow. One such result is the proof of Hamilton's little loop conjecture with surgeries in Chapter 8. Some applications are also found by several authors for Kähler Ricci flow, as indicated in Section 6.2.

We should mention that the reduced distance and volume are still needed so far for the proof of the geometrization conjecture. Actually, they are needed, but only in the proof of Perelman's no collapsing Theorem II with surgeries, i.e. Proposition 6.3 (a) in [P2]. A short proof of the nonsurgery version is given in Section 6.3.

Let us briefly describe the content of Chapters 6–9. In Chapter 6, we show that Perelman's $W$ entropy is just the formula in a log Sobolev inequality and the monotonicity of a family of $W$ entropies implies certain uniform Sobolev inequalities along a smooth Ricci flow. By earlier result of Carron [Ca] and Akutagawa [Ak], local noncollapsing of Ricci flow follows as a corollary. In Section 6.3, we also report two new results which are not directly related to the proof of Poincaré conjecture. One is a uniform Sobolev inequality with critical exponents. The other is a localized uniform Sobolev inequality which implies the smooth version of Perelman's no local collapsing theorem II [P1] as a special case. These results are not published elsewhere. The content in Sections 6.4 and 6.5 are taken from the papers [KZ] and [Z1] respectively. The former is a differential Harnack inequality for all positive solutions of the conjugate heat equation. The later is certain upper bound for the fundamental solution of the conjugate heat equations.

In Chapter 7, we present Perelman's classification of backward limits of ancient $\kappa$ solutions and the canonical neighborhood property for 3 dimensional Ricci flow.

The main goal of Section 7.2 is to establish certain Gaussian type upper bound for the heat kernel (fundamental solutions) of the conjugate heat equation associated with 3 dimensional ancient $\kappa$ solutions to the Ricci flow. As an application, in Section 7.3, using the $W$ entropy associated with the heat kernel, we give a different and shorter proof of Perelman's classification of backward limits of these ancient solutions. The question of whether or not the classification can also be done in this way has been raised in [Tao] e.g.

In Chapter 8, using the idea of Sobolev imbedding developed in Chapter 6 and being inspired by the last section of [P2] and [KL], we prove a uniform Sobolev inequality which is independent of the number of surgeries. As one application, a strong finite time $\kappa$ noncollapsing result for Ricci flow with surgeries is proven. This gives the first proof of Hamilton's little loop conjecture with surgeries in 3 dimension case. We mention that the smooth version of the little loop conjecture is proven by Perelman [P1] using the monotonicity of his $W$ entropy. The main work in the surgery case involves the analysis of eigenvalues of the minimizer equation of the W entropy over a horn like manifold. As a result it is shown that the best constant of the associated log Sobolev inequalities differ at most by the change in volume if the underlying manifold undergoes surgeries (cut and paste). The proof, without using reduced distance or volume, is short and seems more accessible. Its main advantage over the weaker $\kappa$ noncollapsing proved by Perelman is that the relevant geometric information concentrates only on one time slice, thus avoiding the complication associated with surgeries which can happen shortly before this time slice. This chapter is based on [Z4].

In Chapter 9, with the help of strong $\kappa$ noncollapsing, we will give a detailed proof of Perelman's existence theorem of Ricci flow with surgeries. Something new about this chapter is the proof of Lemma 9.1.1, which describes the evolution of regions near surgery caps. This lemma is a key step in proving there are finitely many surgeries in finite time. Here we provide a proof which is considerably different from the one outlined in [P2], Lemma 4.5. The reason is that we are unable to follow the original proof by Perelman, which seems to require a little more justification. The proof also allows one to bypass sophisticated uniqueness theorems for noncompact Ricci flows.

This result, together with the finite time extinction theorem in [P3] or [CM] imply the Poincaré conjecture.

As described above, some of the materials in the book are not directly related to the proof of the Poincaré conjecture. As an experienced worker solely interested in the Poincaré conjecture, one can start with Section 5.4, then go to Sections 6.1 and 6.2, and then read Chapters 7, 8 and 9.

Due to limited space and time, we will only provide in detail those results or proofs that are different from Perelman's original one. These include the proof of strong $\kappa$ noncollapsing with surgeries, backward limit of $\kappa$ solutions and the proof of the canonical neighborhood property with surgeries. Those parts which are more or less the same as the original ones are only sketched or referred to other sources.

Owing to the vastness and depth of the topic, some necessary selections of the material is necessary. This selection only reflects the current personal preference and limited knowledge and is in no way a snub to the materials that are left out. At the beginning of each section, some basic background material will be sketched with little or no proof. However standard references will be given. Coming to the proofs, we will strive to present the main ideas. Sometimes this is done at the price of sacrificing generality. The reader is led to references for further development and generalizations. Due to my limited knowledge and time constraint, the book contains many imperfections and omission. It is hoped that improvements will be made constantly. The author also welcomes all constructive suggestions and corrections, which can be sent to qizhang@math.ucr.edu.

To close the introduction, we list a number of notations to be used throughout the book. More notations will be introduced in each chapter or section.

$\mathbf{R}$ or $\mathbf{R^n}$: Euclidean space of dimension 1 or $n$; $\mathbf{M}$ or $M$: a Riemann manifold; $g$ or $g_{ij}$: the Riemann metric; $Ric$ or $R_{ij}$: the Ricci curvature; $Rm$: the full curvature tensor; $R$: the scalar curvature.

$\nabla$: covariant derivative or gradient; $\Delta$: Laplace operator; Hess: Hessian; $\nabla^2$, or $\nabla^2_{i,j}$, or $\nabla_i \nabla_j$: second covariant derivative.

$dx$, $dg$, $d\mu$ or $d\mu(g), dg(x,t)$ etc.: volume element; $d(x,y)$: distance; $d(x,y,t), d(x,y,g(t))$: distance with respect to metric $g(t)$; $B(x,r)$:

geodesic ball of radius $r$ centered at $x$; $B(x, r, t)$ or $B(x, r, g(t))$: geodesic ball of radius $r$ centered at $x$, with respect to metric $g(t)$; $|B(x, r)|$: volume of the ball $B(x, r)$ under a given metric; $|B(x, r, g(t))|_h$: volume of the ball under the metric $h$.

# Chapter 2

# Sobolev inequalities in the Euclidean space

In this chapter we will define the Sobolev space $W^{k,p}(D)$ where $D$ is a domain in $\mathbf{R}^n$ and prove a basic imbedding theorem. Here $k$ is a positive integer and $p \geq 1$. We will also present the closely related Poincaré inequality.

## 2.1 Weak derivatives and Sobolev space $W^{k,p}(D)$, $D \subset \mathbf{R}^n$

Let $u$ be an locally integrable function in $D$, a domain in $\mathbf{R}^n$. Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be a $n$-tuple of nonnegative integers. The symbol $D^\alpha$ denotes the differential operator $\partial_{x_1}^{\alpha_1} \ldots \partial_{x_n}^{\alpha_n}$.

**Definition 2.1.1** *A function $v \in L^1_{loc}(D)$ is called the $\alpha$-th weak derivative of $u$ if*

$$\int_D v\phi dx = (-1)^{|\alpha|} \int_D uD^\alpha \phi dx$$

*for all $\phi \in C_0^\infty(D)$. Here $|\alpha| = \alpha_1 + \cdots + \alpha_n$.*

**Definition 2.1.2** *For $p \geq 1$ and $k$ a nonnegative integer, we define the Sobolev space*

$$W^{k,p}(D) \equiv L^p(D) \cap \{u \mid D^\alpha u \in L^p(D),\ |\alpha| \leq k\},$$

*with the norm*

$$\|u\| = \|u\|_{k,p} = \left(\int_D \Sigma_{|\alpha|\le k}|D^\alpha u|^p dx\right)^{1/p}.$$

*The space $W_0^{k,p}(D)$ is defined as the closure of $C_0^\infty(D)$ under the above norm.*

In general, a function in $W^{k,p}(D)$ may not be smooth or even continuous. For example $u(x) = \ln(1/|x|)$ is in $W^{1,2}(B(0,1))$ where $B(0,1)$ is the ball of radius 1, centered at the origin in $\mathbf{R}^3$. Therefore, in studying these weakly differentiable functions, it is important to know that these functions can be approximated by smooth functions.

The first approximation result says that any $L^p$ function is the $L^p$ limit of a sequence of smooth functions.

**Proposition 2.1.1** *Let $u \in L^p(\mathbf{R}^n)$, $1 \le p < \infty$. For any, $\epsilon > 0$, define*

$$u_\epsilon(x) = \eta_\epsilon * u(x) = \int_{\mathbf{R}^n} \eta_\epsilon(x-y)u(y)dy,$$

*where $\eta_\epsilon = \epsilon^{-n}\eta(\cdot/\epsilon)$. Here the function $\eta$ is a nonnegative function in $C_0^\infty(B(0,1))$ such that $\int_{\mathbf{R}^n}\eta(x)dx = 1$.*

*Then $u_\epsilon \in C^\infty(\mathbf{R}^n) \cap L^p(\mathbf{R}^n)$ and $\|u_\epsilon - u\|_p \to 0$ when $\epsilon \to 0$.*

Proof. The derivative only hits the smooth function $\eta$. Hence $u_\epsilon$ is smooth. By Hölder's inequality

$$|u_\epsilon(x)| \le \left(\int \eta_\epsilon(x-y)dy\right)^{(p-1)/p} \left(\int \eta_\epsilon(x-y)|u(y)|^p dy\right)^{1/p}.$$

Integrating the $p$-th power of the above inequality and noting the total integral of $\eta_\epsilon$ is 1, we know that

$$\|u_\epsilon\|_p \le \|u\|_p.$$

Hence $u_\epsilon \in C^\infty(\mathbf{R}^n) \cap L^p(\mathbf{R}^n)$.

Next pick a continuous, compactly supported function $w$ such that $\|u - w\|_p < \delta$, an arbitrary positive number. Then

$$\begin{aligned} |w_\epsilon(x) - w(x)| &\le \int \eta_\epsilon(x-y)|w(y) - w(x)|dy \\ &\le \epsilon^{-n}\|\eta\|_\infty \int_{B(x,\epsilon)} |w(x) - w(y)|dy. \end{aligned}$$

Since $w$ is continuous and compactly supported, we know that

$$\|w_\epsilon - w\|_p < \delta$$

when $\epsilon$ is sufficiently small. Therefore

$$\|u_\epsilon - u\|_p \le \|u-w\|_p + \|w_\epsilon - w\|_p + \|w_\epsilon - u_\epsilon\|_p \le 2\|u-w\|_p + \|w_\epsilon - w\|_p < 3\delta.$$

□

**Remark 2.1.1** *One can choose*

$$\eta(x) = ce^{-1/(1-|x|^2)}, \quad |x| \le 1; \quad \eta(x) = 0, \quad |x| > 1.$$

*Here $c$ is the constant such that the total integral of $\eta$ is* 1*.*

The next proposition says that a function in $W^{k,p}(D)$ can be approximated in the $W^{k,p}$ norm by functions which are smooth in the interior of $D$. Note that here the norm is stronger than the $L^p$ norm in the previous proposition.

**Proposition 2.1.2** *The space $\{u \in C^\infty(D) \mid \|u\|_{k,p} < \infty\}$ is dense in $W^{k,p}(D)$.*

Proof. In order to prove the proposition, we need to recall a standard result called the *partition of unity* (for a proof, see [Zi] p53 e.g.)

*Let $\mathbf{J}$ be a collection of open subsets of $\mathbf{R}^n$ whose union contains a set $D$. There is a sequence of nonnegative functions $\{\phi_i\}$ in $C_0^\infty(\mathbf{R}^n)$ such that $0 \le \phi_i \le 1$ and*

*(i) for each $\phi_i$ there is a $U \in \mathbf{J}$ such that $supp\phi_i \subset U$,*

*(ii) for any compact $K \subset D$, there are only finitely many functions in $\{\phi_i\}$, whose supports intersect with $K$,*

*(iii) $\Sigma_{i\ge 1}\phi_i(x) = 1$ for each $x \in D$.*

Now we come back to the proof of the proposition. Let $\Omega_i$, $i = 1, 2, \ldots$, be a sequence of open subsets of $D$, whose union is $D$. The collection $\{\Omega_i - \bar{\Omega}_{i-1}\}$ is an open cover for $D$. Let $\{\phi_i\}$ be a partition of unity, subordinating to this cover. We can make the support of $\phi_i$ to be contained in $\Omega_i - \bar{\Omega}_{i-1}$.

Let $u \in W^{k,p}(D)$, $\epsilon > 0$. For each $i \ge 1$, by the last proposition, we can find $\epsilon_i > 0$ such that $supp(\phi_i u)_{\epsilon_i} \subset \Omega_i - \bar{\Omega}_{i-1}$ and $\|(\phi_i u)_{\epsilon_i} - \phi_i u\|_{k,p} < \epsilon 2^{-i}$. Define

$$v = \Sigma_{i\ge 1}(\phi_i u)_{\epsilon_i}.$$

Then $v \in C^\infty(D)$ since for each $x \in D$ the above is actually a finite sum. By Minkowski inequality

$$\|u - v\|_{k,p} \le \Sigma_{i\ge 1}\|(\phi_i u)_{\epsilon_i} - \phi_i u\|_{k,p} < \Sigma_{i\ge 1}\epsilon 2^{-i} = \epsilon.$$

□

**Remark 2.1.2** *In general functions in $W^{k,p}(D)$ can not be approximated by functions in $C^\infty(\bar{D})$. More conditions on the domain $D$ have to be imposed. See [Ad], Theorem 3.18 e.g.*

## 2.2 Main imbedding theorem for $W_0^{1,p}(D)$

**Theorem 2.2.1** ***(Sobolev Imbedding)*** *Let $\Omega$ be a domain in $\mathbf{R}^n$, then,*

$$W_0^{1,p}(\Omega) \hookrightarrow \begin{cases} L^{\frac{np}{n-p}}(\Omega), & 1 \le p < n; \\ L^{\varphi}(\Omega), & p = n, \varphi(t) = \exp(|t|^{p/(p-1)}) - 1, \quad \Omega \text{ bounded}; \\ C^{1-\frac{n}{p}}(\Omega), & p > n. \end{cases} \tag{2.2.1}$$

*Furthermore, there exists a constant $C = C_{(n,p)}$ such that, for any $u \in W_0^{1,p}(\Omega)$,*

$$\begin{cases} \|u\|_{np/(n-p)} \le C\,\|Du\|_p, & p < n; \\ \sup_\Omega |u| \le C|\Omega|^{1/n-1/p}\|Du\|_p, & p > n. \end{cases} \tag{2.2.2}$$

**Remark 2.2.1** *For the case $n = p$, $L^{\varphi(t)}(\Omega)$ is usually called an Orlicz space, where $\varphi(t) = \exp(|t|^{p/(p-1)}) - 1$. A function $u$ in this space satisfies $\int_\Omega[\exp(|u(x)|^{p/(p-1)}) - 1]dx < \infty$.*

We divide the proof of the theorem into three parts. The first one is:
*The case $1 \le p < n$.*

PROOF. We first establish the estimate for $C_0^1(\Omega)$ functions when $p = 1 : W_0^{1,1}(\Omega) \hookrightarrow L^{\frac{n}{n-1}}(\Omega)$. Denote any $x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$. Then for any $i, 1 \le i \le n$,

$$\begin{aligned} & u(x) = \int_{-\infty}^{x_i} \partial_i u(x_1, \ldots, x_{i-1}, s, x_{i+1}, \ldots, x_n)\, ds \\ \Rightarrow \quad & u(x) \le \int_{-\infty}^{+\infty} |\partial_i u(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)|\, dx_i \\ \Rightarrow \quad & |u(x)|^{n/(n-1)} \le \left( \prod_{i=1}^{n} \int_{\mathbf{R}} |\partial_i u|\, dx_i \right)^{1/(n-1)} . \end{aligned}$$

Here and later $\partial_i = \partial_{x_i}$. Now we integrate the above inequality successively over each $x_i$, $i = 1, \ldots, n$, and apply the generalized Hölder inequality with exponents $p_1 = p_2 = \ldots = p_{n-1} = n - 1$. In order to illustrate the idea, we carry out the calculation in detail for $n = 3$. Integrating with respect to $x_1$, we obtain,

$$\begin{aligned} \int_{\mathbf{R}} |u(x)|^{3/2}\, dx_1 \le & \left( \int_{\mathbf{R}} |\partial_1 u(x_1, x_2, x_3)|\, dx_1 \right)^{1/2} \cdot \\ & \left( \int_{\mathbf{R}} \left[ \int_{\mathbf{R}} |\partial_2 u(x_1, x_2, x_3)|\, dx_2 \right] \right. \\ & \times \left. \left[ \int_{\mathbf{R}} |\partial_3 u(x_1, x_2, x_3)|\, dx_3 \right] \right)^{1/2} dx_1 \\ \le & \left( \int_{\mathbf{R}} |\partial_1 u(x_1, x_2, x_3)|\, dx_1 \right)^{1/2} \cdot \\ & \left( \int_{\mathbf{R}} \int_{\mathbf{R}} |\partial_2 u(x_1, x_2, x_3)|\, dx_1 dx_2 \right)^{1/2} \cdot \\ & \left( \int_{\mathbf{R}} \int_{\mathbf{R}} |\partial_3 u(x_1, x_2, x_3)|\, dx_1 dx_3 \right)^{1/2} \end{aligned}$$

where we used $\int f^{\frac{1}{2}}\, g^{\frac{1}{2}} \le [\int f]^{\frac{1}{2}}\, [\int g]^{\frac{1}{2}}$.

Integrating the last inequality over $x_2$, we have

$$\begin{aligned}
&\int_{\mathbf{R}}\int_{\mathbf{R}} |u(x)|^{3/2}\,dx_1dx_2\\
&\le \left(\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_2 u(x_1,x_2,x_3)|\,dx_1dx_2\right)^{1/2}\cdot\\
&\qquad \left(\int_{\mathbf{R}}[\int_{\mathbf{R}} |\partial_1 u(x_1,x_2,x_3)|\,dx_1]\,[\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_3 u(x_1,x_2,x_3)|\,dx_1dx_3]\right)^{1/2} dx_2\\
&\le \left(\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_2 u(x_1,x_2,x_3)|\,dx_1dx_2\right)^{1/2}\cdot\\
&\qquad \left(\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_1 u(x_1,x_2,x_3)|\,dx_1dx_2\right)^{1/2}\cdot\\
&\quad \left(\int_{\mathbf{R}}\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_3 u(x_1,x_2,x_3)|\,dx_1dx_2dx_3\right)^{1/2}.
\end{aligned}$$

Further, integrating the above over $x_3$, we deduce

$$\begin{aligned}
\int_{\mathbf{R}^3} |u(x)|^{3/2}\,dx &\le \left(\int_{\mathbf{R}}\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_3 u(x_1,x_2,x_3)|\,dx_1dx_2dx_3\right)^{1/2}\cdot\\
&\quad \left(\int_{\mathbf{R}}\left[\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_2 u(x_1,x_2,x_3)|\,dx_1dx_2\right]^{1/2}\right.\\
&\quad \left.\times\left[\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_1 u(x_1,x_2,x_3)|\,dx_1dx_2\right]^{1/2} dx_3\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
\int_{\mathbf{R}^3} |u(x)|^{3/2}dx &\le \left(\int_{\mathbf{R}}\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_3 u(x_1,x_2,x_3)|\,dx_1dx_2dx_3\right)^{1/2}\\
&\qquad \cdot\left(\int_{\mathbf{R}}\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_2 u(x_1,x_2,x_3)|\,dx_1dx_2dx_3\right)^{1/2}\\
&\qquad \cdot\left(\int_{\mathbf{R}}\int_{\mathbf{R}}\int_{\mathbf{R}} |\partial_1 u(x_1,x_2,x_3)|\,dx_1dx_2dx_3\right)^{1/2}\\
&\le \left(\int_{\mathbf{R}}\int_{\mathbf{R}}\int_{\mathbf{R}} \frac{|\partial_1 u|+|\partial_2 u|+|\partial_3 u|}{3}\,dx_1dx_2dx_3\right)^{3/2}\\
&\le \left(\int_{\mathbf{R}}\int_{\mathbf{R}}\int_{\mathbf{R}} \sqrt{\frac{|\partial_1 u|^2+|\partial_2 u|^2+|\partial_3 u|^2}{3}}\,dx_1dx_2dx_3\right)^{3/2}\\
&= 3^{-\frac{3}{4}}\left(\int_{\mathbf{R}^3} |\nabla u(x)|\,dx\right)^{\frac{3}{2}}
\end{aligned}$$

where we have used the inequalities $(abc)^{\frac{1}{3}} \leq \frac{a+b+c}{3} \leq \sqrt{\frac{a^2+b^2+c^2}{3}}$. Thus,

$$\|u\|_{3/2} = \left(\int_{\mathbf{R}^3} |u(x)|^{3/2}\, dx\right)^{\frac{2}{3}} \leq \frac{1}{\sqrt{3}}\left(\int_{\mathbf{R}^3} |\nabla u(x)|\, dx\right).$$

In general, it holds, by induction,

$$\|u\|_{\frac{n}{n-1}} \leq \left(\prod_{i=1}^{n}\int_\Omega |\partial_i u|\, dx\right)^{\frac{1}{n}} \leq \frac{1}{n}\int_\Omega \sum_{i=1}^{n} |\partial_i u|\, dx \leq \frac{1}{\sqrt{n}}\|\nabla u\|_1$$

i.e.

$$\|u\|_{\frac{n}{n-1}} \leq \frac{1}{\sqrt{n}}\|\nabla u\|_1. \tag{2.2.3}$$

Thus we have proved the theorem for $p = 1$.

For other values of $p$, we proceed as follows. Replace $u$ in (2.2.3) by the power $|u|^\gamma$, we get for $\gamma > 1$,

$$\begin{aligned} \||u|^\gamma\|_{\frac{n}{n-1}} &\leq \frac{\gamma}{\sqrt{n}}\int_\Omega |u|^{\gamma-1}|\nabla u|\, dx \\ &\leq \frac{\gamma}{\sqrt{n}}\|\,|u|^{\gamma-1}\,\|_{p'} \cdot \|\nabla u\|_p, \quad \text{with} \qquad (p' = \frac{p}{p-1}). \end{aligned} \tag{2.2.4}$$

Therefore, for $p < n$, we may choose $\gamma$ to satisfy

$$\frac{n\gamma}{n-1} = \frac{(\gamma-1)p}{p-1} \quad \Rightarrow \quad \gamma = \frac{(n-1)p}{n-p} > 1.$$

Consequently,

$$\|u\|_{\frac{np}{n-p}} \leq \frac{\gamma}{\sqrt{n}}\,\|Du\|_p.$$

This proves the theorem for $p \in [1, np/(n-p))$ when $u$ is a $C^1$ function. The case for general $u$ follows from Proposition 2.1.2. □

*Proof of the theorem for the case* $p = n$.

The proof is due to Yudovich [Yu] and Trudinger [Tr1].

We will need some lemmas.

**Lemma 2.2.1** *Suppose* $u \in W_0^{1,1}(\Omega)$, *then,*

$$u(x) = \frac{1}{n\omega_n}\int_\Omega \sum_{i=1}^{n} \frac{(x_i - y_i)\partial_i u(y)}{|x-y|^n}\, dy \qquad a.e.\ in\ (\Omega).$$

*where* $\omega_n$ *is the volume of the unit ball in* $\mathbf{R}^n$.

PROOF. Suppose that $u \in C_0^1(\Omega)$ and extend $u$ to be zero outside $\Omega$. Then for any $\omega \in S^{n-1}$, $|\omega| = 1$, we have

$$u(x) = -\int_0^\infty \partial_r u(x + r\omega)\, dr. \tag{2.2.5}$$

Also notice $u(x) = \frac{1}{n\,\omega_n}\int_{S^{n-1}} u(x)\, d\omega$, substituting into (2.2.5), we have,

$$u(x) = \frac{1}{n\,\omega_n}\int_{S^{n-1}} \left( -\int_0^\infty \partial_r u(x + r\omega)\, dr \right) d\omega. \tag{2.2.6}$$

Denote $y = x + r\omega$, then $r = |x - y|$, $\omega = \frac{x-y}{|x-y|}$,

$$\partial_r u(x + r\omega) = \sum_i^n \frac{(y_i - x_i)\partial_i u(y)}{|x - y|}.$$

Plugging into (2.2.6),

$$\begin{aligned} u(x) &= -\frac{1}{n\,\omega_n}\int_\Omega \left( \sum_i^n \frac{(y_i - x_i)\partial_i u(y)}{|x - y|^n} \right) dy \\ &= \frac{1}{n\omega_n}\int_\Omega \frac{(x - y)\,\nabla u(y)}{|x - y|^n}\, dy. \end{aligned} \tag{2.2.7}$$

□

**Remark 2.2.2** *There is an alternative way to derive the above formula. For an integrable and smooth function $f$, the integral $\int_\Omega \Gamma(y - x) f(y) dy$ is called the Newtonian potential with density $f$. Here $\Gamma = \frac{c_n}{|x-y|^{n-2}}$ is the fundamental solution of the Laplacian. Using integration by parts, the equality $\Delta u = \Delta u$ yields,*

$$u(x) = -\int_\Omega \Gamma(y - x)\triangle u(y)\, dy = \int_\Omega \nabla_y \Gamma(y - x)\nabla u(y) dy.$$

**Lemma 2.2.2** *Let $\Omega$ be a bounded domain in $\mathbf{R}^n$, $\mu \in (0, 1]$ and define the operator $V_\mu$ on $L^1(\Omega)$ by the Riesz potential*

$$(V_\mu f)(x) = \int_\Omega |x - y|^{n(\mu-1)} f(y)\, dy.$$

*Then the operator $V_\mu$ maps $L^p(\Omega)$ continuously into $L^q(\Omega)$ for any $p \in [1, q]$ and $q$ satisfying*

$$0 \le \delta = \delta(p,q) = \frac{1}{p} - \frac{1}{q} < \mu.$$

*Furthermore, for any $f \in L^p(\Omega)$,*

$$\|V_\mu f\|_q \le \left(\frac{1-\delta}{\mu-\delta}\right)^{1-\delta} \omega_n^{1-\mu} \, |\Omega|^{\mu-\delta} \|f\|_p.$$

Before the proof of the lemma, we need to cite Young's inequality for integration. For more information on Young's inequality, see an updated discussion involving best constant in Lieb and Loss [LL].

**Proposition 2.2.1** *Let $f \in L^q(\mathbf{R}^n)$ and $k \in L^r(\mathbf{R}^n)$ with $\frac{1}{q} + \frac{1}{r} > 1$, then the function*

$$(k * f)(x) = \int_{\mathbf{R}^n} k(x-y) f(y) dy$$

*is defined for almost all $x$ and*

$$\|(k * f)\|_q \le \|k\|_r \|f\|_p \qquad \textit{where} \quad \frac{1}{q} = \frac{1}{p} + \frac{1}{r} - 1.$$

**Remark 2.2.3** *In a special case, $r = q, p = 1$, then we can prove $\|(k*f)\|_q \le \|k\|_q \|f\|_1$ as a consequence of Minkowski integral inequality: for any $1 \le q \le \infty$,*

$$\left[\int \left(\int |k(x-y)| \, dy\right)^q dx\right]^{\frac{1}{q}} \le \int \left(\int |k(x-y)|^q \, dx\right)^{\frac{1}{q}} dy.$$

*Denote $|f(y)|dy = d\omega(y)$, then,*

$$\begin{aligned}
\left[\int \left(\int |k(x-y)| d\omega(y)\right)^q dx\right]^{\frac{1}{q}} &\le \int \left(\int |k(x-y)|^q \, dx\right)^{\frac{1}{q}} d\omega(y) \\
&\le \int \left(\int |k(x-y)|^q \, dx\right)^{\frac{1}{q}} |f(y)| \, dy \\
&\le \int \|k\|_q |f(y)| \, dy = \|k\|_q \|f\|_1.
\end{aligned}$$

**Remark 2.2.4** *For a more special case,* $p = q = 2, r = 1$*, we can prove directly* $\|k * f\|_2 \le \|k\|_2 \, \|f\|_1$ *by the following two ways:*

$$
\begin{aligned}
(i) \qquad & \int \left( \int k(x-y) f(y) \, dy \right)^2 dx \\
& \le \int \left( \int |k(x-y)| \, |f(y)|^{\frac{1}{2}} \, |f(y)|^{\frac{1}{2}} dy \right)^2 dx \\
& \le \int \left( [\int |k|^2 \, |f| \, dy][\int |f| \, dy] \right) dx \\
& \le \|f\|_1 \int [\int |k|^2 \, |f| \, dy] \, dx \\
& \le \|f\|_1 \int [\int |k|^2 \; dx] |f| dy \\
& \le \|f\|_1^2 \, \|k\|_2^2 .
\end{aligned}
$$

$$
\begin{aligned}
(ii) \qquad & \left| \int \left( \int k(x-y) f(y) \, dy \right) \phi(x) \, dx \right| \\
& \le \int \left( \int |k(x-y) \phi(x)| \, dx \right) |f(y)| \, dy \\
& \le \left( \sup_y \int |k(x-y)\phi(x)| \, dx \right) \int |f(y)| \, dy \\
& \le \left( \sup_y \|k\|_2 \, \|\phi\|_2 \right) \|f\|_1 \\
& \le \|k\|_2 \, \|\phi\|_2 \, \|f\|_1 \qquad \textit{for } \forall \, \phi \in L^2(\mathbf{R}^n) \\
\Rightarrow \qquad & \|k * f\|_2 \le \|k\|_2 \, \|f\|_1 .
\end{aligned}
$$

PROOF. (**for Lemma 2.2.2**) Denote

$$h = |x-y|^{n(\mu-1)}.$$

Choose $r \ge 1$ such that

$$\frac{1}{r} = 1 + \frac{1}{q} - \frac{1}{p} \qquad \Rightarrow \qquad \frac{1}{r} + \frac{1}{p} = 1 + \frac{1}{q}$$

Then,

$$|h \, f| = \left( |h|^{\frac{r}{q}} \, |f|^{\frac{p}{q}} \right) |h|^{\frac{r(p-1)}{p}} |f|^{p\delta}$$

since $\frac{r}{q}+\frac{r(p-1)}{p}=r(\frac{1}{q}+1-\frac{1}{p})=\frac{r}{r}=1$, $\frac{p}{q}+p\delta=p(\frac{1}{q}+\frac{1}{p}-\frac{1}{q})=1$.

Now we can either use Hölder inequality (essentially reproving Young's inequality) or just apply Young's inequality to deal with $\|h*f\|_q$.

$(i)$ Apply Hölder inequality with three components $(q,\frac{p}{p-1},\frac{1}{\delta})$ satisfying $\frac{1}{q}+\frac{p-1}{p}+\delta=\frac{1}{q}+\frac{p-1}{p}+(\frac{1}{p}-\frac{1}{q})=1$. Thus,

$$\begin{aligned}\int h(x-y)\,|f(y)|\,dy \le \Big(\int \big(h(x-y)^{\frac{r}{q}}\,|f(y)|^{\frac{p}{q}}\big)^q\,dy\Big)^{\frac{1}{q}}&\\ \cdot\Big(\int \big(h(x-y)^{\frac{r(p-1)}{p}}\big)^{\frac{p}{p-1}}\,dy\Big)^{\frac{p-1}{p}}&\\ \cdot\Big(\int \big(|f(y)|^{p\delta}\big)^{\frac{1}{\delta}}\,dy\Big)^{\delta}&.\end{aligned}$$

Therefore,

$$\begin{aligned}\|h*f\|_q^q &= \|\int |h(\cdot-y)\,f(y)|\,dy\|_q^q \le \Big(\int |f(y)|^p\,dy\Big)^{q\,\delta}\\ &\quad\cdot\int\Big(\int h^r(x-y)\,|f(y)|^p\,dy\Big)\cdot\Big(\int h^r(x-y)\,dy\Big)^{\frac{q(p-1)}{p}}\,dx\\ &\le \|f\|_p^{pq\delta}\cdot\sup_{x\in\Omega}\Big(\int h^r(x-y)\,dy\Big)^{\frac{q(p-1)}{p}}\cdot\int\int h^r(x-y)\,|f(y)|^p\,dy\,dx\\ &\le \|f\|_p^{pq\delta}\cdot\sup_{x\in\Omega}\Big(\int h^r(x-y)\,dy\Big)^{\frac{q(p-1)}{p}}\cdot\|f\|_p^p\cdot\sup_{y\in\Omega}\Big(\int h^r(x-y)\,dy\Big)\\ &= \|f\|_p^{pq\delta+p}\cdot\sup_{x\in\Omega}\Big(\int h^r(x-y)\,dy\Big)^{\frac{q(p-1)}{p}+1}\\ &= \|f\|_p^q\cdot\|h\|_r^q\ \left(\text{because }\ p(q\delta+1)=p\frac{q}{p}=q,\ q(1-\frac{1}{p}+\frac{1}{q})=\frac{q}{r}\right)\end{aligned}$$

i.e.

$$\|h*f\|_q\le\|f\|_p\cdot\|h\|_r.$$

$(ii)$ Alternatively, applying directly Young's inequality, noting $\frac{1}{q}=\frac{1}{r}+\frac{1}{p}-1$, we also have

$$\|V_\mu f\|_q=\|h*f\|_q=\|\int_\Omega h\,f\,dy\|_q\le\|h\|_r\|f\|_p.$$

To finish the proof of the lemma, it suffices to show $\|h\|_r\le\left(\frac{1-\delta}{\mu-\delta}\right)^{1-\delta}\omega_n^{1-\mu}\,|\Omega|^{\mu-\delta}$. We consider the symmetrization of the domain

$\Omega$. Let $B_R(x)$ be the ball centered at $x$ with radius $R$ and with the same volume as $\Omega$, i.e. $|\Omega| = |B_R(x)| = \omega_n R^n$. Then $|\Omega - B_R(x)| = |B_R(x) - \Omega|$. Using the fact that

$$\begin{cases} |x-y|^{nr(\mu-1)} \le R^{nr(\mu-1)} \text{ for } y \in \Omega - B_R(x),\ 0<\mu<1; \\ |x-y|^{nr(\mu-1)} \ge R^{nr(\mu-1)} \text{ for } y \in B_R(x) \bigcap \Omega,\ 0<\mu<1. \end{cases}$$

We have

$$\begin{aligned}
\int_\Omega h^r(x-y)\,dy &= \int_{\Omega - B_R(x)} |x-y|^{nr(\mu-1)}\,dy + \int_{B_R(x)\bigcap\Omega} h^r(x-y)\,dy \\
&\le \int_{\Omega - B_R(x)} R^{nr(\mu-1)}\,dy + \int_{B_R(x)\bigcap\Omega} h^r(x-y)\,dy \\
&\le R^{nr(\mu-1)}|\Omega - B_R(x)| + \int_{B_R(x)\bigcap\Omega} h^r(x-y)\,dy \\
&= R^{nr(\mu-1)}|B_R(x) - \Omega| + \int_{B_R(x)\bigcap\Omega} h^r(x-y)\,dy \\
&\le \int_{B_R(x)-\Omega} |x-y|^{nr(\mu-1)}\,dy + \int_{B_R(x)\bigcap\Omega} |x-y|^{nr(\mu-1)}\,dy \\
&= \int_{B_R(x)} |x-y|^{nr(\mu-1)}\,dy
\end{aligned}$$

i.e.

$$\begin{aligned}
\|h\|_r^r = \int_\Omega h^r(x-y)\,dy &\le \int_{B_R(x)} |x-y|^{nr(\mu-1)}\,dy \\
&\le \int_0^R \int_{S^{n-1}} r^{nr\mu - nr} r^{n-1}\,d\omega dr \\
&\le \frac{n\,\omega_n}{n(r\mu - r + 1)} R^{n(r\mu - r+1)} \qquad \left(\text{because } \int_{S^{n-1}} d\omega = n\omega_n\right) \\
&\le \frac{\omega_n^{r(1-\mu)}}{r\mu - r + 1} |\Omega|^{r\mu - r + 1} \qquad (\text{because } |\Omega| = |B_R(x)| = \omega_n R^n) \\
&\le \left(\frac{1}{r(\mu-\delta)}\right)^{r(1-\delta)} \omega_n^{r(1-\mu)}\,|\Omega|^{r(\mu-\delta)} \\
&\qquad\qquad \left(\text{because } 1-\delta = \frac{1}{r} \Rightarrow r(1-\delta) = 1\right).
\end{aligned}$$

Therefore,

$$\|h\|_r \le \left(\frac{1-\delta}{\mu-\delta}\right)^{1-\delta} \omega_n^{1-\mu}\,|\Omega|^{\mu-\delta} \qquad \text{since } \frac{1-\delta}{\mu-\delta} = \frac{1}{r(\mu-\delta)}.$$

□

**Remark 2.2.5** *We have seen that for a domain $\Omega$, the ball $B_R(0)$ with $|B_R(0)| = |\Omega|$ is called a symmetric rearrangement of $\Omega$. Similarly, for an integrable function $f : \mathbf{R^n} \to \mathbf{R}$, vanishing sufficiently fast near infinity, one can define its symmetric decreasing rearrangement $f^*$ in the following manner. Let $D$ be a measurable set with finite measure. We denote by $D^*$ the symmetric rearrangement of $D$, centered at $0$. Let $\chi_D$ be the characteristic function of $D$. Then we define its symmetric decreasing rearrangement by*

$$\chi_D^* \equiv \chi_{D^*}.$$

*For $f$, its symmetric decreasing rearrangement is defined by*

$$f^*(x) = \int_0^\infty \chi^*_{\{|f|>t\}} dt.$$

*As an example, let us just mention the Riesz-Rearrangement inequality. Given three nonnegative functions $f, g, h \in C_0^\infty(\mathbf{R^n})$ , there holds:*

$$\begin{aligned} I(f,g,h) &\equiv \int_{\mathbf{R}^{2n}} f(x)g(x-y)h(y)\,dxdy \\ &\leq \int_{\mathbf{R}^{2n}} f^*(x)g^*(x-y)h^*(y)\,dxdy \equiv I(f^*,g^*,h^*) \end{aligned}$$

*where $f^*, g^*, h^*$ are the symmetric rearrangement for $f, g, h$ respectively.*

*For a proof of this inequality, one can consult the book [LL]. There are many interesting properties involving symmetric rearrangement. Some of them are useful in finding the best constant of Sobolev inequalities. See Section 2.4 for a brief discussion of this topic.*

**Lemma 2.2.3** *Let $f \in L^p(\Omega)$ and $g(x) = V_{\frac{1}{p}} f = \int_\Omega |x - y|^{n(\frac{1}{p}-1)} f(y)\,dy$. Then there exists constants $c_1, c_2$ depending only on $n$ and $p$ such that,*

$$\int_\Omega \exp\left(\frac{g}{c_1\|f\|_p}\right)^{p'} dx \leq c_2\,|\Omega|,\; p' = \frac{p}{p-1}$$

Proof. Apply the previous lemma by taking $\mu = \frac{1}{p}$, for any $q \geq p$, we have

$$\|g\|_q = \|V_{\frac{1}{p}} f\|_q \leq \left(q(1 - \frac{1}{p} + \frac{1}{q})\right)^{1-\frac{1}{p}+\frac{1}{q}} \omega_n^{1-\frac{1}{p}}\,|\Omega|^{\frac{1}{q}} \|f\|_p$$

$$\Rightarrow \quad \int_\Omega |g|^q dx \leq q^{q-\frac{q}{p}+1}\,\omega_n^{q-\frac{q}{p}}\,|\Omega|\,\|f\|_p^q \quad (\text{because } 0 < 1 - \frac{1}{p} + \frac{1}{q} \leq 1)$$

Now for any $q \geq p-1 \Rightarrow p'q = \frac{qp}{p-1} \geq p$, we can replace $q$ in the previous inequality by $p'q$,

$$\int_\Omega |g|^{p'q} dx \leq (p'q)\,(p'q\,\omega_n \|f\|_p^{p'})^q\,|\Omega| \tag{2.2.8}$$

where we have used $p'q - \frac{p'q}{p} + 1 = p'q(1-\frac{1}{p}) + 1 = q+1$.
We continue to take $q = [p], [p+1], \ldots, k, \ldots$ in the above inequality, for any $k \geq [p]$,

$$\int_\Omega |g|^{p'k} dx \leq (p'k)\,(p'k\,\omega_n \|f\|_p^{p'})^k\,|\Omega|$$

Then by the Taylor expansion of the exponential function,

$$\begin{aligned}
&\int_\Omega \exp\left(\frac{g}{c_1\|f\|_p}\right)^{p'} dx \\
&\qquad = \int_\Omega \sum_{k=0}^{[p]-1} \frac{1}{k!}\left(\frac{g}{c_1\|f\|_p}\right)^{p'k} dx + \int_\Omega \sum_{k=[p]}^{\infty} \frac{1}{k!}\left(\frac{g}{c_1\|f\|_p}\right)^{p'k} dx \\
&\qquad = \qquad (I) \qquad\qquad + \qquad\qquad (II)
\end{aligned}$$

To deal with the first part $(I)$, we can use Hölder inequality to increase the power of the function to be at least $[p]$, the other term is the fractional power of $|\Omega|$, which is dominated by $|\Omega|$. The key issue is to handle the second part $(II)$.

From inequality (2.2.8), for $k \geq [p]$,

$$\int_\Omega \frac{1}{k!}\left(\frac{g}{c_1\|f\|_p}\right)^{p'k} dx \leq (p'k)\frac{k^k}{k!}\left(\frac{p'\,\omega_n}{c_1^{p'}}\right)^k |\Omega|.$$

By Stirling's formula, $\frac{k^k}{k!}\sqrt{2\,\pi k} \approx e^k$ for large $k$, then,

$$\int_\Omega \frac{1}{k!}\left(\frac{g}{c_1\|f\|_p}\right)^{p'k} dx \leq (p'\sqrt{k})\left(\frac{e\,p'\,\omega_n}{c_1^{p'}}\right)^k |\Omega|.$$

We can choose sufficiently large $c_1$ such that $\frac{e\,p'\,\omega_n}{c_1^{p'}} < 1$, so the series will converge uniformly, and part $(II)$ will be dominated by some constant $c_2|\Omega|$. Thus we complete the proof. □

Now we prove the main imbedding theorem for the borderline case $p = n$. It shows that $W_0^{1,n}(\Omega) \hookrightarrow L^q(\Omega)$ is almost, but not quite $L^\infty(\Omega)$. One example is $u = \ln\ln|x|$ in $B(0,e) \subset \mathbf{R^n}$, $n \geq 2$. For clarity, we single out this borderline case as a new theorem.

**Theorem 2.2.2** *Let $u \in W_0^{1,n}(\Omega)$, then there exists constants $c_1$ and $c_2$ depending only on $n$ such that,*

$$\int_\Omega \exp\left(\frac{|u|}{c_1\|\nabla u\|_n}\right)^{\frac{n}{n-1}} dx \leq c_2\,|\Omega|.$$

Proof. Recall from (2.2.7),

$$\begin{aligned} u(x) &= \frac{1}{n\omega_n}\int_\Omega \frac{(x-y)\,\nabla u(y)}{|x-y|^n}\,dy \\ \Rightarrow \quad |u(x)| &\leq \frac{1}{n\omega_n}\int_\Omega |x-y|^{n(\frac{1}{n}-1)}|\nabla u(y)|\,dy \\ &= \frac{1}{n\omega_n}V_{\frac{1}{n}}|\nabla u| \qquad (\mu = \frac{1}{n}). \end{aligned}$$

For any $q \geq n$,

$$\Rightarrow \quad \|u\|_q \leq \frac{1}{n\omega_n}\|V_{\frac{1}{n}}|\nabla u|)\|_q \leq \left(\frac{1-\delta}{\frac{1}{n}-\delta}\right)^{1-\delta}\omega_n^{1-\frac{1}{n}}\,|\Omega|^{\frac{1}{n}-\delta}\|\nabla u\|_n < \infty.$$

Further, $p = n$, $p' = \frac{n}{n-1}$, by the previous lemma,

$$\int_\Omega \exp\left(\frac{|u|}{c_1\|\nabla u\|_n}\right)^{\frac{n}{n-1}} dx \leq \int_\Omega \exp\left(\frac{V_{\frac{1}{n}}\,|\nabla u|}{n\omega_n\,c_1\,\|\nabla u\|_n}\right)^{\frac{n}{n-1}} dx \leq c_2\,|\Omega|.$$

This establishes the theorem in the critical case $p = n$. □

*Proof of the theorem for case $p > n$.*

This is due to C. Morrey.

**Lemma 2.2.4** *Suppose $\Omega \subset \mathbf{R}^n$ is convex and $u \in W^{1,2}(\Omega)$. Then*

$$|u(x) - \bar{u}_\Omega| \leq \frac{d^n}{n|\Omega|}\int_\Omega |x-y|^{1-n}\,|\nabla u(y)|dy.$$

*Here $\bar{u}_\Omega$ is the average of $u$ in $\Omega$ and $d$ is the diameter of $\Omega$.*

Proof. This is similar to the proof of Lemma 2.2.1. We just observe that, for $x, y \in \Omega$,

$$u(x) - u(y) = -\int_0^{|x-y|} \partial_r u(x + r\omega) dr,$$

where $\omega = (y - x)/|y - x|$. Integrating over $\omega \in \mathbf{S}^{n-1}$, the lemma follows. □

The proof of the theorem in case $p > n$ can then be carried out in a few lines. Let $u \in W_0^{1,p}(\Omega)$, $p > n$. For any ball $B_R = B(x_0, R)$, with $x_0 \in \Omega$ and $R > 0$, when necessary, we extend $u$ to $B(x_0, R)$ by setting $u = 0$ outside of $\Omega$. By the last lemma

$$|u(x) - \bar{u}_{B_R}| \leq \frac{2R^n}{n|B_R|} \int_{B_R} |x - y|^{1-n} \, |\nabla u(y)| dy.$$

By Hölder's inequality

$$\begin{aligned} |u(x) - \bar{u}_{B_R}| &\leq \frac{2^n}{n w_n} \left( \int_{B_R} |x - y|^{(1-n)p/(p-1)} dy \right)^{(p-1)/p} \|\nabla u\|_p \\ &\leq c_{n,p} R^{1-(n/p)} \|\nabla u\|_p. \end{aligned}$$

Hence, for all $x, y \in B_R$, we have

$$|u(x) - u(y)| \leq C_{n,p} R^{1-(n/p)} \|\nabla u\|_p.$$

This completes the proof of Theorem 2.2.1. □

The imbedding in the borderline case $p = n$ can be improved. In the paper [Mo2], J. Moser proved the following theorem.

**Theorem 2.2.3** *Let $\alpha$ be any positive number less than or equal to $n|S^{n-1}|^{1/(n-1)}$. For each bounded domain $\Omega \subset \mathbf{R^n}$, there exists a positive constant $c = c(n, \alpha)$ such that*

$$\int_\Omega \exp\left(\alpha[|u(x)|/\|\nabla u\|_n]^{n/(n-1)}\right) dx \leq c|\Omega|$$

*for all $u \in C_0^\infty(\Omega)$.*

*On the other hand, if $\alpha > n|S^{n-1}|^{1/(n-1)}$, then*

$$\sup\{\int_\Omega \exp\left(\alpha|u(x)|^{n/(n-1)}\right) dx \,|\, u \in C_0^\infty(\Omega),\ \|\nabla u\|_n = 1\} = \infty.$$

Another inequality of similar spirit is the Onofri's inequality ( [On]):

**Theorem 2.2.4** *For each $u \in C^\infty(S^2)$, there holds*

$$\frac{1}{4\pi}\int_{S^2} e^{2u} d\mu \leq \exp\left(\frac{1}{4\pi}\int_{S^2}(|\nabla u|^2 + 2u)d\mu\right)$$

*with equality if and only if $e^{2u}g$ has constant curvature. Here $g$ is the standard metric on $S^2$ and $d\mu$ is the standard volume form on $S^2$.*

These borderline embedding results and their generalizations to various settings have found applications in such fields as conformal geometry and complex geometry. We refer the reader to the papers [DiTi], [LZ], [Ru2] and [Ru3] for some recent developments. The last three papers also contain new and simpler proofs of results such as Theorem 2.2.4.

## 2.3 Poincaré inequality and log Sobolev inequality

Next we present a weighted Poincaré inequality which looks similar to but weaker than the Sobolev inequality. It seems weaker because there is no gain in the integrability of a function over the integrability of its gradient. However, as we shall see in the next chapter, even a non-weighted Poincaré inequality actually implies a Sobolev inequality under the doubling condition of the metric balls of the ambient space. This result is due to Saloff-Coste and Grigor'yan [Sal2], [Gr2].

**Theorem 2.3.1** *Let $\eta$ be a nonnegative, continuous function in $\mathbf{R}^n$. Suppose $\eta$ has compact support $\Omega$, $\int_\Omega \eta dx = 1$, and the super level set $\{\eta \geq k\}$ is convex for all $k \geq 0$. Write $r$ as the diameter of $\Omega$ and $L = \int_\Omega u\eta dx$. Then for all $u \in W^{1,p}(\Omega)$, $p \geq 1$, there exists $C = C(n) > 0$ such that*

$$\int |u - L|^p \eta dx \leq C(n)\|\eta\|_\infty r^{n+p} \int |\nabla u|^p \eta dx.$$

Proof. By Jensen's inequality

$$\begin{aligned}\int |u - L|^p \eta dx &= \int \left| \int (u(x) - u(y))\eta(y)dy\right|^p \eta(x)dx \\ &\leq \int\int |u(x) - u(y)|^p \eta(x)\eta(y)dydx.\end{aligned} \tag{2.3.1}$$

Pick $x, y \in \Omega$, then $|x-y| \le r$ and

$$|u(x)-u(y)|^p = \left| \int_0^1 \nabla u(x+s(y-x)) \cdot (y-x) ds \right|^p$$
$$\le r^p \int_0^1 |\nabla u(x+s(y-x))|^p ds.$$

In the convexity assumption on $\eta$, we take $k = \min\{\eta(x), \eta(y)\}$. Thus $x, y$ are in the super level set $\{\eta \ge k\}$. Hence, for any $s \in [0,1]$, $x+s(y-x)$ is also in the super level set. Therefore

$$\eta(x+s(y-x)) \ge \min\{\eta(x), \eta(y)\}.$$

Hence

$$|u(x)-u(y)|^p \le r^p \int_0^1 |\nabla u(x+s(y-x))|^p \eta(x+s(y-x)) ds \, [\min\{\eta(x), \eta(y)\}]^{-1}.$$

It shows

$$|u(x)-u(y)|^p \eta(x)\eta(y) \le r^p \sup \eta \int_0^1 |\nabla u(x+s(y-x))|^p \eta(x+s(y-x)) ds.$$

Now we set $z = y - x$. Integrating the above inequality with respect to $x$ and $y$, we obtain

$$\int\int |u(x)-u(y)|^p \eta(x)\eta(y) dx dy$$
$$\le r^p \sup \eta \int_0^1 \int_{|z| \le r} \int |\nabla u(x+sz)|^p \eta(x+sz) dx dz ds$$
$$\le r^p \sup \eta \int_{|z| \le r} \int |\nabla u(x)|^p \eta(x) dx dz$$
$$= C(n) r^{n+p} \sup \eta \int |\nabla u(x)|^p \eta(x) dx.$$

The proof is finished by substituting this to the right-hand side of (2.3.1). □

The following log Sobolev inequality is another inequality which appears weaker than the Sobolev inequality.

**Theorem 2.3.2** *For all* $v \in W^{1,2}(R^n)$ *such that* $\|v\|_2 = 1$ *and all* $\epsilon > 0$, *there exists a constant* $c = c(n)$ *such that*

$$\int v^2 \ln v^2 \, d\mu \le \epsilon^2 \int |\nabla v|^2 d\mu - \frac{n}{2} \ln \epsilon^2 + c(n).$$

One can find in [Gro] the original proof of the theorem. A short proof of this theorem from the standard Sobolev inequality is presented in Theorem 4.2.1 below. On the other hand, this log Sobolev inequality actually implies the standard Sobolev inequality. We will come back to this fact in Chapter 4.

## 2.4 Best constants and extremals of Sobolev inequalities

The Sobolev inequality: for $u \in W^{1,p}(\mathbf{R}^n)$, $1 \le p < n$,

$$\|u\|_{np/(n-p)} \le C(n,p)\|\nabla u\|_p$$

can be rephrased as

$$\inf\{\|\nabla u\|_p \mid \|u\|_{np/(n-p)} = 1, u \in W^{1,p}(\mathbf{R}^n)\} = C^{-1}(n,p) > 0. \quad (2.4.1)$$

Two important questions thus arise. Question 1: What is the infimum? Question 2: Is the infimum reached by certain function(s)?

The questions was addressed by Talenti [Tal]. He showed that

$$C(n,p) = \pi^{-1/2} n^{-1/2} \left(\frac{p-1}{n-p}\right)^{1-(1/p)} \left[\frac{\Gamma(1+(n/2))\Gamma(n)}{\Gamma(n/p)\Gamma(1+n-(n/p))}\right]^{1/n}.$$

The basic idea is to show that the extremal is reached by a radially symmetric function and then work on the resulting ordinary differential equation. Indeed, let $u \in W^{1,p}(\mathbf{R}^n)$ and $u^*$ be the symmetric decreasing rearrangement. Then, for all $p \in [0,\infty)$,

$$\|\nabla u\|_p \ge \|\nabla u^*\|_p, \qquad \|u\|_p = \|u^*\|_p.$$

Therefore, if the infimum is reached by a function $u$, then $u$ must be radially symmetric with respect to some point. The remaining task is to find such an extremal function. We refer the details to the paper [Tal].

Let us finish this section with a little discussion about the minimizer equation. The derivative (at $u$) of the functional $I(u) = \int |\nabla u|^2 dx$ is a linear functional

$$DI(\phi) = \frac{d}{dt}|_{t=0} \int |\nabla(u+t\phi)|^2 dx = \int 2\nabla u \nabla \phi dx$$

$\phi \in W^{1,2}(\mathbf{R^n})$; and that for the functional $J(u) = \int u^{2n/(n-2)}dx$ is

$$DJ(\phi) = \frac{2n}{n-2} \int u^{(n+2)/(n-2)} \phi dx.$$

The Lagrange multiplier method tells us that

$$DI(\phi) = cDJ(\phi)$$

at a minimum $u$, for all $\phi \in W^{1,2}(\mathbf{R^n})$. That is

$$\int 2\nabla u \nabla \phi dx = c\frac{2n}{n-2} \int u^{(n+2)/(n-2)} \phi dx.$$

Hence $u$ satisfies the equation:

$$\Delta u + \lambda u^{(n+2)/(n-2)} = 0.$$

Here $\lambda$ is a positive constant.

In the 1980s and 1990s, a few authors finally proved that all positive solutions to this equation is radial around some point. See [GNN] for a proof with some extra conditions and [ChLi] for a short proof without extra conditions. In two dimension case, a similar result also holds for the corresponding minimizer equation ( [ChLi]). The proofs rely on the so-called moving plane method which is a clever application of the maximum principle for linear elliptic equations.

One can also ask similar questions about best constant and nature of extremal functions for the borderline case $p = n$. The understanding is much less satisfactory though. An interesting partial result can be found in [CC].

# Chapter 3

# Basics of Riemann geometry

## 3.1 Riemann manifolds, connections, Riemann metric

Riemann geometry is a vast area of mathematics, which can only be properly covered by numerous books. Many books of different levels and types have been written about Riemann, and more generally differential geometry. A quick search for the book title "Riemannian geometry" in Mathscinet turns up 80 results, and for the title "Differential geometry", 497 results.

The purpose of this chapter is merely to introduce some basic notations and concepts most closely related to the main theme: the study of Ricci flow. Let us just mention a few recent books and the most relevant references therein as main references for this section: [BSSG], [CwLx], [Cha2], [GHL], [Jo], [Pet], [SY].

A topological manifold is a Hausdorff topological space such that each point in the space is contained in an open set which is homeomorphic to some open set in $\mathbf{R}^n$, the $n$ dimensional Euclidean space.

**Definition 3.1.1** *(Smooth manifold) A smooth or $C^\infty$ manifold $\mathbf{M}$ is a topological manifold with a collection of smooth local charts in the following sense.*

*(i) Each chart is a pair $(U, \phi)$ where $U$ is an open set in $\mathbf{M}$ and $\phi$ is a homeomorphism from $U$ to an open set in $\mathbf{R}^n$.*

*(ii) The union of the open sets $U$ of all the charts in the collection is $\mathbf{M}$.*

*(iii) For any two charts* $(U_i, \phi_i)$, $i = 1, 2$, *the function below is* $C^\infty$.

$$\phi_1 \circ \phi_2^{-1} : \phi_2(U_1 \cap U_2) \to \phi_1(U_1 \cap U_2).$$

*(iv) The collection of the charts is maximal with the above properties.*

The maximal collection of charts in the definition is called a differential structure.

**Definition 3.1.2** *(atlas and orientable manifold)*
*An atlas for a smooth manifold* $\mathbf{M}$ *is a family of charts* $\{(U_i, \phi_i) \,|\, i \in I\}$ *such that* $\{U_i \,|\, i \in I\}$ *is an open cover of* $\mathbf{M}$. *Here* $I$ *is some set for indices.*

*An atlas* $\{(U_i, \phi_i) \,|\, i \in I\}$ *for a smooth manifold* $\mathbf{M}$ *is called oriented if all chart transition functions* $\phi_i \circ \phi_j^{-1}$, $i, j \in I$, *have positive definite functional determinant.*

*A smooth manifold is called orientable if it possesses an oriented atlas.*

**Definition 3.1.3** *(Tangent space) Let* $\mathbf{M}$ *be a smooth manifold and* $\mathbf{c} : \mathbf{R} \to \mathbf{M}$ *a smooth curve such that* $\mathbf{c}(0) = m$, *a point on* $\mathbf{M}$. *Let* $U$ *be a neighborhood of* $m$ *and* $f : U \to \mathbf{R}$ *be a smooth function. The tangent vector of* $\mathbf{c}$ *at* $m$ *is the operator of directional differentiation along* $\mathbf{c}$ *at* $m$, *called* $D_\mathbf{c}$ *or* $c'(t)$, *defined by*

$$D_\mathbf{c}(f) = \frac{d}{dt} f(\mathbf{c}(t))|_{t=0}.$$

*The linear space of all tangent vectors at* $m$ *is called the tangent space at* $m$ *and is denoted by* $T_m(\mathbf{M})$ *or* $T_m\mathbf{M}$.

**Definition 3.1.4** *(a canonical basis for* $T_m(\mathbf{M})$*) Let* $(U, \phi)$ *be a local chart around* $m \in \mathbf{M}$. *The tangent vector* $\frac{\partial}{\partial x^i} \in T_m(\mathbf{M})$ *is defined as*

$$\frac{\partial}{\partial x^i} f = \frac{\partial}{\partial x^i} (f \circ \phi^{-1})(x^1, \ldots, x^n)|_{\phi(m)},$$

*for all* $f \in C^\infty(U)$. *Here* $f \circ \phi^{-1}$ *is a smooth function whose domain is in* $\mathbf{R}^n$ *and* $(x^1, \ldots, x^n)$ *is the Euclidean coordinate in* $\mathbf{R}^n$.

*The tangent vectors* $\{\frac{\partial}{\partial x^i},\ i = 1, \ldots n\}$ *is a canonical basis for* $T_m(\mathbf{M})$.

**Definition 3.1.5** *(tangent bundle) For a smooth manifold* $\mathbf{M}$*, the set*

$$T(\mathbf{M}) \equiv \cup_{m\in\mathbf{M}} T_m(\mathbf{M}) = \{(m, v)\ |m \in \mathbf{M}, v \in T_m(\mathbf{M})\},$$

*equipped with the following structure of* $2n$ *dimensional smooth manifold, is called the tangent bundle of* $\mathbf{M}$*.*

*Let* $(U, \phi)$ *be a local chart around* $m \in \mathbf{M}$*. For* $v \in T_m(\mathbf{M})$*, we define the mapping*

$$\psi(v) = (\phi(m), \phi_*(v)).$$

*Here* $\phi_*(v)$ *is the tangent vector in* $\mathbf{R}^n$ *defined by*

$$\phi_*(v)(h) = v(h \circ \phi), \quad \text{for all} \quad h \in C^\infty(\mathbf{R}^n).$$

*The pair*

$$(\cup_{m\in U} T_m(\mathbf{M}), \psi)$$

*is a local chart of* $T(\mathbf{M})$*.*

**Remark 3.1.1** *By definition, for the local basis* $\{\frac{\partial}{\partial x^i}\}$*, we have*

$$\phi_*(\frac{\partial}{\partial x^i})(h) = \frac{\partial}{\partial x^i}(h \circ \phi) = \frac{\partial}{\partial x^i}(h \circ \phi \circ \phi^{-1}) = \frac{\partial}{\partial x^i}(h).$$

*Here the last term is just partial derivative in the* $x^i$ *direction in* $\mathbf{R^n}$*. Hence, in the canonical local coordinates for* $T_m(\mathbf{M})$ *and* $\mathbf{R^n}$*, the mapping* $\psi$ *takes the form*

$$\psi(m, a_i \frac{\partial}{\partial x^i}\Big|_m) = (\phi(m), a_1, \ldots, a_n).$$

**Definition 3.1.6** *(cotangent bundle) For a smooth manifold* $\mathbf{M}$ *and* $m \in \mathbf{M}$*, the dual space of* $T_m(\mathbf{M})$*, denoted by* $T_m(\mathbf{M})^*$ *is called the cotangent space at* $m$*. It is the space of all bounded linear functionals on* $T_m(\mathbf{M})$*. The set*

$$T(\mathbf{M})^* \equiv \cup_{m\in\mathbf{M}} T_m(\mathbf{M})^* = \{(m, \eta)\ |m \in \mathbf{M}, \eta \in T_m(\mathbf{M})^*\},$$

*equipped with the following structure of* $2n$ *dimensional smooth manifold, is called the cotangent bundle of* $\mathbf{M}$*.*

*Let* $(U, \phi)$ *be a local chart around* $m \in \mathbf{M}$*. For* $\eta \in T_m(\mathbf{M})^*$*, we define the mapping*

$$\psi(\eta) = (\phi(m), \phi^*(\eta)).$$

*Here $\phi^*(\eta)$ is the cotangent vector in $\mathbf{R}^n$ defined by*

$$\phi^*(\eta)(\phi_*(v)) = \eta(v), \quad \textit{for all} \quad v \in T_m(\mathbf{M}).$$

*The pair*

$$(\cup_{m\in U} T_m(\mathbf{M})^*, \psi)$$

*is a local chart of $T(\mathbf{M})^*$.*

*A canonical local coordinates for $T_m(\mathbf{M})^*$ is $\{dx^i\}$ where*

$$dx^i(\frac{\partial}{\partial x_j}) = \delta^i_j$$

*which is 1 if $i = j$ and 0 otherwise.*

**Definition 3.1.7** *($(p, q)$ tensor) Let $\mathbf{M}$ be a smooth manifold and $m \in \mathbf{M}$. A $p$ times covariant and $q$ times contravariant tensor, or $(p, q)$ tensor in short, is a $(p, q)$ linear form (bounded linear functional) on*

$$T_m(\mathbf{M}) \times \ldots \times T_m(\mathbf{M}) \times T_m(\mathbf{M})^* \times \ldots \times T_m(\mathbf{M})^*.$$

*The space of all the above $(p, q)$ tensors is denoted by $T^q_p(T_m(\mathbf{M}))$.*

*A canonical local coordinates for $T^q_p(T_m(\mathbf{M}))$ is*

$$\left\{dx^{i_1} \otimes \ldots \otimes dx^{i_p} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_q}}\right\}_{i_1,\ldots,i_p,j_1,\ldots,j_q}.$$

*Here $\otimes$ is the tensor product defined generally by*

$$u \otimes v(x, y) = u(x)v(y)$$

*where $x, y$ are elements of vector spaces $V$ and $W$ respectively, and $u, v$ are elements of the dual spaces $V^*$ and $W^*$ respectively.*

*The disjoint union of all $T^q_p(T_m(\mathbf{M}))$, $m \in \mathbf{M}$, denoted by $T^q_p(\mathbf{M})$, equipped with the natural smooth manifold structure defined in the same way as for tangent or cotangent bundles, is called $T^q_p$ tensor bundle.*

*A smooth section of $T^q_p(\mathbf{M})$ is called a tensor field.*

**Remark 3.1.2** *Let $(U, \phi)$ and $(U, \psi)$ be two local charts of $\mathbf{M}$ associated with the local coordinates $\{x^i\}$ and $\{y^i\}$ respectively. Let $\eta$ be a $(p, q)$ tensor field on $U$ represented locally by*

$$\eta = T^{j_1\ldots j_q}_{i_1\ldots i_p} dx^{i_1} \otimes \ldots \otimes dx^{i_p} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_q}},$$

$$\eta = \tilde{T}^{j_1 \ldots j_q}_{i_1 \ldots i_p} dy^{i_1} \otimes \ldots \otimes dy^{i_p} \otimes \frac{\partial}{\partial y^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial y^{j_q}}.$$

*Then*

$$\tilde{T}^{j_1 \ldots j_q}_{i_1 \ldots i_p} = T^{l_1 \ldots l_q}_{k_1 \ldots k_p} \frac{\partial x^{k_1}}{\partial y^{i_1}} \cdots \frac{\partial x^{k_p}}{\partial y^{i_p}} \frac{\partial y^{j_1}}{\partial x^{l_1}} \cdots \frac{\partial y^{j_q}}{\partial x^{l_q}}.$$

Let $\phi$ be a $C^1$ map from one smooth manifold $\mathbf{M}$ to another $\mathbf{N}$. The next definition introduces the derivative of $\phi$ and the induced map between vector fields or tensors on these manifolds.

**Definition 3.1.8** *(derivative of a map, push forward, pull back, differential of a diffeomorphism)*

*1. The derivative of $\phi$ at $m \in \mathbf{M}$, denoted by $D\phi(m)$ or $\phi_*$, is the linear map between $T_m\mathbf{M}$ and $T_{\phi(m)}\mathbf{N}$, given by*

$$D\phi(m)(X)f = X(f \circ \phi)$$

*for all $X \in T_m\mathbf{M}$ and all smooth $f$ defined in a neighborhood of $\phi(m)$.*

*$D\phi(m)(X)$ is called the push forward of $X$ by $\phi$.*

*2. $\phi^* : T^0_p(T_{\phi(m)}\mathbf{N}) \to T^0_p(T_m\mathbf{M})$, the pull back by $\phi$ on $(p, 0)$ tensors, is the linear map given by,*

$$(\phi^*\beta)(X_1, \ldots, X_p) = \beta(\phi_* X_1, \ldots, \phi_* X_p)$$

*for all $\beta \in T^0_p(T_{\phi(m)}\mathbf{N})$ and $X_i \in T_m\mathbf{M}$, $i = 1, \ldots, p$.*

*3. Suppose further that $\phi$ is a diffeomorphism. Then $\phi^* : T_{\phi(m)}\mathbf{N} \to T_m\mathbf{M}$ is defined to be $\phi_*^{-1}$. i.e. $\phi^*(Y) \equiv \phi_*^{-1}(Y)$ for $Y \in T_{\phi(m)}\mathbf{N}$.*

*In general, $\phi^* : T^q_p(T_{\phi(m)}\mathbf{N}) \to T^q_p(T_m\mathbf{M})$, the pull back or differential of $\phi$ on $(p, q)$ tensors, is defined by items 2 and 3 with the rule $\phi^*(\alpha \otimes \beta) = \phi^*(\alpha) \otimes \phi^*(\beta)$.*

The next three definitions introduce the concepts of differential forms which are anti-symmetric (skew symmetric) covariant tensors, and exterior derivatives on these tensors. The space of differential forms is then equipped with interesting algebraic structures. On the other hand, these concepts also have their origin in the theory of integration. Let us recall the Green's integration theorem for a bounded, smooth domain $D$ on $\mathbf{R^2}$.

*Let $P, Q$ be two smooth functions on $\bar{D}$. Then*

$$\int_{\partial D} (Pdx + Qdy) = \int_D (\partial_x Q - \partial_y P) dxdy.$$

The quantities $Pdx + Qdy$, $(\partial_x Q - \partial_y P)dxdy$ can be regarded as a differential 1 form and a differential 2 form on $\mathbf{R^2}$, which will be defined below. Moreover the later can be thought of as the exterior derivative of the former.

Using differential forms is the most convenient way to a coordinate-free definition of integration which is crucial for analysis on manifolds.

**Definition 3.1.9** *(antisymmetrization operator, exterior product) Let $V$ be a vector space and $V^*$ be its dual, the space of bounded linear functionals on $V$. Let $k$ be a positive integer. For an element $f \in \otimes^k V^*$, we define an antisymmetric multilinear functional*

$$(Antf)(v_1, \ldots, v_k) = \frac{1}{k!}\Sigma_p sign(p) f(v_{p(1)}, \ldots, v_{p(k)})$$

*where the sum is over all permutations $p$ of $\{v_1, \ldots v_k\} \subset V$.*

*The space of all antisymmetric elements in $\otimes^k V^*$ is denoted by $\wedge^k V^*$.*

*The exterior product of $f \in \wedge^i V^*$ with $g \in \wedge^j V^*$ is an element in $\wedge^{i+j} V^*$ defined by*

$$f \wedge g = \frac{(i+j)!}{i!j!} Ant(f \otimes g).$$

Examples. 1). If $f, g \in V^*$, then for $u, v \in V$, it holds

$$(f \wedge g)(u, v) = f(u)g(v) - f(v)g(u).$$

(2) Let $dx^1, \ldots, dx^n$ be a basis for $V^*$. Given $w^i = a^i_j dx^j \in V^*$, $i = 1, 2, \ldots n$, then

$$w^1 \wedge \ldots \wedge w^n = det(a^i_j) dx^1 \wedge \ldots \wedge dx^n. \tag{3.1.1}$$

For the proof, just notice that

$$w^1 \wedge \ldots \wedge w^n = a^1_{j_1} \ldots a^n_{j_n} dx^{j_1} \wedge \ldots \wedge dx^{j_n}.$$

By antisymmetry

$$\begin{aligned} w^1 \wedge \ldots \wedge w^n &= sign(j_1, \ldots, j_n) a^1_{j_1} \ldots a^n_{j_n} dx^1 \wedge \ldots \wedge dx^n \\ &= det(a^i_j) dx^1 \wedge \ldots \wedge dx^n. \end{aligned}$$

(3) Let $w^1, \ldots, w^k \in V^*$ and $v_1, \ldots, v_k \in V$, then

$$w^1 \wedge \ldots \wedge w^n(v_1, \ldots, v_k) = det(w^i(v_j)).$$

The proof is an easy exercise of induction.

**Exercise 3.1.1** *Prove the above statement (3).*

**Definition 3.1.10** *(differential forms, vector fields) A section of the cotangent bundle $T(\mathbf{M})^*$ over a smooth manifold $\mathbf{M}$ is a continuous map $f$: $\mathbf{M} \to T(\mathbf{M})^*$ such that $\pi(f(m)) = m$ for all $m$ in $\mathbf{M}$. Here $\pi$ is the projection map from $T(\mathbf{M})^*$ to $\mathbf{M}$, which sends all elements in $T_m(\mathbf{M})$ to $m$.*

*A smooth function on $\mathbf{M}$ is called a 0 form.*

*A section of the cotangent bundle $T(\mathbf{M})^*$ is called a one form.*

*A section of the tangent bundle $T(\mathbf{M})$ is called a vector field.*

*Let $q$ be a positive integer, a section of $\wedge^q T(\mathbf{M})^*$, the fiber bundle of skew-symmetric $q$-linear functionals on $T(\mathbf{M})$, is called a $q$ form.*

*A canonical local coordinates for $\wedge^q T(\mathbf{M})^*$ is*

$$\{dx^{i_1} \wedge \ldots \wedge dx^{i_q} \quad | \quad i_1 < \ldots < i_q\}.$$

The theory of differential forms was developed by E. Cartan. It provides an approach to differentiation and integration, which is independent of coordinates.

**Definition 3.1.11** *(exterior differentiation) The exterior derivative is a linear mapping from $\wedge^q T(\mathbf{M})^*$ to $\wedge^{q+1} T(\mathbf{M})^*$, defined inductively in the following manner.*

*Let $f \in C^\infty(\mathbf{M})$, then, for all vector field $X$ on $\mathbf{M}$,*

$$df(X) = X(f).$$

*Let $\eta \in \wedge^q T(\mathbf{M})^*$ be given in a local chart $(U, \phi)$ as*

$$\eta = \Sigma_{i_1 < \ldots < i_q} \eta_{i_1 \ldots i_q} dx^{i_1} \wedge \ldots \wedge dx^{i_q}.$$

*Then*

$$d\eta = \Sigma_{i_1 < \ldots < i_q} d\eta_{i_1 \ldots i_q} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_q}.$$

**Remark 3.1.3** *It is easy to see that in local coordinates $(U, \phi)$ with $\phi = (x^1, \ldots, x^n)$, it holds*

$$df = \frac{\partial f}{\partial x^i} dx^i \tag{3.1.2}$$

where $\frac{\partial f}{\partial x^i}$ is given in Definition 3.1.3.

Next we explain how to integrate differential forms on differential manifolds. We emphasize that this concept of integration is well defined even if the manifold is not equipped with a metric. At this moment integration is regarded as a linear functional on the space of smooth functions. After we provide the manifold with a metric, say a Riemannian one, then we will choose special forms so that the resulting integration is compatible with the metric. These special forms are referred as volume forms.

**Definition 3.1.12** *(integration of differential forms)*
*Let* $\mathbf{M}$ *be a smooth, orientable manifold and* $(U, \phi)$, $\phi(m) = (x^1(m), \ldots, x^n(m))$, $m \in U$, *be a local chart. Let* $w$ *be a* $n$*-form on* $\mathbf{M}$ *and there exists a smooth function* $f$ *on* $U$ *such that*

$$w|_U = f dx^1 \wedge \ldots \wedge dx^n.$$

*Here* $dx^i(\frac{\partial}{\partial x^j}) = \delta^i_j$, *i.e.* $\{dx^i\}$ *is the canonical basis for the cotangent space.*

*We define*

$$\int_U w = \int_{\phi(U)} f \circ \phi^{-1} dx^1 \ldots dx^n.$$

*Let* $\{(U_i, \phi_i)\}$ *be a family of local charts for* $\mathbf{M}$ *such that* $\{(U_i, h_i)\}$ *is a partition of unity for* $\mathbf{M}$, *then we define*

$$\int_{\mathbf{M}} w = \Sigma_i \int_{\mathbf{M}} h_i w = \Sigma_i \int_{U_i} h_i w.$$

**Remark 3.1.4** *One needs to prove that the above integration is independent of the choice of local charts, or the partition of unity. This is where we are using antisymmetry of forms. Let* $(U, \psi)$, $\psi = (y^1, , , , y^n)$ *be another local chart. These* $y^i$ *are smooth functions on* $\mathbf{M}$. *Therefore the local formula (3.1.2) tells us*

$$dx^i = \frac{\partial x^i}{\partial y^j}\, dy^j.$$

*Here again* $\frac{\partial x^i}{\partial y^j}$ *means* $\frac{\partial x^i \circ \psi^{-1}}{\partial y^j}$. *However they are regarded as functions on* $\mathbf{M}$ *but not as functions on* $\psi(U) \subset \mathbf{R^n}$. *More precisely, for* $m \in \mathbf{M}$

$$\left.\frac{\partial x^i}{\partial y^j}\right|_m = \left.\frac{\partial x^i \circ \psi^{-1}}{\partial y^j}\right|_p, \qquad \psi(m) = p.$$

*Hence the* $n$ *form in the definition can be written as*

$$w|_U = f\, \frac{\partial x^1}{\partial y^{j_1}}\, dy^{j_1} \wedge \ldots \wedge \frac{\partial x^n}{\partial y^{j_n}} dy^{j_n}.$$

*According to (3.1.1), we have*

$$w|_U = f det(\frac{\partial x^i}{\partial y^j}) dy^1 \wedge \ldots \wedge dy^n.$$

*By definition, in the local chart $(U, \psi)$,*

$$\int_U w = \int_{\psi(U)} f \circ \psi^{-1} det(\frac{\partial x^i}{\partial y^j}) dy^1 \ldots dy^n.$$

*Now we have returned to the usual change of variable formula in the Euclidean setting. There is one subtle change of notion though. Here the function $det(\frac{\partial x^i}{\partial y^j})$ is regarded as a function on $\psi(U) \subset \mathbf{R^n}$ via the definition $\frac{\partial x^i}{\partial y^j} = \frac{\partial x^i \circ \psi^{-1}}{\partial y^j}$. If we write $y = (y^1, \ldots, y^n)$ and $x = (x^1, \ldots, x^n)$, then it is clear that*

$$y = \psi \circ \phi^{-1}(x)$$

*since $\phi^{-1}(x)$ and $\psi^{-1}(y)$ are the same point in $\mathbf{M}$. Therefore,*

$$\begin{aligned}
\int_U w &= \int_{\psi(U)} f \circ \psi^{-1} det(\frac{\partial x^i}{\partial y^j}) dy^1 \ldots dy^n \\
&= \int_{\psi(U)} f \circ \psi^{-1}(y) det(\frac{\partial x^i \circ \psi^{-1}(y)}{\partial y^j}) dy^1 \ldots dy^n \\
&\qquad\qquad (note \quad x = \phi \circ \psi^{-1}(y) \ ) \\
&= \int_{\psi(U)} f \circ \psi^{-1}(y) det(\frac{\partial x}{\partial y}) dy^1 \ldots dy^n \\
&= \int_{\phi(U)} f \circ \psi^{-1}(\psi \circ \phi^{-1}(x)) dx^1 \ldots dx^n \\
&\qquad\qquad (Jacobi\ formula,\ Euclidean\ case) \\
&= \int_U f dx^1 \wedge \ldots \wedge dx^n.
\end{aligned}$$

*From here one can prove that the integral thus defined is independent of the choice of the local charts (coordinates). We leave the details as*

**Exercise 3.1.2** *Prove that integration thus defined is independent of the choice of local charts for the manifold.*

The next concept is called covariant derivative of a vector field, or more generally a tensor field. It is a generalization of directional derivative in the Euclidean case. The covariant derivative of a tensor field is also a tensor field which must satisfy the change of coordinates formula in Remark 3.1.2. Simply using the Euclidean directional derivative formula in the manifold case will not produce a tensor.

**Definition 3.1.13** *(covariant derivative, connections) Let $\Gamma(\mathbf{M})$ be the space of differentiable vector fields on a smooth manifold $\mathbf{M}$, and let $T(\mathbf{M})$ be the tangent bundle. A connection $\nabla$ is a map*

$$\nabla : T(\mathbf{M}) \times \Gamma(\mathbf{M}) \to \mathbf{T}(\mathbf{M})$$

*satisfying*

*(i) For $m \in \mathbf{M}$, $X \in T_m(\mathbf{M})$, $Y \in \Gamma(\mathbf{M})$, it holds $\nabla(X, Y) \in T_m(\mathbf{M})$,*

*(ii) $\nabla$ is linear,*

*(iii) For any differentiable function $f : \mathbf{M} \to \mathbf{R}$, it holds*

$$\nabla(X, fY) = X(f)Y + f\nabla(X, Y).$$

*(iv) For any $X, Y \in \Gamma(\mathbf{M})$, if $X$ is of class $C^k$ and $Y$ is of class $C^{k+1}$, for a positive integer $k$, then $\nabla(X, Y)$ is class $C^k$.*

*The covariant derivative of $Y$ with respect to $X$, denoted by $\nabla_X Y$, is $\nabla(X, Y)$.*

*Let $\eta$ be a differentiable one form, the covariant derivative of $\eta$ with respect to $X$ is defined as*

$$(\nabla_X \eta)(Y) = X(\eta(Y)) - \eta(\nabla_X Y)$$

*for all differentiable vector field $Y$.*

*(v) In general, let $T$ be a $(p, q)$ differentiable tensor field, and let $X_1, \dots X_p$ and $\eta_1, \dots, \eta_q$ be differentiable vector fields and one form respectively. The covariant derivative of $T$ with respect to $X$ is defined by*

$$\begin{aligned}(\nabla_X T)(X_1, \dots X_p, \eta_1, \dots, \eta_q) &= X(T(X_1, \dots X_p, \eta_1, \dots, \eta_q)) \\ &- T(\nabla_X X_1, X_2, \dots X_p, \eta_1, \dots, \eta_q)) - \dots \\ &- T(X_1, X_2, \dots X_p, \eta_1, \dots, \eta_{q-1}, \nabla_X \eta_q)).\end{aligned}$$

*Let $T$ be a $(p, q)$ differentiable tensor field, then $\nabla T$ can be regarded as a $(p + 1, q)$ tensor field via the above formula i.e.*

$$(\nabla T)(X, X_1, \dots X_p, \eta_1, \dots, \eta_q) \equiv (\nabla_X T)(X_1, \dots X_p, \eta_1, \dots, \eta_q).$$

**Remark 3.1.5** *The motivation behind the above definition is the Leibnitz rule for differentiation. The term* $X(T(X_1,\dots X_p,\eta_1,\dots,\eta_q))$ *is nothing but the directional derivative of the scalar function* $T(X_1,\dots X_p,\eta_1,\dots,\eta_q)$ *in the direction of* $X$.

**Definition 3.1.14** *(Riemann manifold) A Riemann manifold is a smooth manifold with a Riemann metric, a smooth, positive definite, symmetric* $(2,0)$ *tensor field. The Riemann metric is also called an inner product for vector fields.*

**Theorem 3.1.1** *(The fundamental theorem of Riemann geometry) There is a unique Riemann connection* $\nabla$ *satisfying the following conditions.*

*(i)* $\nabla$ *is a connection defined in Definition 3.1.13.*

*(ii)* $\nabla$ *is torsion free, i.e. for all differentiable* $X,Y\in\Gamma(\mathbf{M})$,

$$\nabla_X Y-\nabla_Y X=[X,Y].$$

*(iii) Let* $g$ *be the Riemann metric. For all differentiable* $X,Y,Z\in\Gamma(\mathbf{M})$,

$$Z(g(X,Y))=g(\nabla_Z X,Y)+g(X,\nabla_Z Y).$$

*Condition (iii) means that* $g$ *is parallel, i.e. the covariant derivative of* $g$ *with respect to any tangent vector is* $0$. *This is clear from the last formula in Definition 3.1.13, by taking* $T=g$.

*Proof of the theorem.* Let $X,Y,Z\in T(\mathbf{M})$. By condition (iii),

$$\begin{aligned}X(g(Y,Z))&=g(\nabla_X Y,Z)+g(Y,\nabla_X Z),\\ Y(g(Z,X))&=g(\nabla_Y Z,X)+g(Z,\nabla_Y X),\\ Z(g(X,Y))&=g(\nabla_Z X,Y)+g(X,\nabla_Z Y).\end{aligned}$$

Adding the first two identities and subtracting the last one, we deduce, after using condition (ii)

$$\begin{aligned}2g(\nabla_X Y,Z)&=X(g(Y,Z))+Y(g(Z,X))-Z(g(X,Y))\\ &\quad+g([X,Y],Z)-g([X,Z],Y)-g([Y,Z],X).\end{aligned}$$

Hence the covariant derivative is uniquely determined by $g$. □

It is often convenient to perform covariant differentiation on local coordinate systems. In the following we list some useful formulas concerning covariant derivatives in local systems.

**Proposition 3.1.1** *(local formulas for Riemann connection, the Christoffel symbols) Let $(U, \phi)$ be a local chart around a point $m \in \mathbf{M}$ and $\frac{\partial}{\partial x^i}$, $i = 1, \dots n$, be a local basis for $T_m(\mathbf{M})$. Then*

$$\nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^k} = \Gamma^i_{jk} \frac{\partial}{\partial x^i},$$

*where*

$$\Gamma^i_{jk} = \frac{1}{2} g^{il} (\partial_j g_{kl} + \partial_k g_{lj} - \partial_l g_{jk}) \tag{3.1.3}$$

*is called the Christoffel symbols. Here $\partial_j g_{kl} \equiv \frac{\partial}{\partial x^j} g_{kl}$, etc. Also $(g^{il})$ is the inverse of $(g_{il})$.*

*Let $T = T^{j_1 \dots j_q}_{i_1 \dots i_p} dx^{i_1} \otimes \dots \otimes dx^{i_p} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{j_q}}$ be a smooth $(p, q)$ tensor. Then*

$$\nabla_{\partial / \partial x^k} T = T^{j_1 \dots j_q}_{i_1 \dots i_p, k} dx^{i_1} \otimes \dots \otimes dx^{i_p} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{j_q}}, \tag{3.1.4}$$

*where*

$$T^{j_1 \dots j_q}_{i_1 \dots i_p, k} = \frac{\partial T^{j_1 \dots j_q}_{i_1 \dots i_p}}{\partial x^k} - \Sigma^p_{r=1} \Gamma^l_{i_r k} T^{j_1 \dots j_q}_{i_1 \dots i_{r-1} l i_{r+1} \dots i_p} + \Sigma^q_{s=1} \Gamma^{j_s}_{lk} T^{j_1 \dots j_{s-1} l j_{s+1} \dots j_q}_{i_1 \dots i_p}.$$

Proof. The above formula for Christoffel symbols is a direct consequence of the fundamental theorem of Riemann geometry. One just needs to observe that the Lie bracket

$$[\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}] = 0.$$

Formula (3.1.4) follows from item v) in Definition 3.1.13 □

**Remark 3.1.6** *The notation $T^{j_1 \dots j_q}_{i_1 \dots i_p, k}$ stands for the $\{i_1 \dots i_p, j_1 \dots j_q\}$ component of the tensor $\nabla_{\partial / \partial x^k} T$. It does not mean the derivative of the corresponding component of the tensor $T$ with respect to $x^k$. However in local orthonormal coordinates (Definition 3.4.6 below), $\Gamma^i_{jk} = 0$ at the origin. Therefore at this one point, these two quantities are equal.*

Diffeomorphisms and vector fields are important objects in differential geometry and topology. A smooth vector field $X$ on a manifold $\mathbf{M}$ generates a one parameter family of diffeomorphisms $\phi_t$ via the differential equation

$$\frac{d\phi_t(p)}{dt} = X(\phi_t(p)), \quad p \in \mathbf{M}. \tag{3.1.5}$$

Here, as usual $\frac{d\phi_t(p)}{dt}$ is regarded as the tangent vector of the curve $c = c(t) = \phi_t(p)$ at the point $\phi_t(p)$. Naturally one would like to understand how does a vector or tensor field change under $\phi_t$. This is described by a differential operator for tensor fields, called the Lie derivative.

We have

**Definition 3.1.15** *(Lie derivative) Let $\alpha$ be a smooth tensor field on $\mathbf{M}$. The Lie derivative of $\alpha$ with respect to $X$ is the tensor field*

$$L_X\alpha \equiv \lim_{h\to 0}\frac{\phi_h^*\alpha - \alpha}{h} = \left.\frac{d\phi_h^*\alpha}{dh}\right|_{h=0}.$$

*Here $\frac{d\phi_t(p)}{dt} = X(\phi_t(p))$, $t > 0$, $\phi_0(p) = p \in \mathbf{M}$. $\phi_h^*$ is the differential of $\phi_h$ acting on the tensor field $\alpha$, defined in Definition 3.1.8.*

A simple and useful observation is the following

**Proposition 3.1.2** *Let $\phi_t$ and $\eta_s$ be the one parameter family of diffeomorphisms generated by smooth vector fields $X$ and $Y$ respectively, i.e. for $p, q \in \mathbf{M}$,*

$$\frac{d\phi_t(p)}{dt} = X(\phi_t(p)), \qquad \frac{d\eta_s(q)}{ds} = Y(\eta_s(q)).$$

*Then $\psi_s \equiv \phi_t \circ \eta_s \circ \phi_{-t}$ is the one parameter family of diffeomorphisms generated by $(\phi_t)_*Y$.*

Proof. Pick a smooth function $f \in C^\infty(\mathbf{M})$. Using (3.1.6) below repeatedly, we have, for a point $p \in \mathbf{M}$,

$$\begin{aligned}\frac{d\psi_s(p)}{ds}(f) &= \frac{d}{ds}f(\psi_s(p)) = \frac{d}{ds}f(\phi_t \circ \eta_s \circ \phi_{-t}(p))\\ &= \frac{d}{ds}(f\circ\phi_t)(\eta_s(\phi_{-t}(p))) = \frac{d}{ds}(\eta_s(\phi_{-t}(p)))(f\circ\phi_t)\\ &= Y(\eta_s(\phi_{-t}(p)))(f\circ\phi_t).\end{aligned}$$

On the other hand, by Definition 3.1.8,

$$((\phi_t)_*Y)(\psi_s(p))f = Y(\phi_{-t}(\psi_s)(p))(f\circ\phi_t) = Y(\eta_s(\phi_{-t}(p)))(f\circ\phi_t).$$

Thus

$$\frac{d\psi_s(p)}{ds} = ((\phi_t)_*Y)(\psi_s(p)),$$

which proves the proposition. □

Practical ways of computing Lie derivatives are provided by the following

**Proposition 3.1.3** *(i) for $f \in C^\infty(\mathbf{M})$, $L_X f = X(f)$;*

*(ii) For a smooth vector field $Y$, $L_X Y = [X, Y] = XY - YX$;*

*(iii) For a $(p, 0)$ tensor $\alpha$,*

$$(L_X\alpha)(X_1, \ldots, X_p) \\ = X(\alpha(X_1, \ldots, X_p)) - \Sigma_{i=1}^p \alpha(X_1, \ldots, X_{i-1}, [X, X_i], X_{i+1}, \ldots, X_p).$$

*(iv) Let $D_X\alpha$ be the covariant derivative of $\alpha$ with respect to $X$. Then*

$$(L_X\alpha)(X_1, \ldots, X_p) \\ = (D_X\alpha)(X_1, \ldots, X_p) + \Sigma_{i=1}^p \alpha(X_1, \ldots, X_{i-1}, D_{X_i}X, X_{i+1}, \ldots, X_p).$$

Proof. (ii) During the proof we need to use the simple formula for the tangent of a smooth curve $c = c(s)$. For all smooth functions $f$ on $\mathbf{M}$,

$$\frac{dc(s)}{ds}(f) = \frac{d}{ds}(f(c(s))). \tag{3.1.6}$$

Let $\phi_t$ and $\eta_s$ be the one parameter family of diffeomorphisms generated by $X$ and $Y$ respectively, i.e. for $p, q \in \mathbf{M}$,

$$\frac{d\phi_t(p)}{dt} = X(\phi_t(p)), \qquad \frac{d\eta_s(q)}{dt} = Y(\eta_s(q)).$$

By the above definition

$$\begin{aligned} L_X Y &= \lim_{t\to 0} \frac{\phi_t^* Y - Y}{t} = \lim_{t\to 0} \frac{(\phi_t)_*^{-1} Y - Y}{t} \\ &= \lim_{t\to 0} \frac{(\phi_{-t})_* Y - Y}{t} = -\lim_{t\to 0} \frac{(\phi_t)_* Y - Y}{t}. \end{aligned}$$

Here we have used Definition 3.1.8 for $\phi_t^*$: the pull back on vector fields, and the semigroup property $(\phi_t)^{-1} = \phi_{-t}$.

Since $p$ is an arbitrary point on $\mathbf{M}$, we can just prove the result for $s = 0$. Using Proposition 3.1.2, we compute, at $t = 0$, the derivatives

$$\begin{aligned} &\frac{d((\phi_t)_* Y)(p) f}{dt} \\ &= \frac{d}{dt}\frac{d}{ds} f(\phi_t \circ \eta_s \circ \phi_{-t}(p)) = \frac{d}{ds}\frac{d}{dt} f(\phi_t \circ \eta_s \circ \phi_{-t}(p)) \\ &= \frac{d}{ds} \lim_{t\to 0} \frac{f(\phi_t \circ \eta_s \circ \phi_{-t}(p)) - f(\eta_s \circ \phi_{-t}(p)) + f(\eta_s \circ \phi_{-t}(p)) - f(\eta_s(p))}{t} \\ &= \frac{d}{ds} \lim_{t\to 0} \frac{[f(\phi_t \circ \eta_s) - f(\eta_s)] \circ \phi_{-t}(p)}{t} + \frac{d}{ds} \lim_{t\to 0} \frac{f(\eta_s \circ \phi_{-t}(p)) - f(\eta_s(p))}{t}. \end{aligned}$$

Hence, at $t = 0$,

$$\begin{aligned}
&\frac{d((\phi_t)_*Y)(p)f}{dt} \\
&= \frac{d}{ds}\lim_{t\to 0}\frac{[f(\phi_t \circ \eta_s) - f(\eta_s)](p)}{t} + \frac{d}{ds}\lim_{t\to 0}\frac{f(\eta_s \circ \phi_{-t}(p)) - f(\eta_s(p))}{t} \\
&= \frac{d}{ds}\frac{d}{dt}f(\phi_t \circ \eta_s(p)) + \frac{d}{ds}\frac{d}{dt}f(\eta_s \circ \phi_{-t}(p)) \\
&\equiv I + II.
\end{aligned}$$

Here we have used the fact that $\phi_0$ is the identity map.

By (3.1.6) again, we have, at $s = t = 0$,

$$\begin{aligned}
I &= \frac{d}{ds}\frac{d}{dt}f(\phi_t \circ \eta_s(p)) = \frac{d}{ds}\frac{d\phi_t(\eta_s(p))}{dt}f = \frac{d}{ds}X(\phi_t(\eta_s(p)))f \\
&= \frac{d}{ds}X(\eta_s(p))f = Y|_{\eta_s(p)}(X(f)) = Y|_p(X(f)).
\end{aligned}$$

Similarly, when $s = t = 0$,

$$\begin{aligned}
II &= \frac{d}{dt}\frac{d}{ds}f(\eta_s \circ \phi_{-t}(p)) = -\frac{d}{dt}\frac{d}{ds}f(\eta_s \circ \phi_t(p)) = -\frac{d}{dt}\frac{d(\eta_s(\phi_t(p))}{ds}f \\
&= -\frac{d}{dt}Y(\eta_s(\phi_t(p)))f = -\frac{d}{dt}Y(\phi_t(p))f = -X(p)(Y(f)).
\end{aligned}$$

Therefore

$$(L_X Y)(p) = -\frac{d((\phi_t)_*Y)(p)}{dt} = [X, Y](p).$$

This proves (ii).

The proof of (iii) follows easily from (ii) and Definition 3.1.8 part 3. Indeed, for a point $m \in \mathbf{M}$, and $t = 0$,

$$\begin{aligned}
&(L_X\alpha)(m)(X_1, \ldots, X_p) = \frac{d[\phi_t^*\alpha(m)(X_1, \ldots, X_p)]}{dt} \\
&= \frac{d}{dt}[\alpha(\phi_t(m))((\phi_t)_*X_1, \ldots, (\phi_t)_*X_p)] \\
&= \frac{d}{dt}[\alpha(\phi_t(m))(X_1, \ldots, X_p)] \\
&\quad + \Sigma_{i=1}^p \alpha(\phi_t(m))(X_1, \ldots, X_{i-1}, \frac{d}{dt}((\phi_t)_*X_i), X_{i+1}, \ldots, X_p) \\
&= X(m)(\alpha(X_1, \ldots, X_p)) - \Sigma_{i=1}^p \alpha(m)(X_1, \ldots, X_{i-1}, L_X X_i, X_{i+1}, \ldots, X_p).
\end{aligned}$$

Finally (iv) is a direct consequence of (iii), the formula for covariant derivatives

$$\begin{aligned}
X(\alpha(X_1, \ldots, X_p)) &= (D_X\alpha)(X_1, \ldots, X_p) \\
&\quad + \Sigma_{i=1}^p \alpha(m)(X_1, \ldots, X_{i-1}, D_X X_i, X_{i+1}, \ldots, X_p)
\end{aligned}$$

and the torsion-free condition

$$D_X X_i - D_{X_i} X = [X, X_i].$$

See Definition 3.1.13 and Theorem 3.1.1. □

Sometimes we need to deal with the Lie derivative involving time dependent vector fields and the associated one parameter family of diffeomorphisms. For example, such a need arises in the study for Ricci flow. The following fact will be useful in Section 5.4 where we will discuss Ricci solitons.

**Proposition 3.1.4** *Let $X = X(t)$, $t > 0$, be a family of smooth vector fields on $\mathbf{M}$, depending smoothly on $t$. Denote by $\phi_{t,s}$ the family of diffeomorphisms associated with $X(t)$. That is*

$$\begin{cases} \frac{d\phi_{t,s}(p)}{dt} = X(t)(\phi_{t,s}(p)), & t > s \geq 0, \\ \phi_{s,s}(p) = p, \qquad p \in \mathbf{M}. \end{cases}$$

*Then the following conclusions are true.*

*(i) If $\alpha$ is a smooth tensor field on $\mathbf{M}$, then*

$$L_{X(t)}\alpha = \lim_{h \to 0} \frac{\phi^*_{h+t,t}\alpha - \alpha}{h}.$$

*(ii) $\frac{d}{dt}(\phi^*_{t,0}\alpha) = \phi^*_{t,0}(L_{X(t)}\alpha)$.*

**Remark 3.1.7** *If no confusion arises, one can write $\phi_{t,0}$ as $\phi_t$.*

Proof. (i) By definition

$$L_{X(t)}\alpha = \lim_{h \to 0} \frac{\eta^*_h\alpha - \alpha}{h}$$

where $\eta_h$ is the one parameter family of diffeomorphisms generated by $X(t)$ for the fixed time $t$, i.e. $\frac{d}{dh}\eta_h = X(t)(\eta_h)$. Therefore we just need to show that

$$\lim_{h \to 0} \frac{\phi^*_{h+t,t}\alpha - \eta^*_h\alpha}{h} = 0.$$

This is equivalent to showing that

$$\lim_{h \to 0} \frac{(\phi_{h+t,t})_* Z - (\eta_h)_* Z}{h} = 0$$

for any smooth vector fields $Z$. Choose a smooth function $f$ on $\mathbf{M}$. Then

$$\lim_{h\to 0}\frac{(\phi_{h+t,t})_*Z-(\eta_h)_*Z}{h}f=Z(\lim_{h\to 0}\frac{f\circ\phi_{h+t,t}-f\circ\eta_h}{h}).$$

Therefore we only have to show

$$\lim_{h\to 0}\frac{f\circ\phi_{h+t,t}-f\circ\eta_h}{h}=0. \tag{3.1.7}$$

By definition, for a point $p\in\mathbf{M}$,

$$\frac{d}{dt}(f\circ\phi_{t,s}(p))=\frac{d\phi_{t,s}(p)}{dt}f=X(t)(\phi_{t,s}(p))f.$$

Hence

$$f\circ\phi_{h+t,t}(p)=f(p)+\int_0^h X(t+l)(\phi_{t+l,t}(p))f\,dl.$$

In the same manner

$$f\circ\eta_h(p)=f(p)+\int_0^h X(t)(\eta_l(p))f\,dl.$$

Taking the difference of the last two identities, we deduce

$$\begin{aligned}&\lim_{h\to 0}\frac{f\circ\phi_{h+t,t}(p)-f\circ\eta_h(p)}{h}\\ &\quad=\lim_{h\to 0}\frac{1}{h}\int_0^h[X(t+l)(\phi_{t+l,t}(p))-X(t)(\eta_l(p))]f\,dl=0,\end{aligned}$$

which proves (3.1.7) and consequently statement (i).

Next we prove (ii).

By the semigroup property $\phi_{h+t,0}=\phi_{h+t,t}\circ\phi_{t,0}$, there hold

$$\begin{aligned}\frac{d}{dt}\phi^*_{t,0}\alpha&=\lim_{h\to 0}\frac{\phi^*_{h+t,0}\alpha-\phi^*_{t,0}\alpha}{h}\\ &=\lim_{h\to 0}\frac{(\phi_{h+t,t}\circ\phi_{t,0})^*\alpha-\phi^*_{t,0}\alpha}{h}=\lim_{h\to 0}\frac{\phi^*_{t,0}\circ\phi^*_{h+t,t}\alpha-\phi^*_{t,0}\alpha}{h}\\ &=\phi^*_{t,0}\left(\lim_{h\to 0}\frac{\phi^*_{h+t,t}\alpha-\alpha}{h}\right)=\phi^*_{t,0}(L_{X(t)}\alpha).\end{aligned}$$

The last step is by (i). Hence statement (ii) is true. □

**Remark 3.1.8** *Let $\alpha$ be a $(p, 0)$ tensor field on a Riemann manifold* $\mathbf{M}$*. Let $X$ be a vector field on* $\mathbf{M}$ *and $\phi$ a diffeomorphism on* $\mathbf{M}$*. Then*

$$\phi^*(L_X\alpha) = L_{\phi_*^{-1}X}\phi^*\alpha.$$

**Exercise 3.1.3** *Prove the formula in the remark.*

## 3.2 Second covariant derivatives, commutation formulas, curvatures

Now we define the second covariant derivative. Let $X, Y$ and $Z$ be smooth vector fields on $\mathbf{M}$. By definition $\nabla Z$ is a $(1, 1)$ tensor field satisfying

$$(\nabla Z)(Y) = \nabla_Y Z.$$

Hence

$$\nabla_X[(\nabla Z)(Y)] = \nabla_X(\nabla_Y Z).$$

A differentiation should follow the Leibnitz rule. So we require

$$(\nabla_X(\nabla Z))(Y) + (\nabla Z)(\nabla_X Y) = \nabla_X(\nabla_Y Z).$$

i.e.

$$(\nabla_X(\nabla Z))(Y) = \nabla_X(\nabla_Y Z) - (\nabla Z)(\nabla_X Y) = \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z.$$

The left-hand side gives rise to the second covariant derivative of a vector field.

**Definition 3.2.1** *(second covariant derivative, Hessian of a function) Let $Z$ be a smooth vector field on* $\mathbf{M}$*. Then the second covariant derivative of $Z$ is a $(2, 1)$ tensor field, denoted by $\nabla^2 Z$, defined by*

$$\nabla^2_{X,Y} Z \equiv (\nabla_X(\nabla Z))(Y) \equiv \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z$$

*for all smooth vector fields $X$ and $Y$.*

*In general, let $T$ be a smooth $(p, q)$ tensor field, then its second covariant derivative is a $(p+2, q)$ tensor field, denoted by $\nabla^2 T$, defined by*

$$\nabla^2 T(X, Y) = \nabla^2_{X,Y} T = (\nabla_X(\nabla T))(Y) = \nabla_X(\nabla_Y T) - \nabla_{\nabla_X Y} T$$

*for all smooth vector fields $X$ and $Y$.*

*In particular, if $T$ is a smooth, scalar function $f$, then $\nabla^2 f$ is called the Hessian of $f$.*

**Remark 3.2.1** *It is important to realize the relation between the second covariant derivative* $\nabla^2_{X,Y}T$ *and the iterated covariant derivative* $\nabla_X(\nabla_Y T)$. *From the above definition, the two differ by* $\nabla_{\nabla_X Y}T$. *However, let* $\{x^1, \ldots, x^n\}$ *be an orthonormal coordinate (Definition 3.4.6), centered at a point* $p$. *Denote* $X_i = \frac{\partial}{\partial x^i}$. *Then, at* $p$, $\nabla_{X_i}X_j = 0$, $i, j = 1, \ldots, n$. *Hence,* $\nabla^2_{X_i,X_j}T = \nabla_{X_i}(\nabla_{X_j}T)$. *Sometimes the symbol* $T_{,i}$ *is used to denote* $\nabla_{X_i}T$. *At* $p$, *one uses the notation* $T_{,ji}$ *to denote the second covariant derivative where*

$$T_{,ji} \equiv \nabla^2_{X_i,X_j}T = (T_{,j})_{,i}.$$

*Notice the reversal of the order of* $i$ *and* $j$ *in the notation* $T_{,ji}$, *which means differentiation in* $x^j$ *direction first and* $x^i$ *direction second.*

*Example.* For a smooth function $f : \mathbf{M} \to \mathbf{R}$, its Hessian is the $(2,0)$ tensor field. In local system $\{x^1, \ldots, x^n\}$,

$$\nabla^2 f = f_{,ij}dx^i \otimes dx^j$$

where

$$f_{,ij} = \frac{\partial}{\partial x^i}\left(\frac{\partial f}{\partial x^j}\right) - \Gamma^k_{ij}\frac{\partial f}{\partial x^k}.$$

The second covariant derivative of a vector field is not symmetric with respect to $X$ and $Y$ in general. The amount for which the second covariant derivative fails to be symmetric is called curvature tensor.

**Definition 3.2.2** *(curvature tensor) The Riemann curvature tensor is a* $(3,1)$ *tensor field defined by*

$$\begin{aligned} R(X,Y)Z &\equiv \nabla^2_{X,Y}Z - \nabla^2_{Y,X}Z \\ &= \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z \end{aligned}$$

*for all vector fields* $X, Y$ *and* $Z$.

*Equivalently, the curvature tensor can also be written as a* $(4,0)$ *tensor via the definition*

$$R(X,Y,Z,W) = g(R(X,Y)Z,W) \equiv < R(X,Y)Z, W > \tag{3.2.1}$$

*for all smooth vector fields* $X, Y, Z, W$.

**Remark 3.2.2** *This is NOT the only way to define the* $(4,0)$ *curvature tensor. Many authors also use the definition*

$$R(X,Y,Z,W) = g(R(X,Y)W,Z). \tag{3.2.2}$$

*Note the reversal of the order of* $W$ *and* $Z$. *Thus the* $(4,0)$ *tensor defined this way has the opposite sign comparing with (3.2.1).*

**Remark 3.2.3** *One can switch the covariant and contravariant component of a tensor using the Riemann metric in the above manner. In local coordinates this is amount to lowering or raising the index in a tensor. Usually there is no need to rename the tensor since the change will be recorded either by the change of index in the local form or by the ways the arguments (vector fields or one forms) are arranged.*

**Remark 3.2.4** *In local coordinates* $\{x^1, \ldots, x^n\}$, *we can write the components of* $(3,1)$ *curvature tensor as* $R^i_{jkl}$ *which is defined by*

$$R(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})\frac{\partial}{\partial x^k} = R^l_{ijk}\frac{\partial}{\partial x^l}.$$

*Explicitly*

$$R^l_{ijk} = \partial_i \Gamma^l_{jk} - \partial_j \Gamma^l_{ik} + \Gamma^p_{jk}\Gamma^l_{ip} - \Gamma^p_{ik}\Gamma^l_{jp}. \tag{3.2.3}$$

*The components of* $(4,0)$ *curvature tensor defined by (3.2.1) can be written as* $R_{ijkl}$ *where*

$$R_{ijkl} =< R(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l} >= g_{hl}R^h_{ijk}. \tag{3.2.4}$$

*Note the superscript is lowered into the fourth subscript. In some papers, such as [Ha1], [CZ], it is lowered into the third subscript, which corresponds to formula (3.2.2). This convention produces a different sign in computation.*

We just saw that the quantity $\nabla^2_{X,Y}Z - \nabla^2_{Y,X}Z$ leads to the curvature tensor when $Z$ is a vector field. It would be curious to see what would emerge when $Z$ is replaced by a $(p,q)$ tensor $T$. Indeed there is the important

**Proposition 3.2.1** *(Ricci identities or commutation formulas) Given a smooth* $(p,q)$ *tensor field* $T$ *and smooth vector fields* $X$ *and* $Y$, *it holds*

$$\nabla^2_{X,Y}T - \nabla^2_{Y,X}T = \nabla_X(\nabla_Y T) - \nabla_Y(\nabla_X T) - \nabla_{[X,Y]}T.$$

*In local coordinate system* $\{x^1, \ldots, x^n\}$, *the above becomes*

$$\begin{aligned} &T^{l_1\ldots l_q}_{k_1\ldots k_p, ji} - T^{l_1\ldots l_q}_{k_1\ldots k_p, ij} \\ &\equiv \nabla^2_{i,j}T^{l_1\ldots l_q}_{k_1\ldots k_p} - \nabla^2_{j,i}T^{l_1\ldots l_q}_{k_1\ldots k_p} \\ &= -\Sigma^p_{h=1}R^r_{ijk_h}T^{l_1\ldots l_q}_{k_1\ldots k_{h-1}rk_{h+1}\ldots k_p} + \Sigma^q_{h=1}R^{l_h}_{ijs}T^{l_1\ldots l_{h-1}sl_{h+1}\ldots l_q}_{k_1\ldots k_p}. \end{aligned}$$

*In this formula, the notation* $T^{l_1\ldots l_q}_{k_1\ldots k_p, ij}$ *etc. means the* $\{k_1\ldots k_p; l_1, \ldots, l_q\}$ *component of the tensor* $\nabla^2_{\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^i}}T$ *etc., defined in Remark 3.2.1.*

PROOF. The first stated formula is an immediate consequence of the definition of second covariant derivatives for tensor fields (Definition 3.2.1).

To prove the second statement, one notices that $T$ can be written as the tensor product of $p$ one forms and $q$ vector fields. So only the proof when $T$ is a one form needs to be sorted out and the rest of the proof for the general formula follows from the product rule.

Let $T$ be a one form and $Z$ be a vector field. Then $T(Z)$ is a scalar function. Therefore, the Lebnitz rule says

$$X(T(Z)) = (\nabla_X T)(Z) + T(\nabla_X Z),$$

and

$$\begin{aligned} Y(X(T(Z))) = &(\nabla_Y(\nabla_X T))(Z) + (\nabla_X T)(\nabla_Y Z) + (\nabla_Y T)(\nabla_X Z) \\ &+ T(\nabla_Y(\nabla_X Z)). \end{aligned}$$

Exchanging $X$ and $Y$, the above becomes

$$\begin{aligned} X(Y(T(Z))) = &(\nabla_X(\nabla_Y T))(Z) + (\nabla_Y T)(\nabla_X Z) + (\nabla_X T)(\nabla_Y Z) \\ &+ T(\nabla_X(\nabla_Y Z)). \end{aligned}$$

Taking the difference of the last two identities, one obtains

$$\begin{aligned} &Y(X(T(Z))) - X(Y(T(Z))) \\ &= (\nabla_Y(\nabla_X T))(Z) - (\nabla_X(\nabla_Y T))(Z) + T(\nabla_Y(\nabla_X Z)) - T(\nabla_X(\nabla_Y Z)). \end{aligned}$$

Moving the first two terms of the right-hand side to the left, we arrive at

$$\begin{aligned} &(\nabla_X(\nabla_Y T))(Z) - (\nabla_Y(\nabla_X T))(Z) - [X, Y](T(Z)) \\ &\quad = T(\nabla_Y(\nabla_X Z)) - T(\nabla_X(\nabla_Y Z)). \end{aligned}$$

Since

$$[X, Y](T(Z)) = (\nabla_{[X,Y]} T)(Z) + T(\nabla_{[X,Y]} Z),$$

the above becomes

$$\begin{aligned} &(\nabla_X(\nabla_Y T))(Z) - (\nabla_Y(\nabla_X T))(Z) - (\nabla_{[X,Y]} T)(Z) \\ &\quad = T(\nabla_Y(\nabla_X Z)) - T(\nabla_X(\nabla_Y Z) + T(\nabla_{[X,Y]} Z)). \end{aligned}$$

By Definition 3.2.2, this implies

$$[\nabla^2_{X,Y} T - \nabla^2_{Y,X} T](Z) = -T(R(X, Y)Z).$$

Now let $T = T_m dx^m$, $X = \frac{\partial}{\partial x^i}$, and $Y = \frac{\partial}{\partial x^j}$ and $Z = \frac{\partial}{\partial x^k}$ in a local coordinate system. Then

$$\begin{aligned}(Z) &= -T_m dx^m (R(X,Y)Z) \\ &= -T_m dx^m (R^l_{ijk} \frac{\partial}{\partial x_l}) = -R^m_{ijk} T_m.\end{aligned}$$

That is

$$T_{k,ji} - T_{k,ij} = -R^m_{ijk} T_m.$$

Here $T_{k,ij}$ means the k-th component of the tensor $\nabla^2_{Y,X} T$. □

The curvature tensor enjoys certain symmetries and antisymmetries. The identities (ii) and (iv) below are usually referred to as the first and second Bianchi identities.

**Proposition 3.2.2** *For $X, Y, Z, T, U \in \Gamma(\mathbf{M})$, it holds*

*(i)* $R(X,Y,Z,T) = -R(Y,X,Z,T) = -R(X,Y,T,Z)$;

*(ii)* $R(X,Y,Z,T) + R(Y,Z,X,T) + R(Z,X,Y,T) = 0$;

*(iii)* $R(X,Y,Z,T) = R(Z,T,X,Y)$;

*(iv)* $(\nabla_X R)(Y,Z,T,U) + (\nabla_Y R)(Z,X,T,U) + (\nabla_Z R)(X,Y,T,U) = 0,$
(*permutation of* $X, Y, Z$).

Proof. The proof of (i) to (iii) is a simple algebra exercise from the definition. Therefore we only present *proof of iv).*

By Definition 3.1.13, $\nabla_X R$, the covariant derivative of the $(4,0)$ tensor $R$, is given by

$$\begin{aligned}&(\nabla_X R)(Y,Z,T,U) \\ &\quad = X < R(Y,Z)T,\ U > - < R(Y,Z)T,\ \nabla_X U > \\ &\quad\quad - < R(Y,Z)\nabla_X T,\ U > - < R(\nabla_X Y, Z)T,\ U > \\ &\quad\quad - < R(Y, \nabla_X Z)T,\ U >\end{aligned} \tag{3.2.5}$$

Here and later $< \cdot, \cdot > \equiv g(\cdot, \cdot)$. Since $\nabla_X g = 0$, we know that

$$X < R(Y,Z)T,\ U > = < \nabla_X (R(Y,Z)T),\ U > + < R(Y,Z)T,\ \nabla_X U > .$$

Therefore the first two terms on the right-hand side of (3.2.5) merge into one to give

$$\begin{aligned}(\nabla_X R)(Y,Z,T,U) &= < \nabla_X (R(Y,Z)T),\ U > - < R(Y,Z)\nabla_X T,\ U > \\ &\quad - < R(\nabla_X Y, Z)T,\ U > - < R(Y, \nabla_X Z)T,\ U > .\end{aligned} \tag{3.2.6}$$

Performing permutations of $X, Y, Z$ on (3.2.6) yields

$$
\begin{aligned}
(\nabla_Y R)(Z,X,T,U) =&< \nabla_Y(R(Z,X)T),\ U > - < R(Z,X)\nabla_Y T,\ U > \\
&- < R(\nabla_Y Z,X)T,\ U > - < R(Z,\nabla_Y X)T,\ U >,
\end{aligned} \tag{3.2.7}
$$

$$
\begin{aligned}
(\nabla_Z R)(X,Y,T,U) =&< \nabla_Z(R(X,Y)T),\ U > - < R(X,Y)\nabla_Z T,\ U > \\
&- < R(\nabla_Z X,Y)T,\ U > - < R(X,\nabla_Z Y)T,\ U > .
\end{aligned} \tag{3.2.8}
$$

Adding (3.2.6), (3.2.7) and (3.2.8), we deduce

$$
\begin{aligned}
&(\nabla_X R)(Y,Z,T,U) + (\nabla_Y R)(Z,X,T,U) + (\nabla_Z R)(X,Y,T,U) \\
&\quad \equiv T_1 - T_2 - T_3.
\end{aligned} \tag{3.2.9}
$$

Here $T_1$ is the sum of the three terms of the form $< \nabla_X(R(Y,Z)T),\ U >$, i.e.

$$
\begin{aligned}
T_1 =&< \nabla_X(R(Y,Z)T) + \nabla_Y(R(Z,X)T) + \nabla_Z(R(X,Y)T),\ U > \\
\equiv&< I, U > .
\end{aligned} \tag{3.2.10}
$$

The term $T_2$ is the sum of the three terms containing the covariant derivatives of $T$, i.e.

$$
\begin{aligned}
T_2 =&< (R(Y,Z)\nabla_X T) + (R(Z,X)\nabla_Y T) + (R(X,Y)\nabla_Z T),\ U > \\
\equiv&< II, U > .
\end{aligned} \tag{3.2.11}
$$

The term $T_3$ is the sum of the terms such that the covariant derivative only appears inside the arguments of $R(\cdot,\cdot)$, i.e.

$$
\begin{aligned}
T_3 =&< R(\nabla_X Y,Z)T + R(Z,\nabla_Y X)T,\ U > \\
&+ < R(\nabla_Y Z,X)T + R(X,\nabla_Z Y)T,\ U > \\
&+ < R(\nabla_Z X,Y)T + R(Y,\nabla_X Z)T,\ U > .
\end{aligned} \tag{3.2.12}
$$

Using the torsion-free property for Riemann connections,

$$
R(\nabla_X Y,Z)T + R(Z,\nabla_Y X)T = R(\nabla_X Y - \nabla_Y X, Z)T = R([X,Y],Z)T,
$$

etc., we reduce (3.2.12) to

$$
\begin{aligned}
T_3 =&< R([X,Y],Z)T + R([Y,Z],X)T + R([Z,X],Y)T, U > \\
\equiv&< III, U > .
\end{aligned} \tag{3.2.13}
$$

If we can show that $I = II + III$ then the proof is done. First let us compute $I$. From definition of curvature tensor,

$$R(Y,Z)T = \nabla_Y \nabla_Z T - \nabla_Z \nabla_Y T - \nabla_{[Y,Z]} T.$$

Hence

$$\nabla_X (R(Y,Z)T) = \nabla_X \nabla_Y \nabla_Z T - \nabla_X \nabla_Z \nabla_Y T - \nabla_X \nabla_{[Y,Z]} T.$$

Similarly

$$\nabla_Y (R(Z,X)T) = \nabla_Y \nabla_Z \nabla_X T - \nabla_Y \nabla_X \nabla_Z T - \nabla_Y \nabla_{[Z,X]} T,$$

$$\nabla_Z (R(X,Y)T) = \nabla_Z \nabla_X \nabla_Y T - \nabla_Z \nabla_Y \nabla_X T - \nabla_Z \nabla_{[X,Y]} T.$$

Adding the last three identities together and pairing the terms containing $\nabla_X T$, $\nabla_Y T$, and $\nabla_Z T$ together, we deduce

$$\begin{aligned} I &= \nabla_X (R(Y,Z)T) + \nabla_Y (R(Z,X)T) + \nabla_Z (R(X,Y)T) \\ &= (\nabla_X \nabla_Y - \nabla_Y \nabla_X) \nabla_Z T + (\nabla_Y \nabla_Z - \nabla_Z \nabla_Y) \nabla_X T \\ &\qquad + (\nabla_Z \nabla_X - \nabla_X \nabla_Z) \nabla_Y T \\ &\qquad - \nabla_X \nabla_{[Y,Z]} T - \nabla_Y \nabla_{[Z,X]} T - \nabla_Z \nabla_{[X,Y]} T \\ &= R(X,Y) \nabla_Z T + \nabla_{[X,Y]} \nabla_Z T + R(Y,Z) \nabla_X T \\ &\qquad + \nabla_{[Y,Z]} \nabla_X T + R(Z,X) \nabla_Y T + \nabla_{[Z,X]} \nabla_Y T \\ &\qquad - \nabla_X \nabla_{[Y,Z]} T - \nabla_Y \nabla_{[Z,X]} T - \nabla_Z \nabla_{[X,Y]} T. \end{aligned}$$

Rearranging the like terms, we have

$$\begin{aligned} &\nabla_X (R(Y,Z)T) + \nabla_Y (R(Z,X)T) + \nabla_Z (R(X,Y)T) \\ &= R(X,Y) \nabla_Z T + R(Y,Z) \nabla_X T + R(Z,X) \nabla_Y T \\ &\qquad + (\nabla_{[X,Y]} \nabla_Z - \nabla_Z \nabla_{[X,Y]}) T + (\nabla_{[Y,Z]} \nabla_X - \nabla_X \nabla_{[Y,Z]}) T \\ &\qquad\quad + (\nabla_{[Z,X]} \nabla_Y - \nabla_Y \nabla_{[Z,X]}) T \end{aligned} \tag{3.2.14}$$

By the definition of curvature tensor again, we know that

$$(\nabla_{[X,Y]} \nabla_Z - \nabla_Z \nabla_{[X,Y]}) T = R([X,Y],Z) T + \nabla_{[[X,Y],Z]} T.$$

Similarly

$$(\nabla_{[Y,Z]} \nabla_X - \nabla_X \nabla_{[Y,Z]}) T = R([Y,Z],X) T + \nabla_{[[Y,Z],X]} T,$$

$$(\nabla_{[Z,X]} \nabla_Y - \nabla_Y \nabla_{[Z,X]}) T = R([Z,X],Y) T + \nabla_{[[Z,X],Y]} T.$$

Note that the left-hand side of the last three terms are just the 4th, 5th and 6th term of the right-hand side of (3.2.14). Therefore

$$\begin{aligned} I &= \nabla_X(R(Y,Z)T) + \nabla_Y(R(Z,X)T) + \nabla_Z(R(X,Y)T) \\ &= R(X,Y)\nabla_Z T + R(Y,Z)\nabla_X T + R(Z,X)\nabla_Y T \\ &\quad + R([X,Y],Z)T + R([Y,Z],X)T + R([Z,X],Y)T \\ &= II + III \end{aligned} \tag{3.2.15}$$

where we just used the Jacobi identity $[[X,Y],Z] + [[Y,Z],X] + [[Z,X],Y] = 0$.

By (3.2.15), (3.2.10), (3.2.11), (3.2.13), we know that

$$T_1 = T_2 + T_3,$$

which, according to (3.2.9), proves the second Bianchi identity. □

The full curvature tensor is usually not easy to handle, especially in high dimensions. It is often convenient to deal with the trace, which is a tensor now called the Ricci curvature (tensor). The Ricci curvature can be regarded as a $(2,0)$ tensor whose trace is called the scalar curvature. The components of the curvature tensor lead to the concept of sectional curvature. Thus we have

**Definition 3.2.3** *(Ricci, Scalar and Sectional curvatures) The* $(2,0)$ *Ricci curvature tensor is the trace of the curvature tensor, i.e., for* $X, Y \in T_p(\mathbf{M})$

$$Ric(X,Y) = Trace(\cdot \to R(\cdot, X)Y) = \Sigma_{i=1}^n R(e_i, X, Y, e_i)$$

*where* $\{e_i\}$ *is an orthonormal basis for* $T_p(\mathbf{M})$.

*The scalar curvature* $R$ *is the trace of the Ricci curvature, i.e.*

$$R = \Sigma_{j=1}^n Ric(e_j, e_j).$$

*The plane* $E \subset T_p(\mathbf{M})$ *spanned by* $X, Y$ *is called a section of the tangent space. The sectional curvature with respect to* $E$ *is*

$$sec(E) = sec(X,Y) = \frac{R(X,Y,Y,X)}{g(X,X)g(Y,Y) - g(X,Y)^2}.$$

**Remark 3.2.5** *In local coordinates, the Ricci and scalar curvature are given by*

$$R_{ij} = R^k_{kij} = -g^{kl}R_{ikjl}, \qquad R = g^{ij}R_{ij} = g^{ij}g^{kl}R_{kijl} = -g^{ij}g^{kl}R_{ikjl}.$$

*By (3.2.3),*

$$R_{ij} = \partial_k \Gamma^k_{ij} - \partial_i \Gamma^k_{kj} + \Gamma^k_{kp}\Gamma^p_{ij} - \Gamma^k_{ip}\Gamma^p_{kj}. \tag{3.2.16}$$

**Remark 3.2.6** *The curvature tensor has a rich structure. For example, it can be decomposed into the sum of three parts. The first part is determined solely by the scalar curvature, the second part by the Ricci curvature. The last part, called the Weyl tensor, vanishes in dimension* 3 *and lower. See Chapter 1 of [CLN] e.g. Thus in dimension* 3*, the full curvature tensor is determined solely by the Ricci curvature. This fact makes* 3 *dimensional Ricci flow quite special.*

The following twice contracted the second Bianchi identity and is particularly useful in the study of Ricci flow. Written in local coordinates, it is

**Proposition 3.2.3**

$$2g^{ij}\nabla_i R_{jk} = \nabla_k R.$$

PROOF. In local coordinates the second Bianchi identity can be written as

$$\nabla_i R_{jklm} + \nabla_j R_{kilm} + \nabla_k R_{ijlm} = 0.$$

Multiplying this by $g^{im}g^{jl}$ and using the fact that the covariant derivative of the inverse metric is zero, we have

$$g^{im}\nabla_i(g^{jl}R_{jklm}) + g^{jl}\nabla_j(g^{im}R_{kilm}) + \nabla_k(g^{im}g^{jl}R_{ijlm}) = 0.$$

By definition of Ricci and scalar curvatures,

$$g^{im}g^{jl}R_{ijlm} = R, \qquad g^{jl}R_{jklm} = -R_{km}, \qquad g^{im}R_{kilm} = -R_{kl}.$$

Hence

$$g^{im}\nabla_i R_{km} + g^{jl}\nabla_j R_{kl} - \nabla_k R = 0.$$

The identity follows by rearranging the indices. □

## 3.3 Common differential operators on manifolds

In the next paragraph we present the concepts of gradient, divergence and Laplace operators on Riemann manifolds. These, like in the Euclidean case, are perhaps the three most widely used differential operators.

First we define the gradient of a differentiable scalar function $f : \mathbf{M} \to \mathbf{R}$. This is a confusing notion historically. At first look, it seems reasonable to define the gradient of $f$ as the covariant derivative of $f$.

The covariant derivative of $f$ is a $(1,0)$ tensor, i.e. a 1 form $df$. So, if $X$ is a vector field, then

$$\nabla_X f = X(f) = df(X).$$

Here $\nabla_X$ stands for covariant derivative and $df$ is regarded as a one form: linear functionals on the tangent spaces. However in the Euclidean setting, traditionally the gradient of a function is a vector field. Let us recall that the gradient of a smooth function in $\mathbf{R}^n$ is defined by

$$< gradient\ f, X >= X(f)$$

for all smooth vector fields $X$ on $\mathbf{R}^n$. Here the brackets $<,>$ stands for the Euclidean inner product.

Following this Euclidean tradition, one has to transplant $df$ to the tangent space via the Riemann metric.

**Definition 3.3.1** *(gradient of a scalar function) The gradient of a smooth scalar function $f : \mathbf{M} \to \mathbf{R}$, denoted by $\nabla f$, is a vector field ($(0,1)$ tensor field), such that*

$$g(\nabla f, X) = X(f) = df(X)$$

*for all smooth vector field $X$ on $\mathbf{M}$.*

Following tradition, most people still use $\nabla f$ to denote the gradient of $f$. This should not be confused with covariant derivative.

The following formula for the gradient in local coordinates is often useful. Let $(U, \phi)$ be a local chart for $\mathbf{M}$ with $\phi(m) = (x^1(m), \ldots, x^n(m))$, $m \in U$. For a smooth scalar function $f$ on $U$

$$\nabla f = g^{ij} \frac{\partial f \circ \phi^{-1}}{\partial x^j} \frac{\partial}{\partial x^i}. \tag{3.3.1}$$

The proof is straightforward from the definition. Indeed, suppose $\nabla f = c^i \frac{\partial}{\partial x^i}$. Let $X = l^i \frac{\partial}{\partial x^i}$ be any smooth vector field. By definition

$$g_{ij} c^i l^j = l^j \frac{\partial (f \circ \phi^{-1})}{\partial x^j}.$$

Since $l^j$ is arbitrary, this implies

$$g_{ij} c^i = \frac{\partial (f \circ \phi^{-1})}{\partial x^j}.$$

Multiplying by $g^{kj}$ on the above equality, and summing up, we deduce

$$c^i = g^{ij} \frac{\partial (f \circ \phi^{-1})}{\partial x^j},$$

which gives the local formula.

This local formula sometimes takes a shorter form

$$\nabla f = g^{ij} f_j \partial_i \tag{3.3.2}$$

which is convenient during heavy computations. Obviously $f_j$ stands for $\frac{\partial (f \circ \phi^{-1})}{\partial x^j}$ and $\partial_i$ for $\frac{\partial}{\partial x^i}$.

**Definition 3.3.2** *(divergence of a tensor)*

*(i) The divergence of a $C^1$ vector field $X$ on a Riemann manifold is the scalar*

$$div\ X = trace(\nabla X) \equiv \Sigma_{i=1}^n g(\nabla_{e_i} X, e_i)$$

*where $\nabla$ is the covariant derivative of $X$, and $\{e_i\}$ is an orthonormal basis of the tangent space.*

*(ii) The divergence of a $C^1$, $(p, q)$ tensor field $T$ on a Riemann manifold is a $(p, q-1)$ tensor given by*

$$\begin{aligned}(div\ T)(X_1, \ldots, X_p, \eta_1, \ldots, \eta_{q-1}) &= trace(\nabla T) \\ &= \Sigma_{i=1}^n g(\nabla_{e_i} T(X_1, \ldots, X_p, \eta_1, \ldots, \eta_{q-1}),\, e_i)\end{aligned}$$

*where $\nabla_{e_i} T(X_1, \ldots, X_p, \eta_1, \ldots, \eta_{q-1})$ is regarded as a vector field.*

**Remark 3.3.1** *The reason $\nabla_{e_i} T(X_1, \ldots, X_p, \eta_1, \ldots, \eta_{q-1})$ is a vector field is that $\nabla_{e_i} T$ is a $(p, q)$ tensor. So it is a linear combination of the direct sums of $p$ one forms and $q$ vector fields. After acting on $p$ vector fields and $(q-1)$ one forms, $\nabla_{e_i} T$ has only one component left, a vector field.*

The divergence of a vector field takes the following form in local coordinates:

For a $C^1$ vector field $X = \xi^i \frac{\partial}{\partial x^i}$,

$$divX = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} (\sqrt{g}\ \xi^i) \tag{3.3.3}$$

where $\sqrt{g} \equiv \sqrt{det(g_{ij})}$.

Here is a quick proof of this formula. By definition $\nabla X$ is a $(1,1)$ tensor given by

$$\nabla X = \left(\frac{\partial \xi^i}{\partial x^j} + \xi^k \Gamma^i_{kj}\right) \frac{\partial}{\partial x^i} \otimes dx^j.$$

The trace of a $(1,1)$ tensor $T^i_j$ is just $T^i_i$. Therefore

$$divX = \frac{\partial \xi^i}{\partial x^i} + \xi^k \Gamma^i_{ki}.$$

By the local formula for the Christoffel symbols (3.1.3)

$$\Gamma^i_{ki} = \frac{1}{2} g^{ij} \frac{\partial g_{ij}}{\partial x^k} = \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g}}{\partial x^k}. \tag{3.3.4}$$

This shows

$$divX = \frac{\partial \xi^i}{\partial x^i} + \xi^k \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g}}{\partial x^k},$$

which implies the desired formula.

Next we introduce the Laplace-Beltrami operator on scalar functions on manifolds, which is a generalization of the Laplace operator on $\mathbf{R^n}$. It perhaps is the most important second order operator in geometric analysis.

**Definition 3.3.3** *(the Laplace-Beltrami operator of a scalar function) Let $u$ be a $C^2$ scalar function on* $\mathbf{M}$. *Then the Laplace-Beltrami operator is the one defined by*

$$\Delta u = div(\nabla u).$$

By the computation of gradient and divergence, in local coordinates, it holds

$$\Delta u = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left(\sqrt{g}\, g^{ij} \frac{\partial u}{\partial x^j}\right). \tag{3.3.5}$$

The $\Delta u$ is also the trace of the Hessian of $u$ as given in Definition 3.2.1. Recall that the Hessian of $u$ is a $(2,0)$ tensor given by

$$(Hess\ u)(X,Y) = \nabla^2_{X,Y} u = \nabla_X(\nabla_Y u) - (\nabla_X Y)u.$$

Therefore in the local coordinates $(U, \phi)$ with $\phi = (x^1, \ldots, x^n)$,

$$(Hess\ u)_{ij} = (Hess\ u)(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) = \frac{\partial}{\partial x^j}\left(\frac{\partial u}{\partial x^i}\right) - \Gamma^k_{ij} \frac{\partial u}{\partial x^k}.$$

Recall the trace of a $(2,0)$ tensor $T_{ij}$ is $g^{il}T_{li}$. So

$$tr(Hess\ u) = g^{il}\left[\frac{\partial}{\partial x^i}\left(\frac{\partial u}{\partial x^l}\right) - \Gamma^k_{li}\frac{\partial u}{\partial x^k}\right].$$

To see this being equal to $\Delta u$, we recall that

$$\frac{\partial g_{ij}}{\partial x^k} = g_{il}\Gamma^l_{jk} + g_{jl}\Gamma^l_{ik}.$$

This is nothing but the local form of formula (iii) in the fundamental theorem of Riemann geometry, which means that the metric tensor $g$ has zero covariant derivative. Since $(g^{ij})$ is the inverse of $(g_{ij})$, it holds

$$g^{ih}g_{hj} = \delta^i_j$$

Differentiating this identity with respect to $x^k$, one can show by simple manipulation of indices that

$$\frac{\partial g^{ij}}{\partial x^k} = -g^{il}\Gamma^j_{lk} - g^{jl}\Gamma^i_{lk}. \tag{3.3.6}$$

From (3.3.5) and (3.3.4),

$$\Delta u = g^{ij}\frac{\partial^2 u}{\partial x^i \partial x^j} + \frac{\partial g^{ij}}{\partial x^i}\frac{\partial u}{\partial x^j} + g^{ij}\Gamma^l_{il}\frac{\partial u}{\partial x^j}.$$

Taking $k = i$ in (3.3.6) and summing up, and then substituting the sum into the above expression for $\Delta u$, we deduce

$$\Delta u = g^{ij}\frac{\partial^2 u}{\partial x^i \partial x^j} - g^{il}\Gamma^k_{li}\frac{\partial u}{\partial x^k} = tr(Hess\ u). \tag{3.3.7}$$

## 3.4 Geodesics, exponential maps, injectivity radius, Jacobi fields, index forms

In this section we will define the concept of geodesics. There are two approaches to this. One is to regard a geodesic as generalization of straight lines in the Euclidean space in the sense that the tangent vector of a straight line is a constant, i.e. the derivative of the tangent vector field along a straight line is zero. The other approach is to view geodesics as locally distance minimizing curves.

**Definition 3.4.1** *(covariant derivative for a vector field along a curve) Let $c = c(t)$ be a $C^1$ curve on $\mathbf{M}$ and $X = X(c(t))$ be a $C^1$ vector field along $c$. Fixing $t$, let $V_1, \ldots, V_n$ be $C^1$ vector fields in an open neighborhood $U$ of $c(t)$, such that $V_1(p), \ldots, V_n(p)$ form a basis for the tangent space $T_p(\mathbf{M})$ for each $p \in U$. Suppose*

$$X(c(t)) = \xi^i(t) V_i(c(t)).$$

*Then*

$$\nabla_{c'(t)} X(c(t)) \equiv \dot{\xi}^i(t) V_i(c(t)) + \xi^i(t) \nabla_{c'(t)} V_i|_{c(t)}.$$

The definition is clearly independent of the choice of the basis $V_i$. If the tangent vector $c'(t)$ for the curve $c = c(t)$ can be expanded to a vector field $Y$, defined in an open neighborhood of $c = c(t)$, then the above definition for $\nabla_{c'(t)} X(c(t))$ coincides with $\nabla_Y X$ restricted to $c = c(t)$. However, this expansion may not always be possible.

**Definition 3.4.2** *(geodesics) Let $(\mathbf{M}, g)$ be a Riemann manifold and $\nabla$ be the Riemann connection. A parameterized curve $c = c(t)$ on $\mathbf{M}$ is a geodesic if $\nabla_{c'} c' = 0$.*

Note that the tangent vector has constant magnitude, i.e. $|c'(t)| = \sqrt{g(c'(t), c'(t))}$ is a constant. Indeed, for $f = f(c(t)) \equiv g(c'(t), c'(t))$

$$\begin{aligned} \frac{d}{dt} f(c(t)) &= \nabla_{c'(t)} [g(c'(t)), c'(t))] \\ &= 2g(\nabla_{c'(t)} c'(t), c'(t))) = 0. \end{aligned}$$

Here we have used the Leibniz rule and the fact that $\nabla g = 0$.

Next we present the well-known geodesic equation in local coordinates. Let $(U, \phi)$ be a local chart of $\mathbf{M}$ such that $\phi = (\phi_1, \ldots, \phi_n)$ being a local diffeomorphism from $U \subset \mathbf{M}$ to a domain in $\mathbf{R^n}$. Let $c = c(t)$ be a curve in $U$, then

$$c(t) = \phi^{-1}(x^1(t), \ldots, x^n(t))$$

where $(x^1(t), \ldots, x^n(t))$ is the parametric equation for a curve in $\phi(U) \subset \mathbf{R^n}$. This image of $c = c(t)$ under $\phi$ is often regarded the same as $c = c(t)$ itself. Note that $x^i(t) = \phi_i(c(t))$. The geodesic equation in local coordinates is actually a system of equations for $x^i$.

Let $f : U \to \mathbf{R}$ be a smooth function. By Definition 3.1.3

$$c'(t)(f) = \frac{d}{dt}(f \circ \phi^{-1})(x^1(t), \ldots, x^n(t)) \qquad \text{Euclidean derivative}$$

$$= \frac{\partial(f \circ \phi^{-1})}{\partial x^i} \dot{x}^i(t) \qquad \text{Euclidean chain rule}$$

$$= \frac{\partial f}{\partial x^i} \dot{x}^i(t). \qquad \text{Definition 3.1.4}$$

Therefore

$$c'(t) = \dot{x}^i(t) \frac{\partial}{\partial x^i}.$$

Now we compute, from the definition of covariant derivative

$$\nabla_{c'(t)} c'(t) = \nabla_{c'(t)} (\dot{x}^i(t) \frac{\partial}{\partial x^i})$$

$$= \frac{d^2 x^i(t)}{dt^2} \frac{\partial}{\partial x^i} + \dot{x}^i(t) \nabla_{c'(t)} \frac{\partial}{\partial x^i}.$$

If $c = c(t)$ is a geodesic, then $\nabla_{c'(t)} c'(t) = 0$. Hence $x^i$ satisfy the system of equations

$$\frac{d^2 x^i}{dt^2} + \Gamma^i_{jk}(c(t)) \frac{dx^j}{dt} \frac{dx^k}{dt} = 0. \tag{3.4.1}$$

By standard theory of ordinary differential equations, this system always has a local in-time solution. It is also easy to see that a geodesic is a critical point in for the first variation of the arclength. We will return to this point in Proposition 3.4.1 below.

In multivariable calculus suitable coordinates such as rectangular and spherical ones play important roles since they simplify complicated calculations and expressions. For differential geometry, due to the complexity of objects one is facing, finding suitable local coordinates is even more important.

First, let us introduce natural local coordinates called the exponential map, using the geodesics. The geodesic equation given above is a system of second order, quadratic nonlinear ordinary differential equations. Since the manifold is smooth, the coefficients of the equation $\Gamma^k_{ij}$ are also smooth. By standard theory of ordinary differential equations, for any point $p \in \mathbf{M}$ and a vector $v \in T_p\mathbf{M}$, there exists a unique geodesic $c = c(t)$, $t \in [0, t_0]$ for some $t_0 > 0$, such that $c(0) = p$ and

$c'(0) = v$. Here $t_0$ may depend on $p$ and $v$. If $t_0 < \infty$, then we consider the curve

$$l(t) = c(t_0 t).$$

Clearly $l(0) = c(0) = p$ and $l'(0) = t_0 c'(0) = t_0 v$. Also the curve $l$ is well defined for $t \in [0, 1]$. Moreover

$$\nabla_{l'(t)} l'(t) = t_0^2 \nabla_{c'(s)} c'(s) = 0, \qquad s = t_0 t.$$

Therefore $l$ is a geodesic, existing at least for the time interval $[0, 1]$ such that $l(0) = p$ and $l'(0) = t_0 v$. Note that $v$ is an arbitrary vector in $T_p\mathbf{M}$. From the theory of ordinary differential equations, it is not hard to prove that there exists an maximal open set $D \subset T_p\mathbf{M}$, satisfying the following properties.

(a) $0 \in D$; (b) for each $v \in D$, there exists a unique geodesic $l$, such that $l(0) = p$ and $l'(0) = v$; (c) exists at least on the time interval $[0, 1]$.

Now we are ready to give

**Definition 3.4.3** *(exponential map) Give $p \in \mathbf{M}$, the exponential map $exp_p$ is a map from $D \subset T_p\mathbf{M}$ to $\mathbf{M}$ given by*

$$exp_p(v) = l(1)$$

*where $l$ is the geodesic such that $l(0) = p$ and $l'(0) = v$.*

**Remark 3.4.1** *$exp_p$ is a local diffeomorphism in a neighborhood of the zero tangent vector.*

*By the inverse function theorem, it suffices to prove that, $Dexp_p$, the derivative of $exp_p$ at $0 \in T_p\mathbf{M}$ is nonsingular. We will prove that $Dexp_p$ at $0$ is actually the identity map.*

*To prove this fact, let us recall a definition of derivative of a smooth map, say $F$, from a smooth manifold $M_1$ to another smooth manifold $M_2$. The derivative of $F$ at $m_1$ is the linear map from $T_{m_1}M_1$ to $T_{F(m_1)}M_2$ such that*

$$b'(0) = DF_{m_1} a'(0).$$

*Here $a = a(\tau)$ is any smooth curve on $M_1$ such that $a(0) = m_1$ and $b = b(\tau)$ is the image of $a$ under $F$, i.e. $b(\tau) = F(a(\tau))$. Note that this definition is nothing but a restatement of the chain rule.*

*Now we take $M_1 = T_p\mathbf{M}$, $M_2 = \mathbf{M}$ and $F = exp_p$ with domain $D \subset T_p\mathbf{M}$. Pick any $v \in D$, let $a = \tau v$. Then $b = b(\tau) = exp_p(\tau v)$. By the above definition*

$$\frac{d}{d\tau} exp_p(\tau v)\big|_{\tau=0} = [Dexp_p]\big|_0 v.$$

*From its definition, for* $\tau \in [0, 1]$*,*

$$exp_p(\tau v) = c_\tau(1)$$

*where* $c = c_\tau(t)$ *is the geodesic such that* $c_\tau(0) = p$ *and* $c'_\tau(0) = \tau v$*. Here the derivative is with respect to* $t$*. By uniqueness of local solutions of the geodesic equation, we know that*

$$c_\tau(t) = l(\tau t)$$

*where* $l = l(t)$ *is the geodesic such that* $l(0) = p$ *and* $l'(0) = v$*. Therefore*

$$exp_p(\tau v) = c_\tau(1) = l(\tau). \tag{3.4.2}$$

*i.e.* $\exp_p(\tau v)$ *is just the geodesic starting from* $p$*, with initial tangent vector equaling* $v$*. Hence*

$$\frac{d}{d\tau} exp_p(\tau v)\big|_{\tau=0} = v$$

*which shows* $v = [Dexp_p]\big|_0 v$*. Since* $v$ *is an arbitrary vector in the domain* $D$ *containing* $0$*, we know that* $[Dexp_p]|_0$ *is the identity map.*

The exponential map $exp_p$ is a natural local diffeomorphism between the tangent space $T_p\mathbf{M}$ and the manifold $\mathbf{M}$. Of course we can not expect $exp_p$ to be a global diffeomorphism in general. The next concept, called injectivity radius, measures to what extent $exp_p$ stays as a diffeomorphism.

**Definition 3.4.4** *(injectivity radius) The injectivity radius at* $p \in \mathbf{M}$*, denoted by* $inj(p)$*, is the radius of the largest open ball in* $T_p\mathbf{M}$*, centered at* $0$ *(tangent vector), on which* $exp_p$ *is a diffeomorphism.*

*The infimum of* $inj(p)$ *for all* $p \in \mathbf{M}$*, which is denoted by* $Inj(\mathbf{M})$*, is called the injectivity radius of the whole manifold* $\mathbf{M}$*.*

A closely related concept is called the conjugate point.

**Definition 3.4.5** *(conjugate point) Let* $p \in \mathbf{M}$ *and* $exp_p : T_p\mathbf{M} \to \mathbf{M}$ *be the exponential map. A point* $x \in \mathbf{M}$ *is a conjugate point of* $p$ *if* $x$ *is a singular value of* $exp_p$*, i.e. there exists* $v \in T_p\mathbf{M}$ *such that* $x = exp_p(v)$ *and the linear map* $Dexp_p : T_v(T_p\mathbf{M}) \to T_x\mathbf{M}$ *is singular.*

Knowledge of injectivity radius and conjugate points is important to the understanding of the structure of a manifold and to the study of Ricci flow. We will present a few results on the lower bound of injectivity radius in Section 3.6.

In differential geometry, many objects such as connections, curvatures, etc. involve rather complicated expressions. Therefore, it is imperative to use efficient local coordinate systems to simplify the notations and computation. One of the most useful coordinate systems is given by

**Definition 3.4.6** *(Local orthonormal coordinates) Let $p \in \mathbf{M}$ and $exp_p$ be the exponential map defined on $D \subset T_p\mathbf{M}$. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis for $T_p\mathbf{M}$, i.e. $g|_p(e_i, e_j) = \delta_{ij}$. Given $v = y^i e_i \in T_p\mathbf{M}$, define*

$$J(v) = (y^1, \ldots, y^n).$$

*Then the map $\phi \equiv J \circ exp_p^{-1}$ is a local diffeomorphism which maps $U \equiv exp_p(D)$ to $\mathbf{R^n}$.*

*The local chart $(U, \phi)$ is called local orthonormal coordinates around $p$.*

**Remark 3.4.2** *Under local orthonormal coordinates, there hold, at the point $p$ (where $y = (y^1, \ldots, y^n) = 0$)*

*1). $g_{ij} = g(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}) = \delta_{ij}$*

*2). $\Gamma^i_{jk} = 0$.*

*The proof, though very short, needs certain care about the concepts. Let $f$ be a smooth function on $\mathbf{M}$. By definition*

$$\begin{aligned}\frac{\partial f}{\partial y^i} = \frac{\partial [f \circ \phi^{-1}]}{\partial y^i} &= \frac{\partial [f \circ exp_p \circ J^{-1}](y^1, \ldots, y^n)}{\partial y^i} \\ &= \frac{\partial [f \circ exp_p(y^1 e_1 + \cdots + y^n e_n)]}{\partial y^i}.\end{aligned}$$

*So, at $p$ ($y = 0$), we have*

$$\frac{\partial f}{\partial y^i} = \frac{d[f \circ exp_p(s e_i)]}{ds}\bigg|_{s=0} = \frac{d\ exp_p(s e_i)]}{ds}\bigg|_{s=0} f = e_i f.$$

*Here we just used the definition of the tangent of a curve, i.e. $c'(s) f = [f \circ c(s)]'$. Since $f$ is arbitrary, we know $\frac{\partial}{\partial y^i} = e_i$, which is the chosen orthonormal basis for $T_p\mathbf{M}$. Hence we have proven that*

$$g_{ij} = g(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}) = \delta_{ij}.$$

*To prove the second statement, let us pick $v = y^i e_i \in T_p\mathbf{M}$. Then $c = c(t) = exp_p(tv)$ is a local geodesic such that $c(0) = p$ and $c'(0) = v$. The parametric equation for $c = c(t)$ in the local orthonormal system is*

$$\phi(c) = \phi(c(t)) = J \circ exp_p^{-1}(c(t)) = J \circ exp_p^{-1}(exp_p(tv)) = t(y^1, \ldots, y^n).$$

*Therefore $x^i \equiv ty^i$ satisfies the geodesic equation (3.4.1) which implies*

$$\Gamma^i_{jk}(exp(tv))y^j y^k = 0.$$

*Note that this equation does not mean $\Gamma^i_{jk}(exp(tv)) = 0$ if $t \neq 0$. This is because for a different choice of $\{y^i\}$, the point $exp(tv)$ may be different. However, the same equation implies $\Gamma^i_{jk}(p) = 0$ by taking $t = 0$ and arbitrary $y^i$. This proves the second claim.*

After all these equations and computations about geodesics, let us take a more intuitive view of geodesics, i.e. as local distance minimizing curves. We recall the concept of length of a curve.

**Definition 3.4.7** *(length (arclength) of a curve) Let $c = c(s)$, $s \in [a, b] \subset \mathbf{R}$ be a piecewise $C^1$ curve on a Riemann manifold $\mathbf{M}$. Then the length of $c$ is*

$$|c| = \int_a^b \sqrt{g(c'(s), c'(s))}ds.$$

A closely related concept is the energy of a curve.

**Definition 3.4.8** *(energy of a curve) Let $c = c(s)$, $s \in [a, b] \subset \mathbf{R}$ be a piecewise $C^1$ curve on a Riemann manifold $\mathbf{M}$. Then the energy of $c$ is*

$$e(c) = \frac{1}{2}\int_a^b g(c'(s), c'(s))ds.$$

**Definition 3.4.9** *(variation of a curve) A variation of a smooth curve $c : [a, b] \to \mathbf{M}$ is a smooth, two variable function $B$ that maps $[a, b] \times (-\epsilon, \epsilon)$ to $\mathbf{M}$, such that $B(s, 0) = c(s)$. Here $\epsilon > 0$. The tangent of the curve $B(\cdot, t)$ at $(s, t)$ is denoted by $\partial_s B$, and that of the curve $B(s, \cdot)$ at $(s, t)$ is denoted by $\partial_t B$. For any fixed $t$, the curve $s \to B(s, t)$ is denoted by $c_t$.*

**Proposition 3.4.1** *Let $B(s, t) = c_t(s)$ be a variation of the curve $c = c(s)$. Then, at $t = 0$,*

$$\frac{d}{dt}e(c_t) = g(\partial_t B, c'(s))\big|_{s=a}^{s=b} - \int_a^b g(\partial_t B, \nabla_{c_t'(s)} c_t'(s))ds,$$

$$\frac{d}{dt}l(c_t) = \frac{b-a}{l}\left( g(\partial_t B, c'(s))\Big|_{s=a}^{s=b} - \int_a^b g(\partial_t B, \nabla_{c_t'(s)} c_t'(s))ds \right).$$

*Here $l$ is the length of the curve $c = c(s)$ on $[a, b]$.*

Proof. We just give a proof of the first formula and leave the second one as exercise. We start from

$$\frac{d}{dt}e(c_t) = \frac{1}{2}\frac{d}{dt}\int_a^b g(c_t'(s), c_t'(s))ds.$$

In the following computation, we will use properties of covariant derivatives of vector fields on a Riemann manifold. But, $c_t'(s)$ and $\partial_t B$ are not vector fields on $\mathbf{M}$. They are defined only in a lower dimensional subset in general. However using similar idea as in Definition 3.4.1, one can still apply the rules in the fundamental theorem of Riemann geometry. Hence

$$\begin{aligned}\frac{d}{dt}e(c_t) &= \frac{1}{2}\int_a^b \partial_t[g(c_t'(s), c_t'(s))]ds = \int_a^b g(\nabla_{\partial_t B} c_t'(s), c_t'(s))ds \\ &= \int_a^b g(\nabla_{\partial_t B}\partial_s B, \partial_s B)ds = \int_a^b g(\nabla_{\partial_s B}\partial_t B, \partial_s B)ds \\ &= \int_a^b \partial_s[g(\partial_t B, \partial_s B)]ds - \int_a^b g(\partial_t B, \nabla_{\partial_s B}\partial_s B)ds.\end{aligned}$$

Here we have used the identity (torsion free condition in the fundamental theorem of Riemann geometry)

$$\nabla_{\partial_s B}\partial_t B - \nabla_{\partial_t B}\partial_s B = [\partial_s B, \partial_t B] = 0. \tag{3.4.3}$$

The proposition is thus proven since $\partial_s B = c_t'(s)$. □

From the first variation formula of the arclength, we know that a distance minimizing curve joining two points $p, q \in \mathbf{M}$ must be a geodesic. If $p, q$ lie in sufficiently small neighborhood, a geodesic connecting the two and lying in the neighborhood is also distance minimizing. However, this may not be the case if $p, q$ are not close. It is interesting to study the borderline case, that is the furthest point beyond which a geodesic is no longer distance minimizing. These points form the so-called cut-locus described by

**Definition 3.4.10** *Let $\mathbf{M}$ a complete manifold (c.f. Definition 3.4.14), $p \in \mathbf{M}$ and $v \in T_p\mathbf{M}$ with $\|v\| = 1$. Write $c_v(t) = exp_p(tv)$, the geodesic such that $c_v(0) = p, c_v'(0) = v$. Define*

$$l_v = \sup\{t > 0 \mid c_v \quad \text{is distance minimizing on} \quad [0, t]\}$$

*The cut-locus of p is the set* $\{exp(l_v v)\,|\,v \in T_p\mathbf{M}, \|v\| = 1, l_v < +\infty\}$.

*Every point in the cut-locus is called a cut point.*

Next we consider the second variation formula.

**Proposition 3.4.2** *Given a length parameterized geodesic* $c = c(s)$ *with length* $l$, *let* $B(s,t,u) = c_{t,u}(s)$ *be a smooth variation with two parameters, i.e.* $B$ *is a smooth map from* $[0,l]\times(-\epsilon,\epsilon)\times(-\delta,\delta)$ *to* $\mathbf{M}$, *and* $B(s,0,0) = c(s)$. *Then, at* $t = u = 0$,

$$\frac{\partial^2}{\partial u \partial t} e(c_{t,u}) = g(\nabla_{\partial_u B}\partial_t B, c'(s))\Big|_{s=0}^{s=l} + \int_0^l \left[g(\nabla_{c'(s)}\partial_t B, \nabla_{c'(s)}\partial_u B) + R(\partial_u B, c'(s), \partial_t B, c'(s))\right] ds$$

*and*

$$\frac{\partial^2}{\partial u \partial t} l(c_{t,u}) = g(\nabla_{\partial_u B}\partial_t B, c'(s))\Big|_{s=0}^{s=l} + \int_0^l \left[g(\nabla_{c'(s)}\partial_t B, \nabla_{c'(s)}\partial_u B) + R(\partial_u B, c'(s), \partial_t B, c'(s))\right] ds - \int_0^l g(c'(s), \nabla_{c'(s)}\partial_t B) g(c'(s), \nabla_{c'(s)}\partial_u B) ds.$$

Proof. As in the proof of the first variation formula

$$\frac{1}{2}\frac{\partial}{\partial t} g(\partial_s B, \partial_s B) = g(\nabla_{\partial_s B}\partial_t B, \partial_s B).$$

Therefore

$$\frac{1}{2}\frac{\partial^2}{\partial u \partial t} g(\partial_s B, \partial_s B) = g(\nabla_{\partial_u B}\nabla_{\partial_s B}\partial_t B, \partial_s B) + g(\nabla_{\partial_s B}\partial_t B, \nabla_{\partial_u B}\partial_s B).$$

By (3.4.3) and the Definition 3.2.2 (curvature tensor), the above becomes

$$\frac{1}{2}\frac{\partial^2}{\partial u \partial t} g(\partial_s B, \partial_s B) = g(\nabla_{\partial_s B}\nabla_{\partial_u B}\partial_t B, \partial_s B) + R(\partial_u B, \partial_s B, \partial_t B, \partial_s B) + g(\nabla_{\partial_s B}\partial_t B, \nabla_{\partial_s B}\partial_u B).$$

If $t = u = 0$, then $\partial_s B = c'(s)$ and $\nabla_{\partial_s B}\partial_s B = 0$ since $c = c(s)$ is a geodesic. Hence

$$\frac{\partial}{\partial s} g(\nabla_{\partial_u B}\partial_t B, \partial_s B) = g(\nabla_{\partial_s B}\nabla_{\partial_u B}\partial_t B, \partial_s B) + g(\nabla_{\partial_u B}\partial_t B, \nabla_{\partial_s B}\partial_s B) = g(\nabla_{\partial_s B}\nabla_{\partial_u B}\partial_t B, \partial_s B).$$

Therefore

$$\frac{1}{2}\frac{\partial^2}{\partial u \partial t} g(\partial_s B, \partial_s B) = \frac{\partial}{\partial s} g(\nabla_{\partial_u B}\partial_t B, \partial_s B) + R(\partial_u B, \partial_s B, \partial_t B, \partial_s B) + g(\nabla_{\partial_s B}\partial_t B, \nabla_{\partial_s B}\partial_u B).$$

The first formula in the proposition follows by integration.

The second formula is proven similarly. □

**Exercise 3.4.1** *Prove the second formula in the proposition.*

In the second variation formula for the energy and curve length, the principal term is

$$I \equiv \int_0^l \left[ g(\nabla_{c'(s)}\partial_t B, \nabla_{c'(s)}\partial_u B) + R(\partial_u B, c'(s), \partial_t B, c'(s)) \right] ds$$

If we write $X = \partial_t B$ and $Y = \partial_u B$, then the above becomes

$$I = \int_0^l \left[ g(\nabla_{c'(s)}X, \nabla_{c'(s)}Y) + R(Y, c'(s), X, c'(s)) \right] ds$$

If we regard the energy or arclength as functionals of curves, this quantity can be regarded as Hessians of the functional, which play a similar role as the Hessian of a function.

**Definition 3.4.11** *(index form) Let $c = c(s)$ be a length parameterized geodesic with length $L$. Then the bilinear form $I$ defined by*

$$I(X,Y) = \int_0^L [g(X', Y') + g(X, R(c', Y)c')]ds = \int_0^L [g(X', Y') + R(X, c', Y, c')]ds$$

*is called the index form of $c$. Here $X$ and $Y$ are vector fields along $c = c(s)$, and $X', Y'$ are the covariant derivative of $X, Y$ along $c'(s)$, respectively.*

Observe that

$$g(\nabla_{c'(s)}X, \nabla_{c'(s)}Y) = \frac{\partial}{\partial s} g(X, \nabla_{c'(s)}Y) - g(X, \nabla_{c'(s)}\nabla_{c'(s)}Y).$$

After integration, we obtain

$$I(X,Y) = g(X, \nabla_{c'(s)}Y)|_0^L - \int_0^L g(X, \nabla_{c'(s)}\nabla_{c'(s)}Y + R(Y, c'(s))c'(s))ds. \tag{3.4.4}$$

This formula provides the motivation for

**Definition 3.4.12** *(Jacobi field) A Jacobi field along a geodesic $c = c(t)$ is a vector field $J$ along $c$, satisfying the second order equation*

$$J''(t) + R(J(t), c'(t))c'(t) = 0$$

*where $J''(t) \equiv \nabla_{c'(t)} \nabla_{c'(t)} J(t)$ and $R$ is the curvature tensor.*

**Remark 3.4.3** *Note by (3.4.4), the index form is decided by the information on the end points if one of the vector fields is a Jacobi field, i.e.*

$$I(X, Y) = g(X, \nabla_{c'(s)} Y)|_0^L \tag{3.4.5}$$

*if $Y$ is a Jacobi field.*

*This property allows one to simplify many expressions such as volume form, Laplace operator when using certain Jacobi fields. The basic volume comparison theorems are derived in this manner. We will return to this point with detail a little later.*

Another important property of Jacobi fields is that they describe the derivative of the exponential map. More precisely, we have

**Proposition 3.4.3** *Let $u, v \in T_p\mathbf{M}$, $p \in \mathbf{M}$. Let $c$ be the geodesic $exp_p(sv)$. Then*

$$Dexp_p\big|_{sv}(su) = Y(s)$$

*where $Y$ is the Jacobi field along $c = c(s)$ such that $Y(0) = 0$ and $Y'(0) = u$.*

Proof. Recall that $exp_p$ maps a ball in $T_p\mathbf{M}$ to $\mathbf{M}$. Hence $Dexp_p\big|_{sv}$ is a map from $T_{sv}(T_p\mathbf{M})$ to $T_{exp_p(sv)}\mathbf{M}$. Since $T_p\mathbf{M}$ is linear, we know that $T_{sv}(T_p\mathbf{M})$ and $T_p\mathbf{M}$ are isomorphic. Therefore there is no need to distinguish vectors in these two spaces.

For small numbers $s$ and $t$ consider the variation

$$B(s, t) = exp_p(sv + stu).$$

Then the chain rule tells us

$$\frac{\partial B}{\partial t} = Dexp_p\big|_{sv+stu}(su).$$

Define the vector field $Y = Y(s) = \frac{\partial B}{\partial t}(s, t)$ at $t = 0$. By Definition 3.4.1,

$$Y'(s) = \nabla_{c'(s)}(Dexp_p\big|_{sv}(su)) = Dexp_p|_{sv}u + s\nabla_{c'(s)}(Dexp_p|_{sv}u).$$

Since $Y$ clearly satisfies the initial conditions $Y(0) = 0$ and $Y'(0) = u$, we just need to show that $Y$ is a Jacobi field. This is true since

$$\begin{aligned}\nabla_{c'(s)}\nabla_{c'(s)}Y(s) &= \nabla_{c'(s)}\nabla_{Y(s)}c'(s)\\ &= \nabla_{Y(s)}\nabla_{c'(s)}c'(s) + R(c'(s), Y(s))c'(s)\\ &= -R(Y(s), c'(s))c'(s).\end{aligned}$$

Here we just used the fact that $\nabla_{c'(s)}c'(s) = 0$ since $c$ is a geodesic. □

The next two propositions are two immediate applications of Jacobi fields.

**Proposition 3.4.4** *Let $p, q$ be two points in a complete manifold $\mathbf{M}$ (c.f. Definition 3.4.14), connected by a geodesic $c$ such that $c(a) = p$ and $c(b) = q$ with $b > a$. Then $q$ is a conjugate point of $p$ if and only if there exists a nontrivial Jacobi field along $c$ such that $J(a) = J(b) = 0$.*

Proof. We suppose $q = exp_p v$ for some $v \in T_p\mathbf{M}$. If $q$ is conjugate to $p$, then by Definition 3.4.5 there exists $\xi \in T_v(T_p\mathbf{M})$ such that $Dexp_p|_v(\xi) = 0$. Define a family of curves on $\mathbf{M}$

$$\beta(s, t) = exp_p(s(v + t\xi)).$$

Then, as in the proof of the last proposition, we know that the vector field

$$Y(s) = \frac{\partial \beta}{\partial t}(s, 0)$$

is a Jacobi field along the curve $c = c(s) = exp_p(sv)$. By the chain rule

$$Y(s) = \frac{\partial \beta}{\partial t}(s, 0) = Dexp_p|_{sv}(s\xi).$$

Clearly $Y(0) = 0$ and $Y(1) = Dexp_p|_v\xi = 0$ by assumption. But $Y'(0) = Dexp_p|_0\xi = \xi$. So $Y$ is nontrivial. This proves one direction of the proposition.

On the other hand suppose there exists a nontrivial Jacobi field along $c = exp_p(sv)$ such that $J(0) = 0$ and $J(1) = 0$. Then $J'(0) \neq 0$. Define $\beta(s, t) = exp_p(s(v+tJ'(0)))$. By the proof of the last proposition again, we know that $J(s) = \frac{\partial \beta}{\partial t}(s, t)|_{t=0}$ and

$$Dexp_p|_v J'(0) = J(1) = 0$$

which shows that $q = c(1)$ is a conjugate point of $p$. □

Another application of Jacobi fields is that they can characterize the local orthonormal and the associated geodesic spherical coordinates which simplify many computations.

Let $B(p,r) \subset \mathbf{M}$ be the geodesic ball centered at $p$ with radius $r$, i.e.

$$B(p,r) = \{exp_p(sv) \,|\, s \in [0,r], v \in T_p\mathbf{M}, \|v\| = 1\}.$$

Suppose the ball does not intersect the cut-locus of $p$. By Proposition 3.6.2 part (3) below, the inverse exponential map $exp_p^{-1}$ exists and is a local chart on $B(p,r)$. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of $T_p\mathbf{M}$. Every point $q$ in $B(p,r)$ is represented in this chart by the coordinates $(x^1, \ldots, x^n)$, i.e. $q = exp_p(x^1 e_1 + \cdots + x^n e_n)$. Recall from Definition 3.4.6 that $(x^1, \ldots, x^n)$ is a local orthonormal coordinate of $q$.

**Definition 3.4.13** *(geodesic spherical coordinates) Let $\{x^1, \ldots, x^n\}$ be a local orthonormal coordinate of $q$. Then $(v,s) \in S^{n-1} \times [0,\infty)$, the coordinates of $(x^1, \ldots, x^n)$ in the spherical system in $\mathbf{R^n}$ is called a geodesic spherical coordinate of $q$.*

**Proposition 3.4.5** *Let $q = exp_p(sv) \in B(p,r)$ which does not intersect with the cut-locus of $p$. Here $v$ is a unit vector in $T_p\mathbf{M}$ and $s > 0$. Let $(x^1, \ldots, x^n)$ be the above local orthonormal coordinate of $q$.*

*Then*

$$\frac{\partial}{\partial x^i}|_q = \frac{1}{s} Y_i(s),$$

*where $Y_i$ is the Jacobi field along the curve $c = c(t) = exp_p(tv)$ such that $Y_i(0) = 0$ and $Y_i'(0) = e_i$, $i = 1, \ldots, n$.*

Proof. For any fixed $s \in [0,r]$, let $q = exp_p(sv)$. By Proposition 3.6.2 part (3) again, the ball $B(p,r)$ does not contain any conjugate point of $p$. Hence $Dexp_p|_{sv}$ is nonsingular and thus

$$\{Dexp_p|_{sv}e_1, \ldots, Dexp_p|_{sv}e_{n-1}, Dexp_p|_{sv}e_n\}$$

is a basis of $T_q\mathbf{M}$. In fact it is nothing but the local basis $\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\}$. One just needs to verify that, for any smooth function $f$ on $B(p,r)$, by definition,

$$\frac{\partial f}{\partial x^i}|_q = \left[\frac{d}{dl}|_{l=0} exp_p(sv + le_i)\right] f = [Dexp_p|_{sv}e_i]\; f.$$

Let $c = c(t)$ be the geodesic $exp_p(tv)$. According to Proposition 3.4.3,

$$Dexp_p\big|_{sv}(se_i) = Y_i(s)$$

where $Y_i$ is the Jacobi field along $c = c(t)$ such that $Y_i(0) = 0$ and $Y_i'(0) = e_i$, $i = 1, \ldots, n$. Hence

$$\{\frac{\partial}{\partial x_1}|_q, \ldots, \frac{\partial}{\partial x_n}|_q\} = \{\frac{1}{s}Y_1(s), \ldots, \frac{1}{s}Y_n(s)\}$$

is the canonical basis of $T_q\mathbf{M}$. □

Having discussed a few local properties of the exponential map, we turn to the concept of complete manifolds, which is based on a global property of the exponential map.

**Definition 3.4.14** *A Riemann manifold is geodesically complete if any geodesic can be extended to a geodesic defined on all real lines.*

The following theorem, called Hopf-Rinow theorem, provides a useful description of geodesically complete manifold.

**Theorem 3.4.1** *(Hopf-Rinow theorem) The following statements for a Riemann manifold* $(\mathbf{M}, g)$ *are equivalent.*

*(i) Let* $d$ *be the distance function associated with* $g$*, i.e. for* $p, q \in \mathbf{M}$,

$$d(p, q) = \inf\{\textit{ length of smooth curves joining } p, q\}.$$

*Then* $\mathbf{M}$ *is a complete metric space with respect to* $d$*.*

*(ii) For one point* $p \in \mathbf{M}$*, the exponential map* $exp_p$ *is defined on whole* $T_p\mathbf{M}$*.*

*(iii) For any point* $q \in \mathbf{M}$*, the exponential map* $exp_q$ *is defined on whole* $T_q\mathbf{M}$*.*

*(iv)* $\mathbf{M}$ *is geodesically complete.*

*Moreover any one of the above statements implies*

*(v) Any two points in* $\mathbf{M}$ *can be joined by a minimal geodesic, i.e. a geodesic whose length is the distance between the two points.*

**Remark 3.4.4** *By virtue of this theorem, one usually calls a geodesically complete manifold a complete manifold.*

Before proving the theorem we state and prove two lemmas. One of them is the Gauss lemma. There are at least two proofs to the lemma. We will take the proof that uses Jacobi fields. The following notations will be used in the proof of the lemmas. Let $p \in \mathbf{M}$. We define, for $r > 0$, $B_0(0, r) = \{v \in T_p\mathbf{M} \mid \|v\| < r\}$. Also $B(p, r)$ is the metric ball on $\mathbf{M}$, i.e. $B(p, r) = \{x \in \mathbf{M} \mid d(x, p) < r\}$.

**Lemma 3.4.1** *(Gauss lemma) Let $u, v \in T_p\mathbf{M}$ and $c = c(s)$ be the geodesic $exp_p(sv)$. Then*

$$g_{c(s)}(Dexp_p|_{sv}v, Dexp_p|_{sv}u) = g_p(v, u). \tag{3.4.6}$$

*In particular $c'(s)$ is orthogonal to a smooth geodesic sphere centered at $p$ and with radius $d(c(s), p)$, i.e. $c'(s)$ is orthogonal to*

$$\{exp_p(d(c(s), p)v) \,|\, v \in T_p\mathbf{M},\ \|v\| = 1\},$$

*provided that the geodesic sphere is smooth.*

Proof. By the chain rule, we have $c'(s) = Dexp_p|_{sv}v$. By Proposition 3.4.3, we also know that

$$Dexp_p|_{sv}u = \frac{1}{s}Y(s),$$

where $Y(s)$ is the Jacobi field along $c = c(s)$ such that $Y(0) = 0$ and $Y'(0) = u$. Hence

$$h(s) \equiv g_{c(s)}(Dexp_p|_{sv}v, Dexp_p|_{sv}u) = g_{c(s)}(c'(s), \frac{1}{s}Y(s)).$$

Since $Y(s)$ is a Jacobi field and $c = c(s)$ is a geodesic, we have

$$[sh(s)]'' = \frac{\partial^2}{\partial s^2} g_{c(s)}(c'(s), Y(s)) = g_{c(s)}(c'(s), \\ Y''(s)) = -R(Y, c', c', c') = 0.$$

Therefore $h(s)$ is a constant. Using $Y(0) = 0$, we know that

$$\frac{1}{s}Y(s) \to Y'(0) = \nabla_{c'(s)}Y(s)|_{s=0} = u$$

when $s \to 0$. This shows $h(s) = h(0) = g_p(v, u)$, proving (3.4.6).

Finally we just note that $g_p(v, u) = 0$ for any $u \in T_{sv}T_p\mathbf{M}$ such that $u$ is tangent to $\partial B_0(0, s)$, which is the sphere in $T_p\mathbf{M}$, centered at 0, with radius $s$. Here we are using the equivalence of $T_p\mathbf{M}$ and $T_{sv}T_p\mathbf{M}$ again. By the just proven formula (3.4.6), we know that $c'(s)$ is orthogonal to the geodesic sphere. □

**Lemma 3.4.2** *(1) For any point $p \in \mathbf{M}$, there exists $\epsilon > 0$ such that $exp_p$ is a diffeomorphism from the ball $B_0(0, \epsilon) \subset T_p\mathbf{M}$ onto the metric ball*

$$B(p, \epsilon) = \{x \in \mathbf{M} \,|\, d(x, p) < \epsilon\} \subset \mathbf{M}$$

*i.e.*

$$exp_p(B_0(0,\epsilon)) = B(p,\epsilon).$$

*Moreover, for any unit tangent vector* $v \in T_p\mathbf{M}$*, the geodesic* $c(s) = exp_p(sv)$*,* $s \in [0,\epsilon]$ *is distance minimizing.*

*(2) For any* $q \in B(p,\epsilon)^c$ *where* $\epsilon$ *is the same as in (1), there exists* $p_1 \in \partial B(p,\epsilon)$ *such that*

$$d(p,q) = d(p,p_1) + d(p_1,q) = \epsilon + d(p_1,q).$$

PROOF. (1) We choose $\epsilon$ sufficiently small, so that $exp_p$ is a diffeomorphism on $B_0(0,\epsilon)$. Pick a point $p_1 \in exp_p(\partial B_0(0,\epsilon))$. Then there exists a unit vector $v$ in $T_p\mathbf{M}$ such that $p_1 = exp_p(\epsilon v)$. We will show that $d(p,p_1) = \epsilon$, i.e. $p_1 \in \partial B(p,\epsilon)$, and the geodesic $c(s) = exp_p(sv)$, $s \in [0,\epsilon]$ is distance minimizing.

Let $\sigma = \sigma(s)$, $s \in [0,a]$, $a > 0$, be a smooth curve joining $p$ and $p_1$. First we assume $\sigma$ stays in $exp_p(B_0(0,\epsilon))$ for all $s \in [0,a]$. We write, for a function $r : [0,a] \to [0,\infty)$ and unit vectors $v(s) \in T_p\mathbf{M}$ so that

$$\sigma(s) = exp_p(r(s)\,v(s)).$$

By the chain rule

$$\sigma'(s) = r'(s) Dexp_p|_{r(s)v(s)} v(s) + r(s) Dexp_p|_{r(s)v(s)} v'(s).$$

Here $v'(s)$ is just $\frac{d}{ds}v(s)$ since $v(s)$ is regarded as a vector in the Euclidean space. Since $v(s)$ is a unit vector, the Gauss lemma tells us

$$g(Dexp_p|_{r(s)v(s)} v(s), Dexp_p|_{r(s)v(s)} v(s)) = g|_p(v(s),v(s)) = 1.$$

Differentiating $g_p(v(s),v(s)) = 1$ with respect to $s$, we know that $g_p(v(s),v'(s)) = 0$. Applying Gauss lemma again, we have

$$g(Dexp_p|_{r(s)v(s)} v(s), Dexp_p|_{r(s)v(s)} v'(s)) = g|_p(v(s),v'(s)) = 0.$$

Therefore

$$\begin{aligned} g(\sigma'(s),\sigma'(s)) =& |r'(s)|^2 + r^2(s) g(Dexp_p|_{r(s)v(s)} v'(s), Dexp_p|_{r(s)v(s)} v'(s)) \\ &\geq |r'(s)|^2. \end{aligned}$$

This shows

$$L(\sigma) = \int_0^a \sqrt{g(\sigma'(s),\sigma'(s))} ds \geq |\int_0^a r'(s) ds| = r(a) - r(0) = \epsilon.$$

Observe that $Dexp_p|_{r(s)v(s)}$ is nonsingular since $r(s)v(s)$ is sufficiently close to the origin of $T_p\mathbf{M}$. Hence the equality holds if and only if $v'(s) = 0$. This means that the infimum of the lengths of $\sigma$ is $\epsilon$, and if the length of $\sigma$ is $\epsilon$, then $\sigma(s) = exp_p(sv(0))$ and $\sigma$ is a geodesic. Let $\sigma$ be such a geodesic. Since $exp_p$ is a diffeomorphism on $B_0(0,\epsilon)$ and $exp_p(r(a)v(0)) = exp_p(\epsilon v) = p_1$, we know that $r(a)v(0) = \epsilon v$. Therefore $\sigma$ is just the geodesic $c$ at the beginning of the proof, up to reparametrization.

If $\sigma$ goes out of $exp_p(B_0(0,\epsilon))$, then at some point it crosses the boundary. By the above argument, its length is greater than $\epsilon$. In any case, we have shown that $d(p,p_1) = \epsilon$ and $c = exp_p(sv)$ is distance minimizing. We have also shown that $exp_p(\partial B_0(0,\epsilon)) \subset \partial B(p,\epsilon)$. This inclusion obviously holds when $\epsilon$ is replaced by any smaller positive number. Consequently

$$exp_p(B_0(0,\epsilon)) \subset B(p,\epsilon).$$

To finish the proof, we just need to show the reversal of the inclusion holds. Pick a point $q \in [exp_p(B_0(0,\epsilon))]^c$. Let $\sigma = \sigma(s)$, $s \in [0,a]$, $a > 0$, be a smooth curve joining $p$ and $q$. Then there exists a number $\alpha \in (0,a]$ such that the point $\sigma(\alpha) \in exp_p(\partial B_0(0,\epsilon))$. In the last paragraph, we have shown that $L(\sigma|_0^\alpha) \geq \epsilon$. Therefore $L(\sigma|_0^a) \geq \epsilon + L(\sigma|_a^\alpha)$. Minimizing the inequality over all smooth curves connecting $p$ and $q$, we see that

$$d(p,q) \geq \epsilon + d(p_1,q) \geq \epsilon. \tag{3.4.7}$$

Here $p_1$ is certain point on $exp_p(\partial B_0(0,\epsilon))$ whose compactness assures the existence of $p_1$. Hence $q \in B(p,\epsilon)^c$. Therefore

$$[exp_p(B_0(0,\epsilon))]^c \subset B(p,\epsilon)^c$$

and eventually

$$exp_p(B_0(0,\epsilon)) = B(p,\epsilon),$$

proving (1).

PROOF. (2) This is immediate from (3.4.7) and the conclusion that $\partial exp_p(B_0(0,\epsilon)) = \partial B(p,\epsilon)$. □

Now we commence:

*the proof of the Hopf-Rinow theorem.*

The main work is to show just one of (i)–(iv) implies (v). Once this is done, the rest of the proof is quite routine. Clearly (iii) is equivalent to (iv) and (iii) implies (ii). The order of the proof is: (iii) → (v), (iii) → (ii) → (i) → (iii).

PROOF. (iii) to (v).

Suppose $exp_p$ is defined on the whole $T_p\mathbf{M}$ for any $p \in \mathbf{M}$.

Pick two points $p, q \in \mathbf{M}$. Let $\epsilon > 0$ be sufficiently small. By part (2) of the previous lemma, there exists a point $p_1 \in \partial B(p, \epsilon)$ such that

$$d(p, q) = d(p, p_1) + d(p_1, q).$$

By part (1) of the previous lemma, there exists a unit vector $v \in T_p\mathbf{M}$, such that $p_1 = exp_p(\epsilon v)$. By assumption, the $c(t) = exp_p(tv)$ is defined for all $t > 0$. Define

$$I = \{\, t > 0 \,|t + d(c(t), q) = d(p, q)\}; \qquad T = \sup(I \cap [0, d(p, q)]).$$

If we can show that $T = d(p, q)$, then

$$\begin{aligned} d(p, q) \leq & d(p, c(T)) + d(c(T), q) \leq L(c|_0^T) + d(c(T), q) \\ &= T + d(c(T), q) = d(p, q). \end{aligned}$$

Thus $d(c(T), q) = 0$ and $d(p, c(T)) = T$. So $c$ is a distance minimizing geodesic connecting $p$ and $q$.

We use the method of contradiction to show that $T = d(p, q)$. Suppose $T < d(p, q)$. Applying part (1) of the previous lemma for the points $c(T)$ and $q$, we can find $\epsilon > 0$ and $p_2 \in \partial B(c(T), \epsilon)$ such that

$$\epsilon + d(p_2, q) = d(c(T), q).$$

By definition of $T$, it holds $T + d(c(T), q) = d(p, q)$. Therefore

$$\epsilon + d(p_2, q) = d(p, q) - T.$$

This implies

$$\begin{aligned} d(p, q) &\leq d(p, p_2) + d(p_2, q) \\ &\leq d(p, c(T)) + d(c(T), p_2) + d(p_2, q) \leq T + \epsilon + d(p_2, q) = d(p, q). \end{aligned} \tag{3.4.8}$$

Hence all inequalities here become equalities. In particular

$$d(p, p_2) = d(p, c(T)) + d(c(T), p_2) = T + \epsilon.$$

Let $\gamma : [0, \epsilon]$ be the minimal geodesic connecting $c(T)$ and $p_2$. Then the concatenated curve $c_1 \equiv c|_0^T \cup \gamma|_0^\epsilon$ is at least a piecewise smooth curve with length $T + \epsilon$. Since this curve is distance minimizing, the first variation formula for the distance shows that $c_1$ is a smooth geodesic.

Note $c$ and $c_1$ coincides in an open interval. The uniqueness of the solutions of the geodesic equation tells that $c_1$ is an extension of $c|_0^T$, i.e. $c_1 = exp_p(tv)$, $t \in [0, T+\epsilon]$. Moreover

$$p_2 = exp_p((T+\epsilon)v) = c(T+\epsilon).$$

We have shown that $c_1$ is a minimal geodesic connecting $p$ and $p_2$. Hence any segment of $c_1$ is a minimal geodesic. It is clear that (3.4.8) holds when $\epsilon$ is replaced by any $\epsilon_1 \in [0, \epsilon]$. Therefore

$$T + \epsilon_1 + d(c(T+\epsilon_1), q) = d(p, q), \quad \epsilon_1 \in [0, \epsilon].$$

This is a contradiction with the definition of $T$. We have proven that (iii) implies (v).

PROOF. (iii) to (ii) to (i).

Since (ii) is a special case of (iii), we just need to prove that (ii) implies (i), i.e. if $exp_p$ is defined on whole $T_p\mathbf{M}$ for one $p$, then $\mathbf{M}$ is a complete metric space.

Let $\{q_j\}$ be a Cauchy sequence. By (v) (as a consequence of (iii)), there exists a sequence of unit vectors $\{v_j\} \subset T_p\mathbf{M}$ such that $q_j = exp_p(d(p, q_j)v_j)$. Since the unit sphere of $T_p\mathbf{M}$ is compact, a subsequence of $\{v_j\}$, converges to a unit vector $w \in T_p\mathbf{M}$. Because $|d(p, q_j) - d(p, q_k)| \le d(q_j, q_k)$, we know that $\{d(p, q_j)\}$ is a Cauchy sequence of real numbers. We can regard $q_j$ as a point on the geodesic starting from $p$ with initial velocity $v_j$. Suppose $d(p, q_j) \to r$ when $j \to \infty$. Then $q_j \to exp_p(rw)$ by the continuous dependence of finite time solutions of the geodesic equation on the initial value. Hence $\mathbf{M}$ is complete metric space.

Finally we need the

PROOF. (i) to (iii).

Suppose $\mathbf{M}$ is a complete metric space. Pick $p \in \mathbf{M}$ and $v \in T_p\mathbf{M}$. Let $T$ be the supremum of $t$ such that $exp_p(tv)$ is defined. Suppose $T$ is finite. Choose a sequence $\{t_i\}$ such that $t_i \to T$ when $i \to \infty$. Then the sequence $\{exp_p(t_i v)\}$ is a Cauchy sequence in $\mathbf{M}$. Let $q$ be the limit of this sequence. From geodesic equation, we can extend $exp_p(tv)$ smoothly beyond $T$, using $q$ as the initial point. This shows $T = \infty$. The proof of Hopf-Rinow theorem is done. $\square$

In geometric analysis, one often needs to compute the Laplacian of the distance function. Using Jacobi fields, one can convert the second order differentiation into the index form which is an integral expression involving the curvature of the manifold. Needless to say, this has important implications. For instance, one can obtain bounds on the Laplacian of the distance function if there are certain bounds on the curvature. This kind of result is referred to as Laplacian comparison theorems. One can also derive similarly volume comparison results, which will be presented in the next section.

**Proposition 3.4.6** *For $p, x \in \mathbf{M}$, a complete Riemann manifold of dimension $n$, let $r = r(x) = d(x,p)$ be the distance function. Let $c : [0,a] \to \mathbf{M}$ be a geodesic joining $p$ and $x$ parametrized by arclength. Suppose $c$ does not intersect with the cut-locus of $p$. Let $\{e_1, \ldots, e_{n-1}, c'(0)\}$ be an orthonormal basis of $T_p\mathbf{M}$ and $\{e_1(s), \ldots, e_{n-1}(s), c'(s)\}$ be an orthonormal basis of $T_{c(s)}\mathbf{M}$ where $e_i(s)$ is the parallel transport of $e_i$ along $c$. For $i = 1,,,, n-1$, let $X_i$ be the Jacobi field along $c$ such that $X_i(0) = 0$ and $X_i(a) = e_i(a)$, which exist by Remark 3.4.5. Then the following identities hold:*

$$\Delta r(x) = \Sigma_{i=1}^{n-1} \int_0^{d(x,p)} (|X_i'|^2 + R(c'(s), X_i, c'(s), X_i) ds = \Sigma_{i=1}^{n-1} I(X_i, X_i) \tag{3.4.9}$$

*where $X_i'(s) = \nabla_{c'(s)} X_i(s)$, $I$ is the index form;*

$$\Delta r(x) = \partial_s|_{s=a} \sqrt{det(g(X_i, X_j))}, \quad a = d(x,p); \tag{3.4.10}$$

$$\Delta r(x) = \frac{n-1}{r} + \partial_s|_{s=a} \ln \sqrt{det g_e}. \tag{3.4.11}$$

*Here $(g(X_i, X_j))$ is the $n$ by $n$ matrix with $X_1, \ldots X_{n-1}$ being the above Jacobi fields and $X_n = c'(s)$. Also $det g_e$ is the determinant of the matrix $g(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})$ under a canonical basis in the exponential coordinates.*

PROOF. For simplicity, we denote $c'(s)$ by $X_n$. By (3.3.7) and the definition of the gradient of a function, at the point $x$,

$$\begin{aligned} \Delta r &= trace\, Hess\ r = X_i X_i r - (\nabla_{X_i} X_i) r \\ &= X_i < X_i, grad\, r > - < \nabla_{X_i} X_i, grad\, r > \\ &= < X_i, \nabla_{X_i} \frac{\partial}{\partial r} > \qquad \text{since} \quad grad\ r = \frac{\partial}{\partial r}. \end{aligned}$$

Here and later all terms are summed from $i = 1$ to $i = n$, unless stated otherwise. Note that we may extend $X_i$ to be a vector field in a neighborhood of $x$ so that $[X_i, \frac{\partial}{\partial r}] = 0$. Therefore, the fundamental theorem of Riemann geometry tells us

$$\begin{aligned}\Delta r &= < X_i, \nabla_{\frac{\partial}{\partial r}} X_i > = < X_i, \nabla_{c'(a)} X_i > \\ &= \int_0^{d(x,p)} \frac{d}{ds} < X_i, \nabla_{c'(s)} X_i > ds \\ &= \int_0^{d(x,p)} \left[ < \nabla_{c'(s)} X_i, \nabla_{c'(s)} X_i > + < X_i, \nabla_{c'(s)} \nabla_{c'(s)} X_i > \right] ds\end{aligned}$$

Recall that $X_n = c'(s)$ and hence $\nabla_{c'(s)} X_n = 0$ since $c$ is a geodesic. Hence

$$\Delta r = \Sigma_{i=1}^{n-1} \int_0^{d(x,p)} \left[ < \nabla_{c'(s)} X_i, \nabla_{c'(s)} X_i > + < X_i, \nabla_{c'(s)} \nabla_{c'(s)} X_i > \right] ds.$$

Since $X_i$, $i = 1, \ldots, n-1$ are Jacobi fields, we have, by definition,

$$\nabla_{c'(s)} \nabla_{c'(s)} X_i + R(X_i, c'(s)) c'(s) = 0.$$

The proof of the first identity (3.4.9) is done by combining the last two equalities.

Next we prove the second identity. Write the $n$ by $n$ matrix

$$(g(X_i, X_j)) \equiv B.$$

We compute

$$\begin{aligned}\partial_s \sqrt{detB} &= \frac{1}{2} (detB)^{-1/2} \partial_s detB \\ &= \frac{1}{2} (detB)^{1/2}\, tr(B' B^{-1}).\end{aligned}$$

Here $B' = (\partial_s g(X_i, X_j))$. When $s = a$, the matrix $B$ is the identity matrix by construction. Therefore

$$\partial_s|_{s=a} \sqrt{detB} = \frac{1}{2}\ trB' = \Sigma_{i=1}^n g(X_i'(a), X_i(a)).$$

Note $X_n = c'(s)$ and hence $g(X_n'(a), X_n(a)) = 0$. For $i = 1, \ldots, n-1$, $X_i$ are Jacobi fields. So $g(X_i'(a), X_i(a)) = I(X_i(a), X_i(a))$ (see (3.4.5)). Now the first formula (3.4.9) tells us

$$\Delta r(x) = \partial_s|_{s=a} \sqrt{detB}.$$

This is the second identity on $\Delta r$.

Finally, we prove the last equality involving $\Delta r$. Let $Z_i = Z_i(s)$, $i = 1, ..n-1$ be Jacobi fields along the same curve $c$ such that $Z_i(0) = 0$ and $Z_i'(0) = e_i$. We claim that there exist a constant matrix $(b_{ij})$ such that $X_i(s) = b_{ij}Z_j(s)$. The reason is that the Jacobi equation which both $X_i$ and $Z_i$ satisfy is second order and linear. If we find $(b_{ij})$ such that $X_i'(0) = b_{ij}Z_j'(0)$, then the two Jacobi fields $X_i(s)$ and $b_{ij}Z_j(s)$ will have the same initial value 0 and the same initial derivative. Hence they must be the same throughout. Denote by $A = A(s)$ the $n$ by $n$ matrix: $(g(Z_i(s), Z_j(s)))$ where $Z_i$, $i = 1, \ldots, n-1$ are the above Jacobi fields and $Z_n(s) = c'(s)$. Then there exists a constant matrix $P$ such that

$$B(s) = (g(X_i, X_j)) = PA(s).$$

Observe that

$$\begin{aligned}\partial_s detB &= detB\, tr(B'B^{-1}) = detB\, tr(PA'A^{-1}P^{-1})\\ &= detB\, tr(A'A^{-1}).\end{aligned}$$

When $s = a$, we have $detB = 1$. Therefore

$$\partial_s detB = \frac{detA\, tr(A'A^{-1})}{detA} = \frac{\partial_s detA}{detA}.$$

Hence

$$\partial_s \sqrt{detB} = \frac{\partial_s \sqrt{detA}}{\sqrt{detA}}.$$

By Proposition 3.4.5, for $i = 1, \ldots, n-1$, $Z_i(s) = s\frac{\partial}{\partial x^i}$ and also $Z_n(s) = c'(s)$. Here $\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^{n-1}}, c'(s)\}$ is a canonical basis for $T_{c(s)}\mathbf{M}$ under the exponential coordinates. Consequently

$$\sqrt{detA} = s^{n-1}\sqrt{detg_e}.$$

This shows, for $r = a$,

$$\Delta r(x) = \partial_s|_{s=a}\sqrt{detB} = \frac{\partial_s\sqrt{detA}}{\sqrt{detA}} = \frac{n-1}{r} + \partial_s|_{s=a} \ln \sqrt{detg_e}.$$

This proves (3.4.11). □

**Remark 3.4.5** *The Jacobi field $X_i$ exists and is explicitly given by $X_i(s) = \frac{\partial \beta}{\partial t}(s,t)|_{t=0}$ where $\beta(s,t) = exp_p(sc'(0) + stv_i)$ and $v_i \in T_p\mathbf{M}$*

*satisfies* $Dexp_p|_{ac'(0)}(av_i) = e_i(a)$*. Here is why. In the proof of Proposition 3.4.3, it was shown that* $\frac{\partial\beta}{\partial t}(s,0)$ *is a Jacobi field along the curve* $c$*. Also* $X_i(0) = 0$ *and* $X_i(a) = Dexp_p|_{ac'(0)}(av_i) = e_i(a)$*.*

The following theorem states that the index form reaches minimal value on a Jacobi field along a distance minimizing geodesic.

**Theorem 3.4.2** *Let* $c : [0,a] \to \mathbf{M}$ *be a distance minimizing geodesic, i.e.* $c$ *does not intersect with the cut-locus of* $c(0)$*. If* $Y$ *is a Jacobi field and* $X$ *a vector field along* $c$ *with the same values as* $Y$ *at the ends of* $c$*, then* $I(X,X) \geq I(Y,Y)$*.*

PROOF. Note that $X - Y$ vanishes at the ends of $c$, which is distance minimizing. By the second variation formula of distance (Proposition 3.4.2),

$$I(X - Y, X - Y) \geq 0.$$

Using integration parts as in the proof of (3.4.4), it is easy to check that $I(Y,Y) = g(Y',Y)|_0^a$ and $I(X,Y) = g(Y',X)|_0^a = I(Y,X)$. Therefore $I(X,Y) = I(Y,Y)$ and

$$I(X,X) - I(Y,Y) = I(X - Y, X - Y) \geq 0.$$

□

The next result, often called the basic index theorem, says that a geodesic with no conjugate points also satisfies the conclusion of the previous theorem. Unlike the previous theorem, it is not assumed that the geodesic is distance minimizing. An immediate consequence of the theorem is that a geodesic with no conjugate points is length minimizing among curves in a sufficiently small neighborhood of itself.

**Theorem 3.4.3** *(Basic index theorem) Let* $c : [0,a] \to \mathbf{M}$ *be a geodesic, connecting points* $p$ *and* $q$*. Suppose the curve* $c$ *does not contain any conjugate point of* $p$*. If* $Y$ *is a Jacobi field and* $X$ *a vector field along* $c$ *such that* $Y(0) = X(0) = 0$ *and* $Y(a) = X(a)$*, then* $I(X,X) \geq I(Y,Y)$*. Furthermore, the equality holds if and only if* $X = Y$*.*

PROOF. Pick a basis $\{v_1, \ldots, v_n\}$ for $T_q\mathbf{M}$. By the remark before the previous theorem, there exist Jacobi fields $Y_i$ along $c$, $i = 1, \ldots, n$ such that $Y_i(0) = 0$ and $Y_i(a) = v_i$. By the assumption that there exists no conjugate points along $c$, we know that $Y_i$ is unique. Here we are using the fact that the Jacobi equation is linear. If there are two Jacobi fields with the same end points then their difference is a nontrivial Jacobi field vanishing at the ends. Thus the ends are conjugate with each other. Expanding $Y_i$ along a parallel orthonormal frame along $c$, and consider the differential equation satisfied by the coefficient matrix, it is easy to see that $\{Y_1(s), \ldots, Y_n(s)\}$ is a basis for $T_{c(s)}\mathbf{M}$ when $s \neq 0$.

Now we pick a Jacobi field $Y$ and a vector field $X$ along $c$ such that $Y(0) = X(0) = 0$ and $Y(a) = X(a)$. There exist constants $b_i$ and functions $f_i$ such that

$$Y(s) = \Sigma_{i=1}^n b_i Y_i(s), \quad X(s) = \Sigma_{i=1}^n f_i(s) Y_i(s), \quad f_i(a) = b_i.$$

Since $Y_i$ is Jacobi, formula (3.4.5) shows

$$I(Y, Y) = \Sigma_{i,j=1}^n b_i b_j g(Y_i'(a), Y_j(a)).$$

Next we compute

$$\begin{aligned} I(X, X) &= \int_0^a [g(X', X') + R(X, c', X, c')] ds \\ &= \int_0^a \Sigma_{i,j=1}^n f_i f_j g(Y_i'(s), Y_j'(s)) ds + \int_0^a \Sigma_{i,j=1}^n f_i' f_j' g(Y_i(s), Y_j(s)) ds \\ &\quad + 2\int_0^a \Sigma_{i,j=1}^n f_i' f_j g(Y_i'(s), Y_j(s)) ds \\ &\quad + \int_0^a \Sigma_{i,j=1}^n f_i f_j R(Y_i(s), c'(s), Y_j(s), c'(s)) ds. \end{aligned} \tag{3.4.12}$$

Here we have used the equality $g(Y_i'(s), Y_j(s)) = g(Y_i(s), Y_j'(s))$, which is easily verified by differentiation. The first integral on the second line of the computation satisfies

$$\begin{aligned} &\int_0^a \Sigma_{i,j=1}^n f_i f_j g(Y_i'(s), Y_j'(s)) ds \\ &= \int_0^a \Sigma_{i,j=1}^n f_i f_j [\partial_s g(Y_i'(s), Y_j(s)) - g(Y_i''(s), Y_j(s))] ds \\ &= \Sigma_{i,j=1}^n f_i(a) f_j(a) g(Y_i'(a), Y_j(a)) - 2\int_0^a \Sigma_{i,j=1}^n f_i' f_j g(Y_i'(s), Y_j(s)) ds \\ &\quad + \int_0^a \Sigma_{i,j=1}^n f_i f_j R(Y_i(s), c'(s), c'(s), Y_j(s)) ds \end{aligned}$$

where we have used the Jacobi equation and $g(Y_i'(s), Y_j(s)) = g(Y_i(s), Y_j'(s))$ again. Plugging this to (3.4.12), we see that all but two terms cancel to give

$$I(X,X) = \Sigma_{i,j=1}^n f_i(a)f_j(a)g(Y_i'(a),Y_j(a)) + \int_0^a \Sigma_{i,j=1}^n f_i' f_j' g(Y_i(s),Y_j(s))ds \geq I(Y,Y).$$

The equality holds if and only if $\Sigma_{i=1}^n f_i'(s)Y_i(s) \equiv 0$, i.e. $f_i(s) = f_i(a) = b_i$, or $X = Y$. □

## 3.5 Integration and volume comparison

In the following, we define the canonical measure on a Riemann manifold. This is amount to choosing the canonical volume form or a suitable weight for integration.

**Definition 3.5.1** *(canonical volume form and Riemann integral) Let* $\mathbf{M}$ *be an orientable Riemann manifold equipped with the Riemann metric g. Let* $(U,\phi)$, *with* $\phi = (x^1,\ldots,x^n)$ *be a local chart. Then the canonical volume form on U is*

$$d\mu = \sqrt{det(g_{ij})}dx^1 \wedge \ldots \wedge dx^n$$

*Let h be a smooth function on* $\mathbf{M}$. *Then the Riemann integral of h on U is*

$$\int_U h d\mu \equiv \int_U h\sqrt{det(g_{ij})}dx^1 \wedge \ldots \wedge dx^n.$$

*The Riemann integral on the whole manifold is defined via partition of unity and the localized definition for Riemann integrals.*

**Remark 3.5.1** *(1) The second integral in the above definition is the integral of the form* $h\sqrt{det(g_{ij})}dx^1 \wedge \ldots \wedge dx^n$ *as defined in Definition 3.1.12, which says*

$$\int_U h\sqrt{det(g_{ij})}dx^1 \wedge \ldots \wedge dx^n = \int_{\phi(U)} \left[h\sqrt{det(g_{ij})}\right] \circ \phi^{-1} dx^1 \ldots dx^n.$$

*(2) The appearance of the function* $\sqrt{det(g_{ij})}$ *in the definition of canonical volume form makes such integral formulas as Green's formula valid. Indeed, let u, v be two smooth functions on a compact Riemann manifold* $\mathbf{M}$, *then*

$$\int_{\mathbf{M}} v\Delta u \; d\mu = -\int_{\mathbf{M}} \nabla v \nabla u \; d\mu, \qquad \nabla v \nabla u \equiv g(\nabla v, \nabla u).$$

*A way to understand the formula is to look at local formulas for* $\Delta$ *and* $d\mu$. *Let* $(U, \phi)$ *be the local chart in the above definition. Recall from (3.3.5) that, for* $\sqrt{g} \equiv \sqrt{det(g_{ij})}$,

$$\Delta u = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left( \sqrt{g}\, g^{ij} \frac{\partial u}{\partial x^j} \right).$$

*Hence*

$$\int_U v\Delta u\, d\mu = \int_{\phi(U)} v \circ \phi^{-1} \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left( \sqrt{g}\, g^{ij} \frac{\partial (u \circ \phi^{-1})}{\partial x^j} \right) \times \sqrt{det(g_{ij})} dx^1 \ldots dx^n.$$

*Here we have used the convention (cf Definition 3.1.4) that* $\frac{\partial u}{\partial x^j} = \frac{\partial u \circ \phi^{-1}}{\partial x^j}$. *Also the functions* $\sqrt{g}$ *and* $g^{ij}$ *on the right-hand side are regarded* $\sqrt{g} \circ \phi^{-1}$ *and* $g^{ij} \circ \phi^{-1}$ *respectively. Hence they are functions on the Euclidean domain* $\phi(U)$.

*Observe that the factor* $\frac{1}{\sqrt{g}}$ *coming out of* $\Delta u$ *cancels with the factor* $\sqrt{det(g_{ij})}$ *which comes out of* $d\mu$. *Therefore*

$$\int_U v\Delta u\, d\mu = \int_{\phi(U)} v \circ \phi^{-1} \frac{\partial}{\partial x^i} \left( \sqrt{g}\, g^{ij} \frac{\partial (u \circ \phi^{-1})}{\partial x^j} \right) dx^1 \ldots dx^n.$$

*If* $v$ *vanishes on the boundary of* $U$, *then using integration by parts in Euclidean space, we see that*

$$\int_U v\Delta u\, d\mu = -\int_{\phi(U)} \frac{\partial}{\partial x^i} (v \circ \phi^{-1}) \left( \sqrt{g}\, g^{ij} \frac{\partial (u \circ \phi^{-1})}{\partial x^j} \right) dx^1 \ldots dx^n.$$

*By the local formula for* $\nabla u$ *in (3.3.2), we deduce*

$$\int_U v\Delta u\, d\mu = -\int_U \nabla v \nabla u\, d\mu.$$

*In general the function* $v$ *may not vanish on the boundary* $\partial U$. *But this does not make the proof too much longer. Let* $\{(U_i, \phi_i)\}$ *be a family of local charts such that* $\{(U_i, h_i)\}$ *is a partition of unity for* $\mathbf{M}$. *Since* $\mathbf{M}$ *is compact, this is a finite family. Therefore*

$$\int_{\mathbf{M}} v\Delta u d\mu = \Sigma_i \int_{U_i} h_i v \Delta u d\mu.$$

*The function $h_i v$ is zero on $\partial U_i$ since $h_i$ is the partition function. Hence*

$$\int_{\mathbf{M}} v\Delta u d\mu = -\Sigma_i \int_{U_i} \nabla(h_i v)\nabla u d\mu$$
$$= -\Sigma_i \int_{U_i} h_i \nabla v \nabla u d\mu - \Sigma_i \int_{U_i} v\nabla h_i \nabla u d\mu.$$

*Since $\Sigma_i h_i = 1$, this shows*

$$\int_{\mathbf{M}} v\Delta u \; d\mu = -\int_{\mathbf{M}} \nabla v \nabla u \; d\mu.$$

It is important to possess an efficient formula for computing volume of a manifold. When a geodesic ball does not intersect the cut-locus of its center, we can use the exponential map and associated Jacobi fields to construct such a formula. The formula below seems to make the matter worse by getting Jacobi fields involved. However the differential equations satisfied by Jacobi fields make the formula useful.

**Proposition 3.5.1** *Let $\mathbf{M}$ be a complete Riemann manifold. Suppose the ball $B(p,r)$ does not intersect the cut-locus of $p$. For each unit vector $v \in T_p\mathbf{M}$, let $\{e_1, \ldots, e_{n-1}, v\}$ be an orthonormal basis for $T_p\mathbf{M}$. Then*

$$|B(p,r)| = \int_{S^{n-1}} \int_0^r \sqrt{\det(g(Y_i(s), Y_j(s))} ds dv.$$

*Here $Y_i$ is the Jacobi field along $c = c(t) = exp_p(tv)$ such that $Y_i(0) = 0$ and $Y_i'(0) = e_i$, $i = 1, \ldots, (n-1)$, and $dv$ is the canonical volume element of $S^{n-1}$, regarded as the unit sphere of $T_p\mathbf{M}$.*

Proof. Since the ball $B(p,r)$ does not intersect the cut-locus, we will use the inverse exponential map $exp_p^{-1}$ as the local chart $\phi$ in the definition of the volume form. Let $\{e_1, \ldots, e_{n-1}, v\}$ be an orthonormal basis of $T_p\mathbf{M}$. Every point $m$ in $B(p,r)$ is represented in this chart by the coordinates $\{x^1, \ldots, x^n\}$, i.e. $m = exp_p(x^1 e_1 + \cdots x^{n-1} e_{n-1} + x^n v)$. For any fixed $s \in [0,r]$, let $q = exp_p(sv)$. Then, by Proposition 3.6.2 part (3), the point $q$ is not a conjugate point of $p$. Hence $Dexp_p|_{sv}$ is nonsingular and

$$\{Dexp_p|_{sv} e_1, \ldots, Dexp_p|_{sv} e_{n-1}, Dexp_p|_{sv} v\}$$

is a basis of $T_q\mathbf{M}$. In fact it is nothing but the local basis $\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\}$. One just needs to verify that, for any smooth function $f$ on $B(p,r)$, by definition

$$\frac{\partial f}{\partial x^i}|_q = \left[\frac{d}{dl}|_{l=0} exp_p(sv + le_i)\right] f = [Dexp_p|_{sv} e_i] f.$$

Let $c = c(s)$ be the geodesic $exp_p(sv)$. According to Proposition 3.4.3,

$$Dexp_p\Big|_{sv}(se_i) = Y_i(s)$$

where $Y_i$ is the Jacobi field along $c = c(t)$ such that $Y_i(0) = 0$ and $Y_i'(0) = e_i$, $i = 1, \ldots, (n-1)$. Also, by the chain rule, it is clear that $Dexp_p|_{sv}v = c'(s)$. Hence

$$\{\frac{\partial}{\partial x^1}|_q, \ldots, \frac{\partial}{\partial x^n}|_q\} = \{\frac{1}{s}Y_1(s), \ldots, \frac{1}{s}Y_{n-1}(s), c'(s)\}$$

is the canonical basis of $T_q\mathbf{M}$. In the local chart generated by $exp_p^{-1}$, the coordinates of $q$ in the spherical system is $(v, s)$ where $v$ is regarded an element in $S^{n-1}$. Therefore the volume element at $q$, under this local chart is

$$d\mu = \sqrt{\det(g(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}))} dx^1 \ldots dx^n = \sqrt{\det(g(Y_i(s), Y_j(s))} dsdv$$

where $dv$ is the canonical volume element of $S^{n-1}$, regarded as the unit sphere of $T_p\mathbf{M}$. After integration, we obtain

$$|B(p,r)| = \int_{S^{n-1}} \int_0^r \sqrt{\det(g(Y_i(s), Y_j(s))} dsdv. \quad \Box$$

**Remark 3.5.2** *The function* $J(v,s) \equiv \sqrt{\det(g(Y_i(s), Y_j(s))}$ *is often called the volume element in the geodesic spherical coordinates.*

**Exercise 3.5.1** *Let* $\mathbf{M}$ *be a complete Riemann manifold and* $p \in \mathbf{M}$. *Prove that the cut-locus* $C_p$ *has zero measure and* $\mathbf{M} - C_p$ *is star shaped.*

Examples. In $\mathbf{R^n}$ with standard flat metric, $J(v,s) = s^{n-1}$. In $S^n$ with the standard metric $J(v,s) = (sins)^{n-1}$ and in $H^n$, the hyperbolic space with sectional curvature $-1$, $J(v,s) = (sinhs)^{n-1}$.

When dealing with a complicated manifold, it may be impossible to compute geometric quantities such as the volume explicitly. Therefore one would like to compare these quantities with those on canonical manifolds. The three basic canonical manifolds are the Euclidean space $\mathbf{R^n}$, the $n$ dimensional sphere $S^n$ and the hyperbolic space $\mathbf{H^n}$. These are manifolds with constant zero, positive and negative sectional curvatures respectively. Volume comparison theorems allow us to bound the volume of the geodesic balls of a manifold by those on the canonical manifolds if the curvatures are bounded suitably. Case (i) of the following theorem with some restriction is due to Bishop originally and to Gromov in its current form. Case (ii) is due to Gunther.

**Theorem 3.5.1** *(Classical volume comparison theorem) Let* $\mathbf{M}$ *be a complete Riemann manifold and* $B(p,r)$ *be the ball of radius* $r$ *and centered at* $p$. *Let* $V^k(r)$ *denote the volume of a ball of radius* $r$ *in the space form* $\mathbf{M}_{0,k}$ *with constant sectional curvature* $k$.

*(i)(Bishop and Gromov) If* $Ric \geq (n-1)kg$ *in* $B(p,r)$, *then*

$$|B(p,r)| \leq V^k(r).$$

*Furthermore, let* $J(v,s)$ *and* $J_k(s)$ *be the volume elements of* $\mathbf{M}$ *and* $\mathbf{M}_{0,k}$ *in the geodesic spherical coordinates. Then the ratio* $\frac{J(v,s)}{J_k(s)}$ *is a nonincreasing function of* $s$.

*In addition, the ratio* $\frac{|B(p,r)|}{V^k(r)}$ *is a nonincreasing function of* $r$.

*(ii) (Gunther) If the sectional curvature is less than or equal to* $k$ *in* $B(p,r)$, *which does not intersect with the cut-locus of* $p$, *then*

$$|B(p,r)| \geq V^k(r).$$

Proof. We will use the inverse of the exponential map $exp_p$ as the local chart in Definition 3.5.1 for the volume. This requires us to compute the pull back of the volume form on $\mathbf{M}$ to the tangent space. The tool to do the computation is the Jacobi field. For any $x \in B(p,r)$, let $v$ be a unit tangent vector in $T_p\mathbf{M}$ such that $x = exp_p(av)$ where $a = d(p,x)$. Let $\{e_1, \ldots, e_{n-1}, v\}$ be an orthonormal basis of $T_p\mathbf{M}$ and $\{e_1(s), \ldots, e_{n-1}(s), c'(s)\}$ be an orthonormal basis of $T_{c(s)}\mathbf{M}$ where $e_i(s)$ is the parallel transport of $e_i$ along $c = exp_p(sv)$, $s \in [0,a]$. For $i = 1,,,,n-1$, denote by $Y_i$ the Jacobi field along $c$ such that $Y_i(0) = 0$ and $Y_i'(0) = e_i$. By Proposition 3.5.1

$$|B(p,r)| = \int_{S^{n-1}} \int_0^r \sqrt{\det(g(Y_i(s), Y_j(s))} ds dv.$$

So we need to compute the function

$$J(v,s) = \sqrt{\det(g(Y_i(s), Y_j(s))}.$$

Since $Y_i$ is a Jacobi field,

$$\frac{d^2}{ds^2}g(c'(s),Y_i(s)) = g(c'(s),Y_i''(s)) = -g(c'(s),R(Y_i(s),c'(s))c'(s)) = 0.$$

Also $g(c'(0),Y_i(0)) = 0$ and $\frac{d}{ds}g(c'(s),Y_i(s))|_{s=0} = g(c'(0),Y_i'(0)) = g(c'(0),e_i) = 0$. Hence $Y_i(s)$ is always orthogonal to $c'(s)$ and consequently

$$Y_i(s) = \Sigma_{j=1}^{n-1}a_{ji}(s)e_j(s)$$

for some scalar functions $a_{ij}$. This shows $g(Y_i(s),Y_j(s)) = \Sigma_{k=1}^{n-1}a_{ki}(s)a_{kj}(s)$. Therefore

$$J(v,s) = detA(s)$$

where

$$A(s) \equiv (a_{ij}(s)).$$

Using the fact that $Y_i$ is a Jacobi field again,

$$\begin{aligned}\Sigma_{j=1}^{n-1}a_{ji}''(s)e_j(s) = Y_i''(s) &= -R(Y_i(s),c'(s))c'(s)\\ &= -\Sigma_{k=1}^{n-1}a_{ki}(s)R(e_k(s),c'(s))c'(s).\end{aligned}$$

It follows that

$$a_{ji}''(s) = -\Sigma_{k=1}^{n-1}\rho_{kj}\, a_{ki}(s) \tag{3.5.1}$$

where

$$\rho_{kj} \equiv < R(e_k(s),c'(s))c'(s),e_j(s) > .$$

By linear algebra, it is easy to check that

$$\partial_s J(v,s) = \partial_s detA = detA\, tr(A'\, A^{-1}) = J(v,s)\, tr(A'\, A^{-1})$$

where $A' \equiv \partial_s A$. To facilitate computation, we use the notation

$$B = A'\, A^{-1}.$$

Then $\partial_s J(v,s) = J(v,s)\, trB$. Differentiating this with respect to $s$ again, we have

$$\partial_s^2 J(v,s) = \partial_s J(v,s)\, trB + J(v,s)\, \partial_s trB = J(v,s)\, (trB)^2 + J(v,s)\, \partial_s trB. \tag{3.5.2}$$

Let us compute $\partial_s trB$. Write $A = (a_{ij})$ and $A^{-1} = (n_{ij})$. Then, denoting $\partial_s$ by $'$ and suppressing obvious summation signs, we compute

$$\partial_s trB = [n_{ik}a_{ki}']' = n_{ik}'a_{ki}' + n_{ik}a_{ki}''.$$

By (3.5.1), this becomes

$$\begin{aligned}\partial_s trB &= [n_{ik}a'_{ki}]' = n'_{ik}a'_{ki} - n_{ik}\rho_{lk}a_{li} = n'_{ik}a'_{ki} - \rho_{kk} \\ &= n'_{ik}a'_{ki} - Ric(c'(s), c'(s)).\end{aligned}$$

Here we have used the definition of $\rho_{lk}$ just after (3.5.1). Differentiating the identity $n_{il}a_{lk} = \delta_{ik}$, it is easy to see that

$$n'_{ik}a'_{ki} = -n_{il}a'_{lm}n_{mk}a'_{ki} = -trB^2.$$

Hence

$$\partial_s trB = -trB^2 - Ric(c'(s), c'(s)).$$

Substituting this to (3.5.2), we obtain

$$\partial_s^2 J(v,s) = J(v,s)\,(trB)^2 - J(v,s)\,[trB^2 + Ric(c'(s), c'(s))].$$

Writing $\Phi = J^{1/(n-1)}(v,s)$, then we arrive at the formula

$$\partial_s^2\Phi = \frac{1}{n-1}[\frac{1}{n-1}(trB)^2 - trB^2]\Phi - \frac{1}{n-1}\Phi\, Ric(c'(s), c'(s)).$$

Since $B$ is a $n-1$ by $n-1$ matrix, it holds $\frac{1}{n-1}(trB)^2 - trB^2 \leq 0$. This inequality can easily be verified by triangularization. Therefore

$$\partial_s^2\Phi \leq -\frac{1}{n-1}\Phi\, Ric(c'(s), c'(s)).$$

By the assumption $Ric \geq (n-1)kg$, we reach the inequality

$$\partial_s^2\Phi \leq -k\Phi. \tag{3.5.3}$$

Define a positive function $\Phi_0 : \mathbf{R} \to \mathbf{R}$ as follows.

$$\Phi_0(s) = \begin{cases} sinh(\sqrt{-k}s), & k < 0; \\ s, & k = 0; \\ sin(\sqrt{k}s), & k > 0. \end{cases} \tag{3.5.4}$$

Then the Sturm-Liouville theorem for ordinary differential equations imply that

$$\frac{\Phi(s)}{\Phi_0(s)}$$

is a nonincreasing function of $s$. Actually one can verify the monotonicity directly. Note that $\Phi_0$ satisfies the equality $\Phi_0''(s)+k\Phi_0(s) = 0$. Both $\Phi$ and $\Phi_0$ are positive except at $s = 0$. By (3.5.3),

$$(\Phi'\Phi_0 - \Phi\Phi_0')' = \Phi''\Phi_0 - \Phi\Phi_0'' \leq 0.$$

Also $(\Phi'\Phi_0 - \Phi\Phi_0')(0) = 0$. Hence, for $s \ge 0$, it holds

$$(\Phi'\Phi_0 - \Phi\Phi_0')(s) \le 0. \tag{3.5.5}$$

Therefore, for $s > 0$,

$$\left(\frac{\Phi}{\Phi_0}\right)' = \frac{\Phi'\Phi_0 - \Phi\Phi_0'}{\Phi_0^2} \le 0.$$

Recall that $J(v,s) = \Phi^{n-1}$ is the volume element in the geodesic spherical coordinates for $\mathbf{M}$ and $J_k(s) \equiv \Phi_0^{n-1}(s)$ is that for the space form with sectional curvature $k$. So we have proven that $\frac{J(v,s)}{J_k(s)}$ is a nonincreasing function of $s$.

It is easy to see that

$$\lim_{s\to 0} J(v,s)/J_k(s) = 1.$$

This and the fact that $J(v,s)/J_k(s)$ is a nonincreasing function immediately imply the volume comparison statement in (i), by Proposition 3.5.1.

One can prove the last statement in (i), i.e. the nonincreasing property of

$$|B(p,r)|/V^k(r),$$

by writing

$$\frac{|B(p,r)|}{V^k(r)} = \frac{\int_0^r \int_{S^{n-1}} J(v,s)dvds}{\int_0^r \int_{S^{n-1}} J_k(s)dvds}.$$

Then one differentiates this ratio with respect to $r$ and using the property

$$J'(v,s)J_0(s) - J(v,s)J_0'(s) \le 0$$

which follows from (3.5.5).

*Proof of (ii).* Let $x \in B(p,r)$ be as in part (i). Clearly we just need to prove $|B(p,r)| \ge V^k(r)$ assuming $r$ is less than the diameter of the space form $\mathbf{M}_{0,k}$. Otherwise, if $r$ is larger, then $V^k(r)$ becomes a constant: the volume of $\mathbf{M}_{0,k}$, and $|B(p,r)|$ is not smaller.

By Proposition 3.4.6, for $i = 1,,,,n-1$, let $X_i$ be the Jacobi field along $c$ such that $X_i(0) = 0$ and $X_i(a) = e_i(a)$, then the following identities hold.

$$\begin{aligned}\partial_s|_{s=a} \ln(s^{n-1}\sqrt{detg_e}) &= \Delta r(x)\\ &= \Sigma_{i=1}^{n-1} \int_0^{d(x,p)} (|X_i'|^2 + R(X, X_i, X, X_i)ds\end{aligned}$$

where $X_i'(s) = \nabla_{c'(s)} X_i(s)$ and $X = c'(s)$. Also $detg_e$ is the determinant of the matrix $g(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})$ under a canonical basis in the exponential coordinates and $a = d(x,p)$. Using the sectional curvature upper bound,

$$\partial_s|_{s=a} \ln(s^{n-1}\sqrt{detg_e}) \geq \Sigma_{i=1}^{n-1} \int_0^{d(x,p)} (|X_i'|^2 - k|X_i|^2) ds$$

Let

$$X_i(s) = \Sigma_{j=1}^{n-1} \xi_{ij}(s) e_j(s).$$

Then, since $e_j(s)$ are orthonormal, parallel vector fields, this becomes

$$\partial_s|_{s=a} \ln(s^{n-1}\sqrt{detg_e}) \geq \Sigma_{i=1}^{n-1} \int_0^{d(x,p)} \Sigma_{j=1}^{n-1} (|\xi_{ij}'(s)|^2 - k|\xi_{ij}(s)|^2) ds \tag{3.5.6}$$

Now fix a point $m_0$ in the space form $\mathbf{M}_{0,k}$ equipped with metric $g_0$ and distance $d_0$. Pick a point $m \in \mathbf{M}_{0,k}$ such that $d_0(m_0, m) = a$. Let $c_0 = c_0(s)$ be a minimal geodesic connecting $m_0$ and $m$. Mirroring $e_i(s)$, we define $E_{0,i}(s)$, $i = 1, \ldots, n-1$ as parallel fields along $c_0$ such that $\{E_{0,1}(s), \ldots, E_{0,n-1}(s), c_0'(s)\}$ form an orthonormal basis for $T_{c_0(s)}\mathbf{M}_{0,k}$. Now we define the vector field along $c_0 = c_0(s)$

$$X_{0,i}(s) = \Sigma_{j=1}^{n-1} \xi_{ij}(s) E_{0,j}(s).$$

which can be regarded as the mirror image of $X_i(s)$ on the space form. Then, (3.5.6) becomes

$$\partial_s|_{s=a} \ln(s^{n-1}\sqrt{detg_e}) \geq \Sigma_{i=1}^{n-1} \int_0^{d(x,p)} (|X_{0,i}'(s)|^2 - k|X_{0,i}(s)|^2) ds$$

where the norm is with respect to the metric $g_0$ on $\mathbf{M}_{0,k}$ now.

Recall from the construction of $X_i$ that $X_i(0) = 0$ and $X_i(a) = e_i(a)$. Therefore $\xi_{ij}(0) = 0$ and $\xi_{ij}(a) = \delta_{ij}$. Hence $X_{0,i}(0) = 0$ and $X_{0,i}(a) = E_{0,i}(a)$. Let $J_{0,i}$ be the Jacobi field along $c_0 = c_0(s)$ such that $J_{0,i}(0) = 0$ and $J_{0,i}(a) = E_{0,i}(a)$. Then, clearly

$$J_{0,i}(s) = \frac{\Phi_0(s)}{\Phi_0(a)} E_{0,i}(s)$$

where $\Phi_0$ is defined in (3.5.4). Let us note that the curve $c_0 = c_0(s)$ from $m_0$ to $m$ is a minimal geodesic since $a \leq r$, which is less than the diameter of $\mathbf{M}_{0,k}$. Applying the index Theorem 3.4.2, we deduce

$$\begin{aligned}\partial_s|_{s=a}\ln(s^{n-1}\sqrt{detg_e}) &\geq \Sigma_{i=1}^{n-1}\int_0^{d(x,p)}(|J'_{0,i}(s)|^2 - k|J_{0,i}(s)|^2)ds\\ &= \Sigma_{i=1}^{n-1}\int_0^{d(x,p)}[(\Phi'_0(s)/\Phi_0(a))^2 - k(\Phi_0(s)/\Phi_0(a))^2]ds\\ &= \partial_s|_{s=a}\ln\Phi_0^{n-1}. \quad \text{(computation from (3.5.4))}\end{aligned}$$

This shows, after remembering $x = c(a)$,

$$a^{n-1}\sqrt{detg_e}|_x \geq \Phi_0^{n-1}(a).$$

Therefore

$$\int_0^r\int_{S^{n-1}} a^{n-1}\sqrt{detg_e}(v,a)dvda \geq \int_0^r\int_{S^{n-1}}\Phi_0^{n-1}(a)dvda$$

where $dv$ is again the volume element on $S^{n-1}$. Hence $|B(x,r)| \geq V_r^k$. □

We finish this section by discussing the Rauch comparison theorem for Jacobi fields. As usual, for a Jacobi field $J(t)$ along a geodesic, we use the notations

$$\|J(t)\| \equiv \sqrt{g_{c(t)}(J(t),J(t))}, \quad J'(t) = \nabla_{c'(t)}J(t).$$

**Theorem 3.5.2** *(Rauch) Let* $\mathbf{M}$ *be a complete manifold and* $c : [0,\infty) \to \mathbf{M}$ *a geodesic. Suppose* $J = J(t)$ *be a Jacobi field along* $c$ *satisfying the initial conditions* $J(0) = 0$, $g(J'(0), c'(0)) = 0$ *and* $\|J'(0)\| = 1 = \|c'(0)\|$. *Let sec be the sectional curvature of* $\mathbf{M}$ *and* $K$ *a positive constant. Then the following conclusions are true.*

*(i) If* $sec \leq 0$, *then* $\|J(t)\| \geq t$.

*(ii) If* $sec \leq -K$, *then* $\|J(t)\| \geq \frac{e^{\sqrt{K}t}-e^{-\sqrt{K}t}}{2\sqrt{K}}$.

*(iii) If* $sec \leq K$, *then* $\|J(t)\| \geq \frac{\sin(\sqrt{K}t)}{\sqrt{K}}$.

Proof. We just give a proof of (i), since the proof of (ii) and (iii) are similar. Write $f(t) = \|J(t)\|$. Direct computation using the Jacobi equation shows

$$\begin{aligned}f''(t) &= \left(\frac{g(J,J')}{\|J\|}\right)' = \frac{(g(J,J'))'\|J\| - g(J,J')\|J\|'}{\|J\|^2}\\ &= -\frac{g(R(J,c')c',J)}{\|J\|^2}f(t) + \frac{1}{\|J\|^3}(\|J\|^2\|J'\|^2 - g(J,J')^2)\\ &\geq -\frac{g(R(J,c')c',J)}{\|J\|^2}f(t).\end{aligned}$$

By the assumption on the initial value of $J$ and the Jacobi equation, it is easy to see that $J(t)$ is orthogonal to $c'(t)$. Hence $\frac{g(R(J,c')c',J)}{\|J\|^2}$ is the sectional curvature of the tangent plane spanned by $c'$ and $J$. By the assumption that $sec \le 0$, we obtain

$$f''(t) \ge 0.$$

Note that $f(0) = \sqrt{g(J(0), J(0))} = 0$. Also

$$f'(0) = \lim_{t\to 0^+} \frac{f(t)}{t} = \lim_{t\to 0^+} \sqrt{g(\frac{J(t)}{t}, \frac{J(t)}{t})} = \|J'(0)\| = 1.$$

Hence $f(t) \ge t$. □

**Exercise 3.5.2** *Prove (ii) and (iii) of the theorem.*

Various generalization and refinement of the Rauch comparison theorem exist. They play important roles in Riemann geometry.

## 3.6 More on conjugate points, cut-locus and injectivity radius

In this section we present a number of useful results on conjugate points, cut-locus and lower bounds of injectivity radius. They are not only useful for differential geometry at large but fundamental for the study of Ricci flow in later chapters of the book.

**Proposition 3.6.1** *Let* $\mathbf{M}$ *be a complete Riemann manifold,* $c(s) = exp_p(sv)$, $s > 0$, *be a unit speed geodesic, i.e.* $g_p(v, v) = 1$.

*(1) Suppose* $q = exp_p(s_0 v)$ *is a conjugate point of* $p$. *Then, for any* $\epsilon > 0$, *the curve* $c = c(s), s \in [0, s_0 + \epsilon]$ *is not distance minimizing.*

*(2) Suppose there is no conjugate point along* $c = c(s)$, $s \in [0, s_0]$. *Then for any piecewise smooth curve* $\sigma$ *such that* $\sigma(0) = p$ *and* $\sigma(s_0) = q$, *which is sufficiently close to* $c = c(s)$ *in* $C^0$ *topology, there holds*

$$L(\sigma) \ge L(c).$$

*The equality holds only when* $\sigma$ *and* $c$ *are the same curve.*

*Proof of (1).* Denote $s_0 + \epsilon$ by $s_1$. We will construct a curve joining $p$ and $c(s_1)$, whose length is shorter than $s_1$.

Since $q = c(s_0)$ is conjugate to $p$, by Proposition 3.4.4 there exists a nontrivial Jacobi field $Y = Y(s)$, $s \in [0, s_0]$ along $c$ such that $Y(0) = 0$

and $Y(s_0) = 0$. Since the component of $Y(s)$ that is normal to $c'(s)$ is also a Jacobi field, we can just choose $Y(s)$ to be normal to the geodesic.

Take a parallel vector field $P = P(s)$ along $c$ such that $P(s_0) = -Y'(s_0)$ which is not zero since $Y$ is nontrivial. Let $\theta : [0, s_1] \to [0, 1]$ be a smooth function such that $\theta(0) = \theta(s_1) = 0$ and $\theta(s_0) = 1$. Define, for $\lambda > 0$, a piecewise smooth vector field along $c$,

$$Z = Z(s) = \begin{cases} Y(s) + \lambda\theta(s)P(s), & s \in [0, s_0]; \\ \lambda\theta(s)P(s), & s \in [s_0, s_1]. \end{cases}$$

Finally define a variation of $c = c(s)$,

$$B(s,t) = exp_{c(s)}(tZ(s)).$$

Then $B(s, 0) = c(s)$, $B(0, t) = p$ and $B(s_1, t) = c(s_1)$. Also

$$\partial_t|_{t=0} B(s,t) = Dexp_{c(s)}|_0 Z(s) = Z(s).$$

Denote by $L(t)$ the length of the curve $B(s, t), s \in [0, s_1]$. It is easy to check that $g(P(s), c'(s)) = 0$ and consequently $g(Z(s), c'(s)) = 0$. Differentiating with respect to $s$, we know that $g(Z'(s), c'(s)) = 0$. Here $Z'(s)$ means $\nabla_{c'(s)} Z(s)$. According to the second variation formula in Proposition 3.4.2,

$$\begin{aligned} \frac{d^2L(t)}{dt^2}|_{t=0} = I(Z, Z) &= \int_0^{s_0} [|Y'|^2 + R(Y, c', Y, c')]ds \\ &+ 2\lambda \int_0^{s_1} [g(Y', (\theta P)') + R(Y, c', \theta P, c')]ds + \lambda^2 I(\theta P, \theta P) \\ &\equiv T_1 + T_2 + T_3. \end{aligned}$$

Here $Y(s)$ is regarded as zero when $s \geq s_0$. Since $Y$ is a Jacobi field vanishing at the end points, we know that $T_1 = 0$.

Note that

$$\partial_s g(Y', \theta P) = g(Y'', \theta P) + g(Y', (\theta P)'),$$

$$g(Y'', \theta P) + R(Y, c', c', \theta P) = 0.$$

Using these two identities and integration by parts, we see that

$$T_2 = 2\lambda g(Y'(s), \theta P(s))|_0^{s_0} = -2\lambda|Y'(s_0)|^2.$$

Therefore

$$\frac{d^2L(t)}{dt^2}|_{t=0} = -2\lambda|Y'(s_0)|^2 + \lambda^2 I(\theta P, \theta P) < 0$$

when $\lambda$ is close to 0. This shows that the geodesic $c = c(s)$, $s \in [0, s_1] = [0, s_0 + \epsilon]$ is not distance minimizing.

*Proof of (2).* This is an immediate consequence of Theorem 3.4.3, the basic index theorem. Indeed, we take $Y \equiv 0$ as the Jacobi field in that theorem. Then for any nontrivial vector field $X$ along $c$ vanishing at the ends, we can use Proposition 3.4.2 to deduce

$$\frac{d^2}{dt^2}L(c_t)|_{t=0} = I(X_1, X_1) > 0.$$

Here $c_t = c(s,t) = exp_{c(s)}(tX(s))$ is the variation generated by $X$, and $X_1(s)$ is the component of $X(s)$, which is orthogonal to $c'(s)$. Since $X$ is arbitrary, any small variation of the geodesic $c = c(s)$ can not be distance minimizing □

**Proposition 3.6.2** *Let* $\mathbf{M}$ *be a complete Riemann manifold,* $c(s) = exp_p(sv)$, $s > 0$, *be a unit speed geodesic. Suppose* $q = c(s_0)$ *is a point in the cut-locus of* $p \in \mathbf{M}$. *Then*

*(1) Either* $q$ *is a conjugate point of* $p$, *or*

*(2) There exists another geodesic* $\sigma$ *joining* $p$ *and* $q$ *such that*

$$L(\sigma) = L(c|_{[0,s_0]}) = d(p,q).$$

*(3) Conversely, if either the statement in (1) or (2) holds, then* $q$ *is in the cut-locus of* $p$.

*(4) The injectivity radius at* $p$ *equals the distance between* $p$ *and its cut-locus.*

Proof. We pick a sequence of positive numbers $\epsilon_i \to 0$ when $i \to \infty$. Let $\sigma_i = \sigma_i(s)$, $s \in [0, s_0 + \epsilon_i]$ be a minimum geodesic connecting $p$ and $c(s_0 + \epsilon_i)$, which is parameterized by arclength. Since $q$ is in the cut-locus, by definition, the curve $c = c(s)$, $s \in [0, s_0 + \epsilon_i]$ is not distance minimizing. Hence

$$d(p, c(s_0 + \epsilon_i)) = L(\sigma_i) < s_0 + \epsilon_i.$$

By the triangle inequality, we also have

$$s_0 - \epsilon_i \leq L(\sigma_i).$$

Since $\sigma_i$ is a geodesic, there exists a unit vector $u_i \in T_p\mathbf{M}$ such that $\sigma_i(s) = exp_p(su_i)$. The sequence $\{u_i\}$ is a subset of $S^{n-1}$ in $T_p\mathbf{M}$. Hence there exists a subsequence, still called $\{u_i\}$ that converges to a unit vector $u \in T_p\mathbf{M}$. Consequently $\sigma_i$ converge in $C^0$ topology to the geodesic $\sigma = exp_p(su)$ connecting $p$ and $q$.

If $u \neq v$, then $\sigma$ and $c$ are two different geodesics connecting $p$ and $q$. Also $L(\sigma_i) \to s_0$ since $\epsilon_i \to 0$, by the last paragraph. So we are in case (2).

If $u = v$, then $\sigma$ and $c$ are the same curve. Therefore, for large $i$, $\sigma_i$ is sufficiently close, in $C^0$ topology, to the curve $c = c(s)$, $s \in [0, s_0+\epsilon_i]$. If $q = c(s_0)$ is not a conjugate point of $p$, then by continuity, for large $i$, the points $c(s)$, $s \in [s_0, s_0+\epsilon_i]$, are not conjugate points of $p$ too. Then by item (1) of Proposition 3.6.1, it is not hard to see that the whole curve $c|_{[0,s_0+\epsilon_i]}$ does not contain any conjugate point of $p$. By item (2) of Proposition 3.6.1, we know

$$L(\sigma_i|_{[0,s_0+\epsilon_i]}) \geq L(c|_{[0,s_0+\epsilon_i]}).$$

Since $c|_{[0,s_0+\epsilon_i]}$ is not distance minimizing by assumption, we know that $\sigma_i$ can not be a minimal geodesic connecting $p$ and $c(s_0+\epsilon_i)$. This is a contradiction with the definition of $\sigma_i$. Hence $q$ is a conjugate point of $p$, i.e. (1) holds.

Next we prove (3). Suppose $q$ is a conjugate point of $p$ along a geodesic. Then item (1) of Proposition 3.6.1 shows that $q$ is a cut point.

Now suppose there exists another geodesic $\sigma$ joining $p$ and $q$ such that

$$L(\sigma) = L(c|_{[0,s_0]}) = d(p, q).$$

Let $c = exp_p(sv)$ for some unit vector $v \in T_p\mathbf{M}$. For a small $a > 0$, denote the point $exp((s_0 + a)v)$ by $q_1$. Suppose $q_1 \in \sigma$. As the points $q$ and $q_1$ are very close, they can be joined by only one minimal geodesic. Thus the extended curve $c|_{[0,s_0+a]}$ must overlap $\sigma$. Hence $c \cup \sigma$ is a closed geodesic. So the curve $c : [0, s_0 + a] \to \mathbf{M}$ is not distance minimizing. The reason is that one can reach $q_1$ through $\sigma$ which is shorter than $s_0 + a$. On the other hand, if $q_1$ is not a point on $\sigma$, then the curve $\sigma \cup c|_{[s_0,s_0+a]}$ is a nonsmooth curve connecting $p$ and $q_1$, whose length is $s_0 + a$. Since this curve is nonsmooth, it can not be distance minimizing. Therefore $d(p, q_1) < s_0 + a = L(c|_{[0,s_0+a]})$. This shows that $c : [0, s_0 + a] \to \mathbf{M}$ is still not distance minimizing. Hence $q$ is a cut point, proving (3).

Finally we prove (4). Denote the cut-locus of $p$ by $Cut_p$. If $Cut_p$ is compact, pick $q \in Cut_p$ such that $d(p, q)$ is the distance between $p$ and

$Cut_p$. By (3) of the proposition, for any point $q_1$ which is closer to $p$ than $q$, there exists a unique minimal geodesic connecting $p$ and $q_1$. Hence $exp_p$ is an imbedding on the ball $\{v \in T_p\mathbf{M} \,|\, \|v\| < d(p,q)\}$. Therefore the injectivity radius at $p$ equals $d(p,q)$. If $Cut_p$ is noncompact, one can choose a point $q_1$ outside of the cut-locus such that $d(p,q_1)$ is arbitrarily close to $d(p, Cut_p)$. The conclusion follows in the same way as the compact case. □

Case (1) and (2) in the proposition may not be mutually exclusive. The two geodesics in part (2) of the proposition may not form a smooth, closed geodesic in general. However, in the special case that $q$ is a point in the cut-locus of $p$, which is closest to $p$, we can form a smooth, closed geodesic connecting $p$ and $q$ if they are not conjugate points with each other.

**Proposition 3.6.3** *(injectivity radius and closed geodesic) Let* $\mathbf{M}$ *be a complete Riemann manifold. For* $p \in \mathbf{M}$*, suppose* $q$ *is a point in the cut-locus of* $p$*, which is closest to* $p$*. Then*

*(1) Either* $q$ *is a conjugate point of* $p$ *along a minimal geodesic connecting* $p$ *and* $q$*, or*

*(2) there exists a closed geodesic connecting* $p$ *and* $q$ *whose length is twice of* $d(p,q)$*.*

Proof. Suppose (1) does not hold, i.e. $q$ is not a conjugate point of $p$ along any minimal geodesic. By the previous proposition, as $q$ is in the cut-locus of $p$, there exist two geodesics $c$ and $\sigma$ joining $p$ and $q$ such that $L(\sigma) = L(c) = d(p,q)$. We need to show that $c \cup \sigma$ form a closed geodesic. By scaling the metric we can assume that $d(p,q) = 1$. We parameterize both $c$ and $\sigma$ arclength so that they are functions from $[0,1]$ to $\mathbf{M}$. $c(0) = \sigma(0) = p$ and $c(1) = \sigma(1) = q$. To prove the smoothness of $c \cup \sigma$ at $q$, we only need to show that $c'(1) = -\sigma'(1)$.

We assume that $c'(1) \neq -\sigma'(1)$, then there exists a vector $w \in T_q\mathbf{M}$ such that $g(w, c'(1)) < 0$ and $g(w, \sigma'(1)) < 0$. For a small $t > 0$, denote $q(t) = exp_q(tw)$. We claim that there exist two geodesics $c_t : [0,1] \to \mathbf{M}$ and $\sigma_t : [0,1] \to \mathbf{M}$, connecting $p$ and $q(t)$. This claim is the result of inverse function theorem. Indeed, since $q$ is not a conjugate point of $p$ and $exp_p c'(0) = q$, $Dexp_p|_{c'(0)}$ is nonsingular. By the inverse function theorem, there exist $v_1 = v_1(t) \in T_p\mathbf{M}$, in a small neighborhood of $c'(0)$, such that $exp_p(v_1) = q(t)$. Therefore we can regard $exp_p(sv_1), s \in [0,1]$ as the geodesic $c_t = c_t(s)$. The other geodesic $\sigma_t$ is found similarly by replacing $c'(0)$ with $\sigma'(0)$ in the above argument.

Thus the function $c = c(s,t) \equiv c_t(s)$, $s \in [0,1]$, is a smooth variation of the geodesic $c = c(s)$ such that $c(s,0) = c(s)$. We compute, in the usual way as in the first variation formula for geodesics

$$\begin{aligned} \frac{d}{dt}L(c(\cdot,t)) &= \frac{1}{2}\int_0^1 \frac{\partial_t g(\partial_s c(s,t), \partial_s c(s,t))}{\|\partial_s c(s,t)\|} ds \\ &= \int_0^1 \frac{g(\nabla_t \nabla_s c(s,t), \partial_s c(s,t))}{\|\partial_s c(s,t)\|} ds \\ &= \int_0^1 \frac{g(\nabla_s \nabla_t c(s,t), \partial_s c(s,t))}{\|\partial_s c(s,t)\|} ds \\ &= \int_0^1 \frac{\partial_s g(\partial_t c(s,t), \partial_s c(s,t))}{\|\partial_s c(s,t)\|} ds. \end{aligned}$$

Here $\nabla_s = \nabla_{\partial_s c(s,t)}$ and $\nabla_t = \nabla_{\partial_t c(s,t)}$ with $\partial_s c(s,t)$ and $\partial_t c(s,t)$ being recognized as tangent vectors of $c(\cdot,t)$ and $c(s,\cdot)$ respectively. Also $\nabla_s c(s,t) = \partial_s c(s,t)$.

When $t = 0$, we have $\|\partial_s c(s,t)\| = 1$ since $c = c(s) = c(s,0)$ is parameterized by arclength. Notice also $\partial_t c(s,t)|_{s=0,t=0} = 0$ since $c(0,t)$ is a fixed point. Consequently

$$\frac{d}{dt}L(c(\cdot,t))|_{t=0} = g(\partial_t c(s,t), \partial_s c(s,t))|_{s=1,t=0}.$$

Noting that $c(1,t) = q(t) = exp_q(tw)$, we arrive at

$$\frac{d}{dt}L(c(\cdot,t))|_{t=0} = g(w, c'(1)) < 0.$$

In the same manner

$$\frac{d}{dt}L(\sigma(\cdot,t))|_{t=0} = g(w, \sigma'(1)) < 0.$$

Hence for sufficiently small $t > 0$, the lengths of $c_t$ and $\sigma_t$ are strictly less than 1. Recall that $c_t(1) = \sigma_t(1) = q(t)$. Since $d(p, Cut_p) = 1$ by assumption, we know that $q(t)$ is not a cut point of $p$.

If $L(c_t) = L(\sigma_t) = d(p, q(t))$, by Proposition 3.6.2 (3), $q(t)$ must be a cut point. This is a contradiction. If either $L(c_t)$ or $L(\sigma_t)$ is strictly greater than $d(p, q(t))$, then the longer geodesic is no longer distance minimizing. But its length is less than 1, contradicting with the fact that $d(p, Cut_p) = d(p,q) = 1$. This contradiction proves that $c \cup \sigma$ is a smooth, closed geodesic. □

The following theorem, due to Klingenberg, provides a lower bound for the injectitivity radius of a manifold with sectional curvature bounded from above.

Let $\mathbf{M}$ be a Riemann manifold with metric $g$, we introduce the notation

$$l(\mathbf{M}, g) = \inf\{L(\sigma) \,|\, \sigma : S^1 \to \mathbf{M}, \sigma \,\text{is a geodesic}, \sigma'(0) \neq 0\}$$

which is nothing but the lower bound of the lengths of closed geodesics on $\mathbf{M}$.

**Theorem 3.6.1** *(Klingenberg) Suppose* $\mathbf{M}$ *is a compact Riemann manifold whose sectional curvature is bounded from above by a positive constant* $K_0$*. Then the injectivity radius has the lower bound*

$$Inj(\mathbf{M}, g) \geq \min\{\frac{\pi}{\sqrt{K_0}}, \frac{1}{2} l(\mathbf{M}, g)\}.$$

PROOF. Since $\mathbf{M}$ is compact, we can find a point $p \in \mathbf{M}$ such that the injectivity radius at $p$ equals the injectivity radius of the whole manifold. Let $q$ be a point in the cut-locus of $p$, which is closest to $p$. According to Proposition 3.6.2, $inj_p$, the injectivity radius at $p$ equals $d(p, q)$.

According to Proposition 3.6.3, there are two possibilities. One is that $p$ and $q$ are joined by a closed geodesic whose length is twice of $d(p, q)$. Whence $inj_p \geq \frac{1}{2} l(\mathbf{M}, g)$. Two is that $p$ and $q$ are conjugate points along a minimal geodesic $\sigma$. Let $\sigma = \sigma(t)$ be parameterized by arclength. By Rauch comparison theorem (Theorem 3.5.2), the sectional upper bound and Proposition 3.4.4, a conjugate point of $p$ can not occur for $t \in [0, \frac{\pi}{\sqrt{K_0}})$. Hence $inj_p = d(p, q) \geq \frac{\pi}{\sqrt{K_0}}$. □

Based on this theorem, one can prove the following local lower bound for the injectivity radius which is utilized very often in the study of Ricci flow.

**Theorem 3.6.2** *(Cheeger-Gromov-Taylor [CGT] and Cheng-Li-Yau [CLY])*

*Let* $B(x_0, 4r_0)$, $r_0 \in (0, \infty)$, *be a geodesic ball in a* $n$ *dimensional complete manifold* $(\mathbf{M}, g)$*. Suppose the sectional curvature sec in* $B(x_0, 4r_0)$ *satisfies*

$$\lambda \leq sec \leq \Lambda$$

*for two constants* $\lambda$ *and* $\Lambda$*. Then for any positive constant* $r$ *satisfying*

$$r \leq \min(r_0, \frac{\pi}{4\sqrt{\max(\Lambda, 0)}}),$$

*there holds*

$$inj(\mathbf{M}, x_0) \geq r \frac{|B(x_0, r)|}{|B(x_0, r)| + V_\lambda^n(2r)}$$

*where $V_\lambda^n(2r)$ is the volume of a geodesic ball of radius $2r$ in the $n$ dimensional simply connected space form with constant sectional curvature $\lambda$.*

Proof. (sketch) The proof goes as the one for Theorem 4.2.2 in [CZ]. By the above Klingenberg's theorem, it suffices to prove

$$l \equiv l_M(x_0) \geq 2r \frac{|B(x_0, r)|}{|B(x_0, r)| + V_\lambda^n(2r)}.$$

Here $l_M(x_0)$ is the length of the shortest geodesic loop passing $x_0$. Pick a number $r$ as in the statement of the theorem such that $r \geq l$. Consider the ball $B(x_0, 4r)$ and the exponential map:

$$exp_{x_0} : \tilde{B}(0, 4r)(\subset T_{x_0}M) \to B(x_0, 4r).$$

By equipping $\tilde{B}(0, 4r)$ with the pull back metric $\tilde{g} = exp_{x_0}^* g$, we know that $exp_{x_0}$ is a local isometry from $\tilde{B}(0, 4r)$ to $B(x_0, 4r)$. Let $p_1 = 0, p_2, \ldots, p_k$ be the pre-images of $x_0$ in $\tilde{B}(0, r)$. Using the Whitehead lemma below, we know that $k - 1 \geq 2[r/l]$ and that the intersections of the balls $\tilde{B}(p_i, r)$ have zero measure. By the classical volume comparison theorem, we have

$$k|B(x_0, r)| \leq | \cup_{i=1}^k \tilde{B}(p_i, r)| \leq |\tilde{B}(0, 2r)| \leq V_\lambda^n(2r).$$

Hence

$$l \geq 2r/(k+1) \geq 2r \frac{|B(x_0, r)|}{|B(x_0, r)| + V_\lambda^n(2r)}.$$

□

**Exercise 3.6.1** *Fill in the details of the proof of the theorem.*

Here we present Whitehead's lemma which was used in the proof of the above theorem.

**Lemma 3.6.1** *(Whitehead) Let $(\mathbf{M}, g)$ be a complete Riemann manifold with sectional curvature bounded from above by $1$. Suppose $x \in M$ and $0 < r < \frac{1}{2} min\{\pi, inj(M)\}$. Then*

*(a). the ball $B(x, r)$ is geodesic convex, i.e. for any $y, z \in B(x, r)$ there exists a unique minimum geodesic connecting $z$ and $y$, which lies in $B(x, r)$.*

*(b). for any* $\{x_1, \ldots, x_k\} \subset B(x,r)$*, there exists a unique center of mass* $y \in B(x,r)$*, i.e.*

$$\Sigma_{i=1}^{k} exp_y^{-1} x_i = 0.$$

For a proof of the lemma, see p103 [CE] e.g.

## 3.7 Bochner-Weitzenbock type formulas

Bochner-Weitzenbock type formulas are extensions of the Ricci identity in Proposition 3.2.1. They provide a link between various kinds of differential operators on manifolds. In this section we discuss a few samples.

**Proposition 3.7.1** *Bochner's formula.*

*Let* $u$ *be a smooth function on a manifold* $(\mathbf{M}, g)$*, then*

$$\frac{1}{2}\Delta|\nabla u|^2 = |\nabla\nabla u|^2 + <\nabla\Delta u, \nabla u> + Ric(\nabla u, \nabla u).$$

Proof. It is convenient to do it in local orthonormal system. We use $u_i, u_{ij}, u_{ijk}$ etc. to denote first, second and third covariant derivatives of $u$. Recall $du = u_i dx^i$ with $u_i = \frac{\partial u}{\partial x^i}$ and $\nabla u = g^{ij}\frac{\partial u}{\partial x^j}\frac{\partial}{\partial x^i} = u^i \frac{\partial}{\partial x^i}$ where $u^i = g^{ij}u_j$. Hence

$$|\nabla u|^2 = g_{ij}u^i u^j = g^{ij}u_i u_j = u_i^2.$$

Now we compute

$$\begin{aligned}
\frac{1}{2}\Delta|\nabla u|^2 &= \frac{1}{2}(u_i^2)_{jj} \\
&= (u_i u_{ij})_j = (u_i u_{ji})_j = u_{ij}u_{ij} + u_i u_{jij} \quad (\text{since} \quad u_{ij} = u_{ji}) \\
&= u_{ij}u_{ij} + u_i(u_{jji} - R^s_{jij}u_s) \\
&= u_{ij}u_{ij} + u_i u_{jji} - R^s_{jij}u_i u_s \\
&= u_{ij}u_{ij} + u_i u_{jji} - R^s_{jij}g_{il}u^l g_{sk}u^k \\
&= u_{ij}u_{ij} + u_i u_{jji} - R_{jijk}g_{il}u^l u^k \\
&= u_{ij}u_{ij} + u_i u_{jji} + R_{ljjk}u^l u^k \\
&= |\nabla\nabla u|^2 + <\nabla\Delta u, \nabla u> + Ric(\nabla u, \nabla u).
\end{aligned}$$

Here, when going from the second to the third line, we have used the Ricci identity in Proposition 3.2.1, applied to the $(1,0)$ tensor $u_j$. □

The Laplace-Beltrami operator on scalar functions can be generalized to act on tensor. However there are at least two ways to do the generalization, resulting in different Laplacians.

**Definition 3.7.1** *(Laplace operators acting on tensors)*

*(i) Rough Laplacian on tensors. Let $T$ be a smooth $(p, q)$ tensor field on $\mathbf{M}$, then the (rough) Laplacian is the second order operator defined by*

$$\Delta T = div\nabla T = trace_g \nabla^2 T = \Sigma_{i=1}^n \nabla^2_{e_i, e_i} T$$

*where $\{e_1, \ldots, e_n\}$ is a orthonormal frame.*

*(ii) Hodge Laplacian on forms. Let $T$ be a smooth $p$ form, i.e. an antisymmetric $(p, 0)$ tensor field on $\mathbf{M}$, then the Hodge Laplacian is the second order operator defined by*

$$\Delta_d T = -(d\delta + \delta d)T$$

*where $\delta$ is the adjoint of the exterior derivative defined in the definition below.*

**Remark 3.7.1** *Some explanation is needed on the equality*

$$trace_g \nabla^2 T = \Sigma_{i=1}^n \nabla^2_{e_i, e_i} T$$

*in the above definition. From Definition 3.2.1, we know that $\nabla^2 T$ is a $(p+2, q)$ tensor. In a local orthonormal system we can write $\nabla^2 T(X_1, \ldots, X_p, \eta_1, \ldots, \eta_q) = T_{ij} dx^i \otimes dx^j$. Here $X_l, \eta_m$ are vector fields and 1 forms respectively. Therefore, since $e_i$ is the dual of $dx^i$, we have*

$$trace_g \nabla^2 T = g^{ij} T_{ij} = \Sigma_{i=1}^n T_{ii} = \Sigma_{i=1}^n \nabla^2 T(e_i, e_i) = \Sigma_{i=1}^n \nabla^2_{e_i, e_i} T.$$

**Definition 3.7.2** *For two $p$ forms $\alpha = \alpha_{i_1 \ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{x_p}$ and $\beta = \beta_{j_1 \ldots j_p} dx^{j_1} \wedge \ldots \wedge dx^{j_p}$, the inner product is the scalar*

$$< \alpha, \beta >= p! g^{i_1 j_1} \ldots g^{i_p j_p} \alpha_{i_1 \ldots i_p} \beta_{j_1 \ldots j_p}.$$

*The $L^2$ inner product is defined by*

$$< \alpha, \beta >_{L^2}= p! \int_{\mathbf{M}} g^{i_1 j_1} \ldots g^{i_p j_p} \alpha_{i_1 \ldots i_p} \beta_{j_1 \ldots j_p} d\mu = \int_{\mathbf{M}} < \alpha, \beta > d\mu.$$

*The operator $\delta$ is the adjoint of the exterior derivative under the $L^2$ inner product, i.e.*

$$< d\eta, \alpha >_{L^2}=< \eta, \delta\alpha >_{L^2}$$

*for any $p$ form $\alpha$ and $p-1$ form $\eta$.*

**Remark 3.7.2** *It is easy to check that the $\delta$ operator has two equivalent forms.*

*1. $\delta\alpha = -p$ div $\alpha$, i.e.*

$$(\delta\alpha)_{i_1\ldots i_{p-1}} = -pg^{jk}\nabla_j\alpha_{ki_1\ldots i_{p-1}};$$

*2. $\delta\alpha = (-1)^{np+n+1} \star d \star \alpha$. Here $\star$ is the Hodge star operator mapping $\Lambda^p T^*(\mathbf{M})$ to $\Lambda^{n-p}T^*(\mathbf{M})$ defined by*

$$\alpha \wedge \star\eta = < \alpha, \eta > d\mu.$$

*The rough Laplacian and the Hodge Laplacian on scalar functions are the same as the Laplace Beltrami operator.*

The relation between the rough Laplacian and the Hodge Laplacian on one forms is given by

**Proposition 3.7.2** *Let $\alpha$ be a smooth one form on $\mathbf{M}$ and $Ric(\alpha)$ be the smooth one form defined by $Ric(\alpha)(X) = Ric(\alpha^*, X)$ for all smooth vector fields $X$. Here $\alpha^*$ is the vector field defined by $g(\alpha^*, X) = \alpha(X)$ for all smooth vector fields $X$. Then*

$$\Delta_d\alpha = \Delta\alpha - Ric(\alpha).$$

Proof. We will give a proof from scratch, using computations in a local orthonormal system $\{x^1, \ldots, x^n\}$ centered at a point $p$.

Let $\alpha = a_i dx^i$ be a one form in a neighborhood of $p$. Then by Proposition 3.1.1

$$\nabla\alpha = a_{i,k}dx^k \otimes dx^i \tag{3.7.1}$$

where $a_{i,k} = \frac{\partial a_i}{\partial x^k} - \Gamma^l_{ik}a_l$. According to the previous remark

$$d\delta\alpha = -d\ trace\nabla\alpha = -d(g^{ik}a_{i,k}) = -\frac{\partial g^{ik}}{\partial x^l}a_{i,k}dx^l - g^{ik}\frac{\partial a_{i,k}}{\partial x^l}dx^l.$$

Since $\frac{\partial g^{ik}}{\partial x^l} = 0$ at the center $p$, we have, at $p$,

$$d\delta\alpha = -g^{ik}\frac{\partial a_{i,k}}{\partial x^l}dx^l. \tag{3.7.2}$$

We need to find a relation between this and the second covariant derivative of $\alpha$.

By Definition 3.2.1, for any smooth vector fields $X$ and $Y$,

$$\nabla^2_{X,Y}\alpha = \nabla_X(\nabla_Y\alpha) - \nabla_{\nabla_X Y}\alpha.$$

We take $X = \frac{\partial}{\partial x^k}$ and $Y = \frac{\partial}{\partial x^l}$. Since $\nabla_X Y = 0$ at $p$, we have

$$\nabla^2_{X,Y}\alpha = \nabla_X(\nabla_Y \alpha).$$

By (3.7.1),

$$\nabla_Y \alpha = \left(\frac{\partial a_i}{\partial x^l} - \Gamma^m_{il} a_m\right) dx^i.$$

Applying (3.7.1) again on the last identity, we obtain, at $p$,

$$\nabla_X(\nabla_Y \alpha) = \frac{\partial}{\partial x^k}\left(\frac{\partial a_i}{\partial x^l} - \Gamma^m_{il} a_m\right) dx^i = \frac{\partial a_{i,l}}{\partial x^k} dx^i$$

This means, at $p$

$$\nabla^2 \alpha = \frac{\partial a_{i,l}}{\partial x^k} dx^k \otimes dx^l \otimes dx^i.$$

Since $\nabla^2\alpha = a_{i,lk} dx^k \otimes dx^l \otimes dx^i$ by definition, we deduce

$$\frac{\partial a_{i,l}}{\partial x^k} = a_{i,lk}. \tag{3.7.3}$$

Substituting this identity to (3.7.2), we reach the formula

$$d\delta\alpha = -g^{ik} a_{i,kl} dx^l. \tag{3.7.4}$$

Next we compute $\delta d\alpha$. Note that

$$\begin{aligned} d\alpha &= \frac{\partial a_i}{\partial x^l} dx^l \wedge dx^i = \frac{\partial a_i}{\partial x^l}(dx^l \otimes dx^i - dx^i \otimes dx^l) \\ &= \left[\frac{\partial a_i}{\partial x^l} - \Gamma^m_{il} a_m\right](dx^l \otimes dx^i - dx^i \otimes dx^l). \end{aligned}$$

Here, in reaching the last identity, we have used the symmetry $\Gamma^m_{ij} = \Gamma^m_{ji}$. Hence

$$d\alpha = a_{i,l}(dx^l \otimes dx^i - dx^i \otimes dx^l).$$

Applying the local derivative formula (3.1.4) to the $(2,0)$ tensor $d\alpha$, we see that, at $p$,

$$\nabla d\alpha = \frac{\partial a_{i,l}}{\partial x^k} dx^k \otimes dx^l \otimes dx^i - \frac{\partial a_{i,l}}{\partial x^k} dx^k \otimes dx^i \otimes dx^l.$$

Here we have used the fact that all terms involving the Christoffel symbols are zero at $p$. Therefore

$$\delta d\alpha = -trace\ \nabla d\alpha = -g^{kl}\frac{\partial a_{i,l}}{\partial x^k} dx^i + g^{ki}\frac{\partial a_{i,l}}{\partial x^k} dx^l.$$

Switching the indices $i$ and $l$ in the second last term, and using (3.7.3), we deduce

$$\begin{aligned}\delta d\alpha &= -g^{ki}\frac{\partial a_{l,i}}{\partial x^k}dx^l + g^{ki}\frac{\partial a_{i,l}}{\partial x^k}dx^l \\ &= -g^{ki}a_{l,ik}dx^l + g^{ki}a_{i,lk}dx^l.\end{aligned} \tag{3.7.5}$$

Adding this to (3.7.4), we obtain

$$\begin{aligned}d\delta\alpha + \delta d\alpha &= -g^{ik}a_{i,kl}dx^l - g^{ki}a_{l,ik}dx^l + g^{ki}a_{i,lk}dx^l \\ &= -g^{ik}(a_{i,kl} - a_{i,lk})dx^l - g^{ki}a_{l,ik}dx^l.\end{aligned}$$

By the Ricci identity Proposition 3.2.1

$$\begin{aligned}\Delta_d\alpha = g^{ik}(a_{i,kl} - a_{i,lk})dx^l + g^{ki}a_{l,ik}dx^l &= -g^{ik}R^m_{lki}a_m dx^l + \Delta\alpha \\ &= -g^{ik}R^m_{lki}g_{rm}a^r dx^l + \Delta\alpha \qquad (\text{where} \quad a^r = g^{rm}a_m) \\ &= -g^{ik}R_{lkir}a^r dx^l + \Delta\alpha \\ &= \Delta\alpha - R_{lr}a^r dx^l = \Delta\alpha - Ric(\alpha).\end{aligned}$$

□

# Chapter 4

# Sobolev inequalities on manifolds and some consequences

In this chapter, we present Sobolev, log Sobolev, parabolic Harnack and some related inequalities on a Riemann manifold. Here are some notations to be used in this chapter. We use $M$ or $\mathbf{M}$ to denote a compact Riemann manifold with a metric $g$, unless stated otherwise; $d(x, y)$, $d\mu$ will denote the distance and volume element respectively; $B(x, r)$, $|B(x, r)|$ mean the geodesic ball centered at $x$ with radius $r$ and its volume respectively; $\nabla$ stands for the gradient of a function, relative to the metric $g$.

## 4.1 A basic Sobolev inequality

Let us start with the following basic Sobolev inequality on compact manifolds as stated in [Heb2].

**Theorem 4.1.1** *Let $(M, g)$ be a smooth, compact Riemann $n$-manifold. For any $p \in [1, n)$, $W^{1,p}(M) \subset L^{np/(n-p)}(M)$, i.e. there exists $A = A(M) > 0$, such that*

$$\left(\int_M u^{np/(n-p)} d\mu\right)^{(n-p)/np} \leq A \left(\int_M |\nabla u|^p d\mu\right)^{1/p} + A \left(\int_M |u|^p d\mu\right)^{1/p}$$

*for all $u \in W^{1,p}(M)$.*

PROOF. By standard approximation, we just need to prove the theorem for smooth functions. First we prove the theorem for $p = 1$ case.

Since $M$ is compact, it can be covered by a finite number of charts $(\Omega_m, \phi_m)$, $m = 1, \ldots, N$ such that for any $m$ the local form of metric $(g_{ij}^{(m)})$ in $(\Omega_m, \phi_m)$ satisfies

$$\frac{1}{2}(\delta_{ij}) \leq (g_{ij}^{(m)}) \leq 2(\delta_{ij}).$$

Let $\{\eta_m\}$ be a smooth partition of unity to the local charts above, i.e. $\eta_m \in C_0^\infty(\Omega_m)$ and $\Sigma_{m=1}^N \eta_m = 1$. In each of the local chart $\Omega_m$, the volume element $d\mu$ can be written as $d\mu = \sqrt{det(g_{ij}^{(m)})}\, dx$ where $dx$ is the volume element in $\mathbf{R}^n$.

For any $u \in C^\infty(M)$ and $m = 1, 2, \ldots, N$,

$$\begin{aligned}&\left(\int |\eta_m u|^{n/(n-1)} d\mu\right)^{(n-1)/n} \\ &\quad \leq \left(2^{n/2} \int_{\mathbf{R^n}} |(\eta_m u) \circ \phi_m^{-1}(x)|^{n/(n-1)} dx\right)^{(n-1)/n}.\end{aligned}$$

By the Sobolev inequality in the Euclidean case (Theorem 2.2.1), we have

$$\left(\int |\eta_m u|^{n/(n-1)} d\mu\right)^{(n-1)/n} \leq c_n \int_{\mathbf{R^n}} |\nabla_e((\eta_m u) \circ \phi_m^{-1}(x))| dx. \quad (4.1.1)$$

Here $\nabla_e$ stands for the Euclidean gradient.

In local coordinates,

$$\begin{aligned}|\nabla(\eta_m u)|^2 &= g_{ij} g^{ik} \partial_k(\eta_m u \circ \phi_m^{-1}) g^{jl} \partial_l(\eta_m u \circ \phi_m^{-1}) \\ &= g^{kl} \partial_k(\eta_m u \circ \phi_m^{-1}) \partial_l(\eta_m u \circ \phi_m^{-1}) \\ &\geq \frac{1}{2} |\nabla_e(\eta_m u \circ \phi_m^{-1})|^2.\end{aligned}$$

The last step is by the assumption on the size of $g_{ij}$ in each coordinate chart. Substituting the above to the right-hand side of (4.1.1) we deduce

$$\left(\int |\eta_m u|^{n/(n-1)} d\mu\right)^{(n-1)/n} \leq c c_n \int |\nabla(\eta_m u)| d\mu. \quad (4.1.2)$$

Summing up (4.1.2) for all $m$ and using Minkowski inequality, we obtain

$$\begin{aligned}\left(\int |u|^{n/(n-1)} d\mu\right)^{(n-1)/n} &\leq \Sigma_{m=1}^N \left(\int |\eta_m u|^{n/(n-1)} d\mu\right)^{(n-1)/n} \\ &\leq cc_n \Sigma_{m=1}^N \int |\nabla(\eta_m u)| d\mu \\ &= cc_n \int |\nabla u| d\mu + cc_n \max \Sigma_{m=1}^N |\nabla \eta_m| \int |u| d\mu.\end{aligned}$$

This proves the theorem when $p = 1$.

For $p \in (1, n)$, we just apply the $p = 1$ case on the function $|u|^{p(n-1)/n}$. □

The following theorem says that the Sobolev imbedding implies the noncollapsing property. It is due Akutagawa [Ak] and Carron [Ca] (see Lemma 2.2 in [Heb2]).

**Theorem 4.1.2** *Let $(M, g)$ be a smooth, complete Riemann $n$-manifold. Suppose that $W^{1,p}(M) \hookrightarrow L^q(M)$, $p \in [1, n)$, $q = \frac{np}{n-p}$, i.e. there exists some constant $A$ such that for any $u \in W^{1,p}(M)$,*

$$\|u\|_q \leq A(\|\nabla u\|_p + \|u\|_p).$$

*Then,*

$$|B(x, r)| \geq \min\{\frac{1}{2A}, \frac{r}{2^{\frac{n+2p}{p}} A}\}^n.$$

Proof. Pick $u \in W^{1,p}(M)$ such that $u = 0$ in $M - B(x, r)$. By Hölder inequality,

$$\begin{aligned} & \|u\|_p \leq |B(x,r)|^{\frac{1}{n}} \|u\|_q \leq |B(x,r)|^{\frac{1}{n}} A(\|\nabla u\|_p + \|u\|_p) \\ \Rightarrow \quad & \|u\|_p - |B(x,r)|^{\frac{1}{n}} A \|u\|_p \leq |B(x,r)|^{\frac{1}{n}} A \|\nabla u\|_p \\ \Rightarrow \quad & 1 - |B(x,r)|^{\frac{1}{n}} A \leq |B(x,r)|^{\frac{1}{n}} A \frac{\|\nabla u\|_p}{\|u\|_p} \\ \Rightarrow \quad & \frac{1}{|B(x,r)|^{\frac{1}{n}}} - A \leq A \frac{\|\nabla u\|_p}{\|u\|_p}. \end{aligned}$$

Case $(i)$, if $|B(x, r)|^{\frac{1}{n}} \geq \frac{1}{2A}$, then we are done.
Case $(ii)$, if $|B(x, r)|^{\frac{1}{n}} \leq \frac{1}{2A}$, then we have $A \leq \frac{1}{2|B(x,r)|^{\frac{1}{n}}}$. Plugging into the above inequality, we deduce

$$\frac{1}{2|B(x,r)|^{\frac{1}{n}}} \leq A \frac{\|\nabla u\|_p}{\|u\|_p} \tag{4.1.3}$$

For any fixed point $x \in M$ such that $d(y) = d(y, x)$ is differentiable at $y$, take

$$u(y) = \begin{cases} r - d(y,x), & \text{if } d(x,y) \le r; \\ 0, & \text{otherwise.} \end{cases}$$

Since $u$ is Lipschitz and $|\nabla u| = |\nabla d| = 1, a.e.$, the following inequality holds:

$$\|\nabla u\|_p \le \left( \int_{B(x,r)} 1 \, d\mu \right)^{\frac{1}{p}} = |B(x,r)|^{\frac{1}{p}}$$

$$\|u\|_p = \left( \int_{B(x,r)} |u|^p \, d\mu \right)^{\frac{1}{p}} \ge \left( \int_{B(x,r/2)} |u|^p \, d\mu \right)^{\frac{1}{p}}$$

Notice for $\forall y \in B(x, r/2)$, $d(x,y) \le \frac{r}{2}$, $u(y) = r - d(x,y) \ge \frac{r}{2}$, then,

$$\|u\|_p \ge \left( \int_{B(x,r/2)} (\frac{r}{2})^p \, d\mu \right)^{\frac{1}{p}} = \frac{r}{2} \cdot |B(x, \frac{r}{2})|^{\frac{1}{p}}.$$

Inserting the above two estimates into (4.1.3), for $\forall r > 0$ such that $B(x,r) \subseteq M$, $B(x,r) \ne M$,

$$\frac{1}{2|B(x,r)|^{\frac{1}{n}}} \le A \frac{|B(x,r)|^{\frac{1}{p}}}{\frac{r}{2} \cdot |B(x,r/2)|^{\frac{1}{p}}} \Rightarrow |B(x,r)| \ge (\frac{r}{4A})^{\frac{np}{n+p}} \cdot |B(x,r/2)|^{\frac{n}{n+p}}.$$

Now we use induction. For a fixed $R > 0$ such that $B(x,R) \subseteq M$, $B(x,R) \ne M$,

$$\begin{aligned} & |B(x,R)| \ge (\frac{R}{4A})^{\frac{np}{n+p}} \cdot |B(x,R/2)|^{\frac{n}{n+p}}, \\ & |B(x,R/2)| \ge (\frac{R/2}{4A})^{\frac{np}{n+p}} \cdot |B(x,R/2^2)|^{\frac{n}{n+p}}, \\ & \cdots\cdots \\ \Rightarrow \quad & |B(x,R)| \ge (\frac{R}{2A})^{p\alpha(m)} \cdot (\frac{1}{2})^{p\beta(m)} \cdot |B(x, \frac{R}{2^m})|^{\gamma(m)}, \text{ where} \end{aligned} \tag{4.1.4}$$

$$\begin{cases} \alpha(m) = \sum_{i=1}^m \left(\frac{n}{n+p}\right)^i \to \frac{\frac{n}{n+p}}{1-\frac{n}{n+p}} = \frac{n}{p} \\ \beta(m) = \sum_{i=1}^m i \cdot \left(\frac{n}{n+p}\right)^i \to \frac{\frac{n}{n+p}}{(1-\frac{n}{n+p})^2} = \frac{n(n+p)}{p^2} \\ \gamma(m) = \left(\frac{n}{n+p}\right)^m. \end{cases}$$

When $m$ is sufficiently large, $|B(x, \frac{R}{2^m})|$ is comparable to the Euclidean case. In fact (see, [GHL] e.g.), let $R_g$ be the scalar curvature, then

$$
\begin{aligned}
& |B(x,r)|_g = \omega_n r^n \left(1 - \frac{R_g(x)}{6(n+2)} r + o(r^2)\right) \geq \frac{\omega_n r^n}{2}. \\
\Rightarrow\quad & |B(x, \frac{R}{2^m})| \geq \frac{\omega_n}{2} (\frac{R}{2^m})^n \\
\Rightarrow\quad & |B(x, \frac{R}{2^m})|^{\gamma(m)} \geq \left(\frac{\omega_n}{2} \cdot 2^{-mn} \cdot R^n\right)^{(\frac{n}{n+p})^m} \\
\Rightarrow\quad & = (\frac{\omega_n}{2})^{(\frac{n}{n+p})^m} \cdot 2^{-mn \cdot (\frac{n}{n+p})^m} \cdot R^{n \cdot (\frac{n}{n+p})^m} \to 1 \text{ as } m \to \infty.
\end{aligned}
$$

Also notice $p\,\alpha(m) \to n$ and $p\,\beta(m) \to \frac{n^2+np}{p}$ as $m \to \infty$. Plugging into (4.1.4), we obtain

$$
|B(x,R)| \geq (\frac{R}{2A})^n \cdot (\frac{1}{2})^{\frac{n^2+np}{p}} \cdot 1 = \left(\frac{R}{2^{\frac{n+2p}{p}} A}\right)^n
$$

Thus we complete the proof. □

In many situations, it is necessary to find the best constant in the Sobolev inequalities. One example is the Yamabe problem of prescribing constant scalar curvature on a compact manifold. It was studied by Yamabe [Ya], Trudinger [Tr2] and eventually solved by Aubin [Au2] and Schoen [Sc]. The next theorem is due to T. Aubin [Au].

**Theorem 4.1.3** *Let $(M,g)$ be a smooth, compact Riemann manifold of dimension $n$. For any $\epsilon > 0$ and any $p \in [1,n)$, there exists a constant $B$ such that for any $u \in W^{1,p}(M)$,*

$$
\left(\int_M u^{np/(n-p)} d\mu\right)^{(n-p)/n} \leq (K(n,p)^p + \epsilon) \int_M |\nabla u|^p d\mu + B \int_M |u|^p d\mu.
$$

*Here $K(n,p)$ is the best constant in the Sobolev inequality in $\mathbf{R}^n$.*

**Remark 4.1.1** *Aubin proved that $B$ depends on $\epsilon$, bounds on the injectivity radius, sectional curvatures. Hebey [Heb1] showed that $B$ can be chosen to depend only on $\epsilon$, the injectivity radius and the lower bound of the Ricci curvature. In the case $p = 2$, the above theorem was improved by Hebey and Vaugon [HV] in the form of the next theorem. However the constant $B$ will also depend on the derivatives of the curvature tensor.*

*Proof of Theorem 4.1.3* The proof is almost the same as that of Theorem 4.1.1. The only change is to select local charts $(\Omega_m, \phi_m)$, $m = 1, \ldots, N$ such that for any $m$ the local form of metric $(g_{ij}^{(m)})$ in $(\Omega_m, \phi_m)$ satisfies

$$(1 - \lambda\epsilon)(\delta_{ij}) \leq (g_{ij}^{(m)}) \leq (1 + \lambda\eta\epsilon)(\delta_{ij}).$$

Here $\lambda > 0$ is sufficiently small. □

**Theorem 4.1.4** *Let $(M, g)$ be a smooth, compact Riemann manifold of dimension $n$. There exists a constant $B$ such that for any $u \in W^{1,2}(M)$,*

$$\left(\int_M u^{2n/(n-2)} d\mu\right)^{(n-2)/n} \leq K(n,2)^2 \int_M |\nabla u|^2 d\mu + B \int_M |u|^2 d\mu.$$

*Here $K(n, 2)$ is the best constant in the Sobolev inequality in $\mathbf{R}^n$.*

We just give a sketch of the proof. The technical details are presented in [Heb2] Chapter 7.

It suffices to prove that there exists a sufficiently large $\alpha > 0$ such that

$$I_\alpha(u) \equiv \frac{\int_M |\nabla u|^2 d\mu + \alpha \int_M u^2 d\mu}{\left(\int_M |u|^{2n/(n-2)} d\mu\right)^{(n-2)/n}} \geq \frac{1}{K(n,2)^2}$$

for all $u \in W^{1,2}(M)$, $u \neq 0$. One uses the method of contradiction. Suppose the conclusion were false, then for any $\alpha > 0$,

$$\inf_{0 \neq u \in W^{1,2}(M)} I_\alpha(u) < \frac{1}{K(n,2)^2}.$$

A minimizer of this functional then satisfies the Euler-Lagrange equation

$$\Delta u_\alpha - \alpha u_\alpha + \lambda_\alpha u_\alpha^{(n+2)/(n-2)} = 0.$$

Here $\lambda_\alpha \in (0, K(n,2)^{-2})$ and $\int_M u_\alpha^{2n/(n-2)} d\mu = 1$. Now ones uses a blow-up argument to prove that for sufficiently large $\alpha$, the above equation does not have a solution. This leads to a contradiction. □

## 4.2 Sobolev inequalities, log Sobolev inequalities, heat kernel upper bound and Nash inequality

We begin the section with the concept of heat kernels.

The heat kernel is a fundamental solution to the heat equation

$$\Delta u - \partial_t u = 0 \tag{4.2.1}$$

defined on $M \times (0, \infty)$. Here $\Delta$ is the Laplace Beltrami operator acting on a scalar function $u = u(x,t)$. One way to define $G$ is to require, for any fixed $y \in M$,

$$\begin{cases} \Delta G(x,t;y) - \partial_t G(x,t;y) = 0, x \in M, t > 0 \\ G(x,0;y) = \delta(x,y). \end{cases} \tag{4.2.2}$$

Here $\Delta$ acts on the $x$ variable and $\delta = \delta(x,y)$ is the Dirac delta function concentrated at $y$. When $M$ is a compact Riemann manifold or certain noncompact ones, it is known that $G$ is smooth when $t > 0$. For this and some other basic properties of the heat kernel, see [Da] e.g.

Similarly one can define the heat kernel on a domain $D \subset M$ satisfying certain boundary conditions. For example the Dirichlet heat kernel on $D$ is the one satisfying:

$$\begin{cases} \Delta_y G(y,t,z) - \partial_t G(y,t,z) = 0, \quad y, z \in D, \quad t > 0 \\ G(y,t,z) = 0, \quad y \in \partial D, z \in D, t > 0, \\ G(y,0,z) = \delta(y,z). \end{cases} \tag{4.2.3}$$

If $\partial D$ is sufficiently smooth, say $C^1$, one can also define the Neumann heat kernel as a function $G$ satisfying

$$\begin{cases} \Delta_y G(y,t,z) - \partial_t G(y,t,z) = 0, \quad y, z \in D, \quad t > 0 \\ \frac{\partial G(y,t,z)}{\partial n_y} = 0, \quad y \in \partial D, z \in D, t > 0, \\ G(y,0,z) = \delta(y,z) \end{cases} \tag{4.2.4}$$

where $n_y$ is the exterior normal of $\partial D$ at $y$.

The first theorem of this section establishes the equivalence of a Sobolev inequality with a number of other inequalities. It is the synopsis of the work of several authors: E. B. Davies, L. Gross and J. Nash.

**Theorem 4.2.1** *Let $M$ be a $n$ dimensional compact Riemann manifold without boundary. Suppose $n \geq 3$. Then the following inequalities are equivalent up to constants.*

*(I) Sobolev inequality: there exist positive constants $A$ and $B$ such that, for all $v \in W^{1,2}(M)$,*

$$\left(\int_M v^{2n/(n-2)} d\mu\right)^{(n-2)/n} \le A \int_M |\nabla v|^2 d\mu + B \int_M v^2 d\mu;$$

*(II) Log-Sobolev inequality: for all $v \in W^{1,2}(M)$ such that $\|v\|_2 = 1$ and all $\epsilon > 0$,*

$$\int_M v^2 \ln v^2 \, d\mu \le \epsilon^2 \int_M |\nabla v|^2 d\mu - \frac{n}{2} \ln \epsilon^2 + BA^{-1}\epsilon^2 + \frac{n}{2} \ln \frac{nA}{2e}.$$

*(III) Heat kernel upper bound: for all $t > 0$,*

$$G(x, t; y) \le \frac{(nA)^{\frac{n}{2}}}{t^{n/2}} \, e^{A^{-1}B\,t};$$

*(IV) Nash inequality: for all $v \in W^{1,2}(M)$,*

$$\|v\|_2^{2+\frac{4}{n}} \le \left(A\|\nabla v\|_2^2 + B\|v\|_2^2\right) \, \|v\|_1^{\frac{4}{n}};$$

*More specifically, the following relations hold:*

*(I) Sobolev inequality $\Rightarrow$ (II) Log-Sobolev inequality $\Rightarrow$ (III) Heat kernel upper bound $\Rightarrow$ (I) Sobolev inequality with $A$ and $B$ replaced by const.$A$ and const.$B$ respectively;*

*(I) Sobolev inequality $\Rightarrow$ (IV) Nash inequality $\Rightarrow$ (III) Heat kernel upper bound.*

*Finally, the conclusion still holds for $v \in W_0^{1,2}(D)$ where $D$ is a Lipschitz domain in $M$.*

Proof. **(for (I) $\Rightarrow$ (II): Sobolev inequality $\Rightarrow$ Log-Sobolev inequality)**

The proof is a quick application of the Jensen's inequality. The assumed Sobolev inequality is: for all $v \in W^{1,2}(M)$,

$$\left(\int_M v^{2n/(n-2)} d\mu\right)^{(n-2)/n} \le A \int_M |\nabla v|^2 d\mu + B \int_M v^2 d\mu.$$

Given $v \in W^{1,2}(M)$ such that $\|v\|_2 = 1$, we introduce the measure

$$dw(x) = v^2(x) \, d\mu(x).$$

Then

$$\int_M v^2 \, d\mu = 1 \Rightarrow \int_M dw = 1.$$

Note $\ln\phi$ is a concave function of $\phi$. Hence we can take $\phi = v^{q-2}$, with $q = \frac{2n}{n-2}$, and apply Jensen's inequality, $\int \ln\phi\, dw \le \ln\int \phi\, dw$, to deduce

$$
\begin{aligned}
\Rightarrow \quad & \int (\ln v^{q-2}) v^2\, d\mu \le \ln \int v^{q-2} v^2\, d\mu = \ln \|v\|_q^q \\
\Rightarrow \quad & \int v^2 \ln v\, d\mu \le \frac{q}{q-2} \ln \|v\|_q \qquad \left(\frac{q}{q-2} = \frac{n}{2}\right) \\
\Rightarrow \quad & \int v^2 \ln v^2\, d\mu \le \frac{n}{2} \ln \|v\|_q^2 \\
& \le \frac{n}{2} \ln\left(A \|\nabla v\|_2^2 + B\|v\|_2^2\right) \\
& \le \frac{n}{2} \ln\left(A \|\nabla v\|_2^2 + B\right). \qquad (\|v\|_2 = 1)
\end{aligned}
$$

We estimate the last quantity by the elementary inequality,

$$\ln x \le \sigma x - 1 - \ln\sigma$$

for all $\sigma > 0$. This holds since for $f(x) \equiv \sigma x - lnx - 1 - \ln\sigma$, we have $f(1/\sigma) = f'(1/\sigma) = 0$; further $f''(x) > 0$ and $f(0^+) = +\infty$, $f(+\infty) = +\infty$. Therefore,

$$\int v^2 \ln v^2\, d\mu \le \frac{n}{2} \ln\left(A\|\nabla v\|_2^2 + B\right) \le \frac{n\,\sigma}{2}(A\|\nabla v\|_2^2 + B) - \frac{n}{2}(1 + ln\sigma)$$

that is,

$$\int v^2 \ln v^2\, d\mu \le \frac{n\,\sigma\,A}{2} \int |\nabla v|^2 d\mu + \frac{n\,\sigma\,B}{2} - \frac{n}{2}(1 + \ln\sigma).$$

Taking $\epsilon^2 = n\sigma A/2$, we can convert the above inequality to

$$\int v^2 \ln v^2\, d\mu \le \epsilon^2 \int |\nabla v|^2 d\mu - \frac{n}{2}\ln\epsilon^2 + BA^{-1}\epsilon^2 + \frac{n}{2}\ln\frac{nA}{2e}. \quad (4.2.5)$$

□

Proof. **(for (II) ⇒ (III): Log-Sobolev inequality ⇒ Heat kernel upper bound due to Davies [Da])**

Let $u$ be a smooth solution to the heat equation, then

$$u(x,t) = \int G(x,t;y) u(y,0)\, d\mu(y)$$

where $G$ is the heat kernel. Notice

$$\sup_{u\neq 0}\frac{\|u(\cdot,t)\|_\infty}{\|u(\cdot,0)\|_1}=\sup_{x,y}G(x,t;y).$$

For any fixed $T$, $t\in[0,T]$, consider the norm $\|u(\cdot,t)\|_{p(t)}$, where we want $p(t)$ to be nondecreasing in $t$, and $p(0)=1$, $p(T)=\infty$. For instance we can take

$$p(t)=\frac{T}{T-t},\quad\Rightarrow\quad p(0)=1,\,p(T)=\infty.$$

Let $u=u(x,t)$ be a positive solution to the heat equation, we estimate the time derivative for the quantity

$$\ln\|u\|_{p(t)}=\ln(\int|u|^{p(t)}d\mu)^{1/p(t)}.$$

The motivation is

$$\int_0^T\frac{\partial}{\partial t}\ln\|u\|_{p(t)}\,dt=\ln\frac{\|u(\cdot,T)\|_\infty}{\|u(\cdot,0)\|_1}$$

By direct calculation,

$$\partial_t\|u\|_{p(t)}=\partial_t\left(\int|u|^{p(t)}d\mu\right)^{\frac{1}{p(t)}}=-\|u\|_{p(t)}\cdot\frac{p'(t)}{p^2(t)}\cdot\ln\|u\|_{p(t)}^{p(t)}$$
$$+\frac{\|u\|_{p(t)}^{1-p(t)}}{p(t)}\left(p'(t)\int u^{p(t)}\ln u\,d\mu+p(t)\int u^{p(t)-1}(\triangle u)\,d\mu\right).$$

Multiplying by $p^2(t)\|u\|_{p(t)}^{p(t)}$ on both sides, and applying integration by part on the $\triangle u$ term, we deduce

$$\begin{aligned}p^2(t)\|u\|_{p(t)}^{p(t)}\,\partial_t\|u\|_{p(t)}=&-p'(t)\|u\|_{p(t)}^{1+p(t)}\ln\|u\|_{p(t)}^{p(t)}\\&+p(t)\,p'(t)\|u\|_{p(t)}\int u^{p(t)}\ln u\,d\mu\\&-p^2(t)(p(t)-1)\|u\|_{p(t)}\int u^{p(t)-2}|\nabla u|^2\,d\mu.\end{aligned}$$

Further, we divide both sides by $\|u\|_{p(t)}$ to reach,

$$\begin{aligned}p^2(t)\|u\|_{p(t)}^{p(t)}\,\partial_t(\ln\|u\|_{p(t)})=&-p'(t)\|u\|_{p(t)}^{p(t)}\ln\|u\|_{p(t)}^{p(t)}\\&+p(t)\,p'(t)\int u^{p(t)}\ln u\,d\mu\\&-4(p(t)-1)\int|\nabla(u^{p(t)/2})|^2\,d\mu.\end{aligned}\tag{4.2.6}$$

Write,

$$v = \frac{u^{p(t)/2}(x,t)}{\|u^{p(t)/2}\|_2}.$$

Then,

$$\begin{aligned}
&v^2 = \frac{u^{p(t)}}{\|u\|_{p(t)}^{p(t)}}, \ \|v\|_2 = 1, \ln v^2 = \ln u^{p(t)} - \ln \|u\|_{p(t)}^{p(t)}, \\
\Rightarrow \quad & p'(t) \int v^2 \ln v^2 d\mu = p'(t) \int \frac{u^{p(t)}}{\|u\|_{p(t)}^{p(t)}} \left( \ln u^{p(t)} - \ln \|u\|_{p(t)}^{p(t)} \right) d\mu \\
&= \frac{p(t)p'(t)}{\|u\|_{p(t)}^{p(t)}} \int u^{p(t)} \ln u \, d\mu - p'(t) \ln \|u\|_{p(t)}^{p(t)}.
\end{aligned}$$

We plug in $v$ on the right-hand side of the above equality (4.2.6). Then

$$p^2(t)\, \partial_t \left(\ln \|u\|_{p(t)}\right) = p'(t) \left[ \int v^2 \ln v^2 d\mu - \frac{4(p(t)-1)}{p'(t)} \int |\nabla v|^2 \, d\mu \right].$$

In comparison with the Log-Sobolev inequality (4.2.5), we choose

$$\varepsilon^2 = \frac{4(p(t)-1)}{p'(t)} = \frac{4t(T-t)}{T} \le T.$$

Thus we can apply (4.2.5) to get,

$$p^2(t)\, \partial_t \left(\ln \|u\|_{p(t)}\right) \le p'(t) \Big( -\frac{n}{2} \ln \frac{4t(T-t)}{T} + BA^{-1} \frac{4t(T-t)}{T} + \frac{n}{2} \ln \frac{nA}{2e} \Big).$$

Note that $\frac{p'(t)}{p^2(t)} = \frac{1}{T}$. Substituting this into the above, we arrive at

$$\partial_t \ln \|u\|_{p(t)} \le \frac{1}{T} \left( -\frac{n}{2} \ln \frac{t(T-t)}{T} + \frac{B}{A} T + \frac{n}{2} \ln \frac{nA}{8e} \right).$$

Integrating with respect to $t$ from 0 to $T$,

$$\begin{aligned}
\ln \frac{\|u(x,T)\|_{p(T)}}{\|u(x,0)\|_{p(0)}} &\le -\frac{1}{T} \int_0^T \frac{n}{2} \ln \frac{t(T-t)}{T} \, dt + \frac{B}{A} T + \frac{n}{2} \ln \frac{nA}{8e} \\
&= -\frac{n}{2} \ln T + n + \frac{B}{A} T + \frac{n}{2} \ln \frac{nA}{8e}.
\end{aligned}$$

Since $p(T) = \infty$, $p(0) = 1$,

$$\|u(x,T)\|_\infty \le \|u(x,0)\|_1 \frac{\exp\left(\frac{B}{A} T + n + \frac{n}{2} \ln \frac{nA}{8e}\right)}{T^{n/2}}.$$

Because

$$u(x,T) = \int G(x,T,y)u(y,0)dy,$$

we deduce

$$G(x,T,y) \le (\frac{nAe}{8})^{n/2} \frac{\exp\left(A^{-1}B\,T\right)}{T^{n/2}}.$$

This implies the heat kernel bound since $e/8 \le 1$. □

PROOF. **(for (I) ⇒ (IV): Sobolev inequality ⇒ Nash inequality)**

The Nash inequality is just an interpolation between the Hölder inequality and Sobolev inequality. However it gives the quickest access to heat kernel upper bound. There is also an issue about the best constant. See Carlen and Loss [CaLo].

We assume

$$\|v\|^2_{2n/(n-2)} \le A\|\nabla v\|_2^2 + B\|v\|_2^2.$$

Using Hölder inequality, by direct calculation,

$$\begin{aligned}\int v^2\, d\mu &= \int v^{2-\frac{4}{n+2}}\, v^{\frac{4}{n+2}}\, d\mu = \int v^{\frac{2n}{n+2}} v^{\frac{4}{n+2}}\, d\mu \\ &\le \left(\int v^{\frac{2n}{n+2}p'}\, d\mu\right)^{\frac{1}{p'}} \left(\int v^{\frac{4p}{n+2}}\, d\mu\right)^{\frac{1}{p}}.\end{aligned}$$

Choose $p = \frac{n+2}{4}$, $p' = \frac{n+2}{n-2}$. Then,

$$\begin{aligned} &\int v^2\, dx \le \left(\int v^{\frac{2n}{n-2}}\, d\mu\right)^{\frac{n-2}{n+2}} \left(\int |v|\, d\mu\right)^{\frac{4}{n+2}} \\ \Rightarrow \qquad &\|v\|_2^{2+\frac{4}{n}} \le \left(\int v^{\frac{2n}{n-2}}\, d\mu\right)^{\frac{n-2}{n}} \left(\int |v|\, d\mu\right)^{\frac{4}{n}}.\end{aligned}$$

Hence

$$\|v\|_2^{2+\frac{4}{n}} \le \left(A\|\nabla v\|_2^2 + B\|v\|_2^2\right)\, \|v\|_1^{\frac{4}{n}} \quad \forall\, v \in W^{1,2}(M). \tag{4.2.7}$$

This is the desired Nash inequality. □

PROOF. **(for (IV) ⇒ (III): Nash inequality ⇒ Heat kernel upper bound)**

Suppose (4.2.7) holds.

Let $G = G(x,t;y)$ be the heat kernel; for any fixed $y \in M$, define

$$v = v(x,t) = G(x,t;y),$$

We have $\int v(x,t)\,d\mu = 1$, i.e. $\|v\|_1 = 1$. By direct calculation,

$$\frac{\partial}{\partial t}\left(\int v^2(x,t)\,d\mu\right) = \int 2v\,v_t\,d\mu = \int 2v\,\triangle v\,d\mu = -2\int |\nabla v|^2\,d\mu$$

Also from Nash inequality (4.2.7) with $\|v\|_1 = 1$,

$$\begin{aligned} \|v\|_2^{2+\frac{4}{n}} &\le \left(A\|\nabla v\|_2^2 + B\,\|v\|_2^2\right) \\ \Rightarrow -\|\nabla v\|_2^2 &\le -\frac{1}{A}\|v\|_2^{2+\frac{4}{n}} + \frac{B}{A}\,\|v\|_2^2 \end{aligned}$$

Combining the above two estimates, we arrive at

$$\frac{\partial}{\partial t}\left(\int v^2(x,t)\,d\mu\right) \le -\alpha\|v\|_2^{2+\frac{4}{n}} + \beta\,\|v\|_2^2$$

where $\alpha = \frac{2}{A}$, $\beta = \frac{2B}{A}$. Denote

$$f(t) = \int v^2(x,t)\,d\mu, \quad g(t) = e^{-\beta t} f(t).$$

Then

$$\begin{aligned} &\frac{\partial}{\partial t} f(t) \le -\alpha[\,f(t)\,]^{1+\frac{2}{n}} + \beta\,f(t) \\ \Rightarrow \quad &\frac{\partial}{\partial t} g(t) = -\beta e^{-\beta t} f(t) + e^{-\beta t} f'(t) \\ &\qquad \le -\beta e^{-\beta t}\,f(t) + e^{-\beta t}\,\beta\,f(t) - \alpha e^{-\beta t}[\,f(t)\,]^{1+\frac{2}{n}} \\ \Rightarrow \quad &\frac{\partial}{\partial s} g(s) \le -\alpha e^{\frac{2\beta s}{n}}[\,g(s)\,]^{1+\frac{2}{n}}, \quad s \in (0,t]. \end{aligned}$$

Integrating from $\frac{t}{2}$ to $t$,

$$\begin{aligned} \Rightarrow \quad & -\frac{n}{2}\left([\,g(s)\,]^{-\frac{2}{n}}\right)_{s=\frac{t}{2}}^{s=t} \le -\frac{n\alpha}{2\beta}[\,e^{\frac{2\beta t}{n}} - e^{\frac{\beta t}{n}}\,] \\ \Rightarrow \quad & \frac{n}{2}\,e^{\frac{2\beta t}{n}}[\,f(t)\,]^{-\frac{2}{n}} \ge \frac{n\alpha}{2\,\beta}\left(e^{\frac{2\beta t}{n}} - e^{\frac{\beta t}{n}}\right) \\ \Rightarrow \quad & e^{\frac{\beta t}{n}}[\,f(t)\,]^{-\frac{2}{n}} \ge \frac{\alpha}{\beta}\left(e^{\frac{\beta t}{n}} - 1\right) \ge \frac{\alpha}{\beta}\cdot\frac{\beta\,t}{n} = \frac{t\alpha}{n} \qquad (4.2.8) \\ \Rightarrow \quad & f(t) \le \frac{(n/\alpha)^{\frac{n}{2}}}{t^{n/2}}\,e^{(\beta\,t)/2} \\ & \text{i.e.} \int G(x,t;y)\,G(x,t;y)\,d\mu(y) \le \frac{(n/\alpha)^{\frac{n}{2}}}{t^{n/2}}\,e^{(\beta\,t)/2}. \end{aligned}$$

Notice $G(x,t;y)$ is symmetric with respect to $x$ and $y$. Thus

$$\int G(x,t;y)\,G(y,t;x)\,d\mu(y) \le \frac{(n/\alpha)^{\frac{n}{2}}}{t^{n/2}}\,e^{(\beta\,t)/2}.$$

By the reproducing property, for any fixed $x \in M$,

$$G(x,2t;x) = \int G(x,t;y)\,G(y,t;x)\,d\mu(y) \le \frac{(n/\alpha)^{\frac{n}{2}}}{t^{n/2}}\,e^{(\beta\,t)/2}.$$

Further,

$$\begin{aligned} G(x,t;y) &= \int G(x,t/2;z)\,G(z,t/2;y)\,d\mu(z) \\ &\le \left(\int G^2(x,t/2;z)\,d\mu(z)\right)^{1/2} \cdot \left(\int G^2(z,t/2;y)\,d\mu(z)\right)^{1/2} \\ &= [G(x,t;x)]^{1/2} \cdot [G(y,t;y)]^{1/2} \\ &\le \frac{(2n/\alpha)^{\frac{n}{2}}}{t^{n/2}}\,e^{(\beta\,t)/2}, \end{aligned}$$

that is, we have the heat kernel upper bound,

$$G(x,t;y) \le \frac{(2n/\alpha)^{\frac{n}{2}}}{t^{n/2}}\,e^{(\beta\,t)/2}$$

Recall $\alpha = 2/A$ and $\beta = 2B/A$. We deduce

$$G(x,t;y) \le \frac{(nA)^{\frac{n}{2}}}{t^{n/2}}\,e^{A^{-1}B\,t}.$$

When $t$ is large the above bound can be improved. Indeed, from the inequality just before (4.2.8), we have, for $t \ge 1$,

$$f(t) = G(x,2t;x) \le \frac{e^{(\beta\,t)/2}}{(e^{\beta t/n}-1)^{n/2}}(\beta/\alpha)^{n/2} \le \frac{e^{B/A}}{(e^{2B/(nA)}-1)^{n/2}}B^{n/2}.$$

Following the above argument, we obtain, for some positive constant $C = C(A,B)$.

$$G(x,t;y) \le \min\left(\frac{(nA)^{\frac{n}{2}}}{t^{n/2}}e^{A^{-1}B\,t}, C(A,B)\right). \tag{4.2.9}$$

□

PROOF. **for (III) $\Rightarrow$ (I): heat kernel upper bound implying Sobolev inequality**.

The assumption in this part is: for some $c_1, c_2 > 0$,

$$G(x,t;y) \le \frac{c_1}{t^{n/2}} e^{c_2 t}$$

for all $x, y \in M$ and $t > 0$. For example, we can take $c_1 = (nA)^{n/2}$ and $c_2 = A^{-1}B$ as in (III).

We follow the presentation in [Da].

Let $H = H(x,t,y)$ be the heat kernel of the equation

$$\Delta u - c_2 u - \partial_t u = 0$$

Since $G(x,t,y) \le c_1 e^{c_2 t}/t^{n/2}$, we know that

$$H(x,t,y) = e^{-c_2 t} G(x,t,y) \le \frac{c_1}{t^{n/2}}.$$

Moreover

$$\int_M H(x,t,y) d\mu(y) \le \int_M G(x,t,y) d\mu(y) = 1.$$

So, for all $f \in L^2(M)$,

$$\begin{aligned} \|H * f\|_\infty &= \sup_{x \in M} |\int H(x,t,y) f(y) d\mu(y)| \\ &\le \sup_x \left( \int_M H^2(x,t,y) d\mu(y) \right)^{1/2} \ \|f\|_2 \\ &\le \frac{\sqrt{c_1}}{t^{n/4}} \int H(x,t,y) d\mu(y) \ \|f\|_2 \\ &\le \frac{\sqrt{c_1}}{t^{n/4}} \ \|f\|_2. \end{aligned}$$

Similarly, by Hölder inequality, for all $q \in [1, n)$ and $q' = q/(q-1)$, it holds

$$\|H * f\|_\infty \le \sup_x \left( \int_M H^{q'}(x,t,y) d\mu(y) \right)^{1/q'} \ \|f\|_q \le \frac{c_1^{1/q}}{t^{n/(2q)}} \ \|f\|_q. \tag{4.2.10}$$

We consider the integral operator

$$L \equiv (\sqrt{-\Delta + c_2})^{-1}. \tag{4.2.11}$$

Since $\Delta$ is a self-adjoint operator, by eigenfunction expansion (Laplace transform), we have this important relation: for $f \in C_0^\infty(M)$,

$$\begin{aligned}(Lf)(x) &= \Gamma(1/2)^{-1} \int_0^\infty t^{-1/2} [e^{(\Delta - c_2)t} f](x,t) dt \\ &= \Gamma(1/2)^{-1} \int_0^\infty t^{-1/2} (H * f)(x,t) dt.\end{aligned}$$

Here $e^{(\Delta - c_2)t} f$ is the semigroup notation for $H * f$.

For a fixed $T > 0$ we write

$$Lf \equiv L_1 f + L_2 f$$

where

$$L_1 f(x) = \Gamma(1/2)^{-1} \int_0^T t^{-1/2} [H * f](x,t) dt,$$

$$L_2 f(x) = \Gamma(1/2)^{-1} \int_T^\infty t^{-1/2} [H * f](x,t) dt.$$

For any $\lambda > 0$, observe that

$$|\{x \,||Lf(x)| \ge \lambda\}| \le |\{x \mid |L_1 f(x)| \ge \lambda/2\}| + |\{x \mid |L_2 f(x)| > \lambda/2\}|. \tag{4.2.12}$$

By (4.2.10) and the definition of $L_2 f$,

$$\|L_2 f\|_\infty \le c_1^{1/q} \int_T^\infty t^{-1/2 - n/(2q)} \|f\|_q dt = c c_1^{1/q} T^{1/2 - n/(2q)} \|f\|_q.$$

Now we choose $T$ so that

$$\frac{\lambda}{2} = c c_1^{1/q} T^{1/2 - n/(2q)} \|f\|_q. \tag{4.2.13}$$

Then (4.2.12) becomes

$$|\{x \,||Lf(x)| \ge \lambda\}| \le |\{x \mid |L_1 f(x)| \ge \lambda/2\}|$$

since the set

$$\{x \mid |L_2 f(x)| > \lambda/2\}$$

is empty. Hence

$$\begin{aligned}|\{x \,||Lf(x)| \ge \lambda\}| &\le |\{x \mid |L_1 f(x)| \ge \lambda/2\}| \\ &\le (\lambda/2)^{-q} \int_M |L_1 f(x)|^q d\mu(x).\end{aligned}$$

By Minkowski inequality and Young's inequality

$$\begin{aligned}\|L_1 f\|_q &\le \Gamma(1/2)^{-1}\int_0^T t^{-1/2}\|H * f(\cdot,t)\|_q dt\\ &\le \Gamma(1/2)^{-1}\int_0^T t^{-1/2}\sup_x \|H(x,t,\cdot)\|_1\ \|f\|_q dt\\ &\le cT^{1/2}\|f\|_q.\end{aligned}$$

This shows

$$|\{x\ ||Lf(x)| \ge \lambda\}| \le c(\lambda/2)^{-q}T^{q/2}\|f\|_q^q.$$

By our choice of $T$ (4.2.13), this is equivalent to

$$|\{x\ ||Lf(x)| \ge \lambda\}| \le c(c_1)^{q/(n-q)}\lambda^{-r}\|f\|_q^r$$

where $r = qn/(n-q)$. Hence $L$ is a linear operator that sends $L^q$ space into weak $L^r$ space, for all $q \in [1,n)$. By the Marcinkiewicz interpolation lemma, we know that $L$ is a bounded operator from $L^2$ to $L^p$ with $p = 2n/(n-2)$ (taking $q = 2$). i.e

$$\|Lu\|_p \le c(c_1)^{1/n}\|u\|_2 \tag{4.2.14}$$

for all $u \in C_0^\infty(M)$. Define

$$v = Lu.$$

Then $u = L^{-1}v$ and

$$\begin{aligned}\|u\|_2^2 &= < L^{-1}v, L^{-1}v > = < L^{-2}v, v > = < -\Delta v + c_2 v, v >\\ &= \int_M (|\nabla v|^2 + c_2 v^2)d\mu.\end{aligned}$$

Substituting this to (4.2.14), we arrive at the Sobolev inequality

$$\|v\|_p^2 \le const.(c_1)^{2/n}(\|\nabla v\|_2^2 + c_2\|v\|_2^2).$$

If we take $c_1 = (nA)^{n/2}$ and $c_2 = A^{-1}B$ as in (III), then we have a Sobolev inequality with the claimed constants:

$$\|v\|_p^2 \le const.A\|\nabla v\|_2^2 + const.B\|v\|_2^2.$$

This proves the theorem when $M$ is a compact manifold without boundary. The proof of the last statement of the theorem (the result for the Dirichlet case) follows verbatim. □

Next we explain a result of A. Grigor'yan [Gr2] and L. Saloff-Coste [Sal2], proven independently in the 1990s. It claims that certain Sobolev inequality is equivalent to a Poincaré inequality and the volume doubling condition; and it is also equivalent to a parabolic Harnack inequality for positive solutions of the heat equation. These results can also be regarded as local version of the previous theorem.

Given $M$, a connected complete noncompact Riemann manifold, denote $B(x,r)$ the geodesic ball of center $x \in M$ and radius $r > 0$, denote $vol(B(x,r))$ by $|B(x,r)|$. We have the following definitions:

**Definition 4.2.1** *We say $M$ satisfies the volume doubling property, if there exists a constant $d_0 > 0$ such that*

$$|B(x,2r)| \leq d_0\, |B(x,r)|, \quad \forall\, x \in M, r > 0 \tag{4.2.15}$$

**Definition 4.2.2** *We call the weak $p$-Poincaré inequalities hold on $M$ if there exists a constant $c_p > 0$ such that $\forall\, f \in C^\infty(M)$, $x \in M, r > 0$,*

$$\int_{B(x,r)} |f(y) - f_r(x)|^p \, d\mu(y) \leq c_p\, r^p \int_{B(x,2r)} |\nabla f(y)|^p \, d\mu(y), \quad \textit{where}$$
$$f_r(x) = \frac{\int_{B(x,r)} f(y)\, d\mu(y)}{|B(x,r)|}; \tag{4.2.16}$$

*in particular, for $p = 2$, the above inequality is called the weak $L^2$ Poincaré inequality, i.e. for some constant $P_2 > 0$ and $\forall\, f \in C^\infty(M)$,*

$$\int_{B(x,r)} |f(y) - f_r(x)|^2 \, d\mu(y) \leq P_2\, r^2 \int_{B(x,2r)} |\nabla f(y)|^2 \, d\mu(y). \tag{4.2.17}$$

**Remark 4.2.1** *The word "weak" in front of the Poincaré inequality reflects the fact that the ball on the right-hand side has twice the radius of the ball on the left-hand side. In the Euclidean case a stronger form of the Poincaré inequality holds, i.e. the radius of the balls on either side is the same.*

The volume doubling property actually has several interesting implications which we summarize below.

**Remark 4.2.2** *On any complete metric space with doubling property (4.2.15), the following inequalities are true.*

*(a). If $y \in B(x,r)$, then*

$$0 < \frac{1}{d_0} \le \frac{|B(y,r)|}{|B(x,r)|} \le d_0. \tag{4.2.18}$$

*The reason is*
$B(y,r) \subseteq B(x,2r) \Rightarrow |B(y,r)| \le |B(x,2r)| \le d_0\,|B(x,r)|$.

*(b). For any $r > s > 0$,*

$$|B(x,r)| \le d_0\,(\frac{r}{s})^{\log_2 d_0}|B(x,s)|.$$

*Indeed, there exists some $i \in \mathbb{Z}^+$ such that $r \in [\,2^{i-1}s, 2^i s\,]$. Hence*

$$|B(x,r)| \le |B(x,2^i s)| \le d_0^i|B(x,s)|.$$

*Now* $\frac{r}{s} \ge 2^{i-1} \Rightarrow i \le 1 + \log_2 \frac{r}{s} \Rightarrow$

$$|B(x,r)| \le d_0 \cdot d_0^{\log_2 \frac{r}{s}}\,|B(x,s)| = d_0 \cdot 2^{(\log_2 d_0)\log_2 \frac{r}{s}}\,|B(x,s)|$$
$$\Rightarrow |B(x,r)| \le d_0 \cdot (\frac{r}{s})^{\log_2 d_0}\,|B(x,s)|.$$

*(c). For any $r > s > 0$, if $x \in B(y,r)$, then*

$$\frac{|B(y,r)|}{|B(x,s)|} = \frac{|B(y,r)|}{|B(x,r)|} \cdot \frac{|B(x,r)|}{|B(x,s)|} \le d_0^2 \cdot (\frac{r}{s})^{\log_2 d_0}. \tag{4.2.19}$$

*Further, if $B(y,r) \cap B(x,s) \ne \emptyset$, then $x \in B(y,r+2s)$. Therefore*

$$\frac{|B(y,r)|}{|B(x,s)|} \le \frac{|B(y,r+2s)|}{|B(x,s)|} \le d_0^2 \cdot \left(\frac{r+2s}{s}\right)^{\log_2 d_0}. \tag{4.2.20}$$

The doubling property (4.2.15) and weak Poincaré inequality (4.2.17) imply a family of Sobolev inequalities on balls. This is illustrated by the following theorem.

**Theorem 4.2.2** *Let $M$ be a connected complete noncompact Riemann $n$-manifold. Suppose there hold on $M$ the doubling property (4.2.15) and weak $L^2$ Poincaré inequality (4.2.17). Then the following Nash inequality is true:*

*Let $\nu = \log_2 d_0$, where $d_0$ is the constant in (4.2.15). Then $\forall f \in C_0^\infty(B(x,r))$*

$$\|f\|_2^{2+\frac{4}{\nu}} \le \frac{C_0\,r^2}{|B(x,r)|^{\frac{2}{\nu}}}\left(\|\nabla f\|_2^2 + r^{-2}\|f\|_2^2\right)\,\|f\|_1^{\frac{4}{\nu}}.$$

*Here $C_0 > 0$ depends only on $d_0$, $P_2$, the constants in the doubling property and weak Poincaré inequality respectively.*

*Moreover, if $\nu > 2$, the following Sobolev inequality is true: $\forall f \in C_0^\infty(B(x,r))$,*

$$\left(\int_M |f|^{2\nu/(\nu-2)} d\mu\right)^{(\nu-2)/\nu} \leq C_S |B(x,r)|^{-2/\nu} r^2 \int_M (|\nabla f|^2 + r^{-2} f^2) d\mu.$$

*Here the constant $C_S$ depends only on $d_0$, $P_2$.*

*If $\nu \leq 2$, then the following Sobolev inequality is true: given any $p > 2$, there exists $C = C(p, d_0, P_2)$ such that, $\forall f \in C_0^\infty(B(x,r))$,*

$$\left(\int_M |f|^{2p/(p-2)} d\mu\right)^{(p-2)/p} \leq C(p, d_0, P_2) |B(x,r)|^{-2/\nu} r^2 \times \int_M (|\nabla f|^2 + r^{-2} f^2) d\mu.$$

The proof of above theorem requires the following two lemmas.

**Lemma 4.2.1** *For $\forall f \in C_0^\infty(B(y,r))$, and any $0 < s \leq r < \infty$, there exists a positive constant $c_3$ depending only on the doubling constant $d_0$ such that,*

$$\|f_s\|_2 \leq \frac{c_3}{\sqrt{|B(y,r)|}} \left(\frac{r}{s}\right)^{\frac{1}{2}\log_2 d_0} \|f\|_1,$$

*where*

$$f_s(x) \equiv \frac{1}{|B(x,s)|} \int_{B(x,s)} f(z) d\mu(z)$$

*is the average of $f$ in the ball $B(x,s)$.*

Proof. First we note that if $B(x,s) \cap B(y,r) = \emptyset$, then because $f \in C_0^\infty(B(y,r))$, we have $f_s(x) = \frac{1}{|B(x,s)|} \int_{B(x,s)} f(z)\, d\mu(z) = 0$. Therefore we can assume $B(x,s) \cap B(y,r) \neq \emptyset$.

Using (4.2.18), we deduce

$$\begin{aligned} \|f_s\|_1 = \int_M |f_s(x)|\, d\mu(x) &\leq \int_M \frac{\int_{B(x,s)} |f(z)|\, d\mu(z)}{|B(x,s)|} d\mu(x) \\ &= \int_M \left(\int_{B(z,s)} \frac{1}{|B(x,s)|} d\mu(x)\right) |f(z)|\, d\mu(z) \\ &\leq \int_M \left(\int_{B(z,s)} \frac{d_0}{|B(z,s)|} dx\right) |f(z)|\, d\mu(z) \\ &= d_0 \|f\|_1 \end{aligned}$$

that is,

$$\|f_s\|_1 \le d_0 \, \|f\|_1.$$

Applying (4.2.20), we have

$$\begin{aligned}
|f_s(x)| &\le \frac{1}{|B(x,s)|} \int_{B(x,s)} |f(z)| \, d\mu(z) \\
&\le \frac{1}{|B(y,r)|} \cdot \frac{|B(y,r)|}{|B(x,s)|} \int_{B(y,r)} |f(z)| \, d\mu(z) \\
&\le \frac{d_0^2 \cdot (\frac{r+2s}{s})^{\log_2 d_0}}{|B(y,r)|} \int_{B(y,r)} |f(z)| \, d\mu(z) \\
&\le \frac{d_0^2 \cdot 3^{\log_2 d_0} \cdot (\frac{r}{s})^{\log_2 d_0}}{|B(y,r)|} \int_{B(y,r)} |f(z)| \, d\mu(z)
\end{aligned}$$

where we have used $f \in C_0^\infty(B(y,r))$; therefore,

$$\|f_s\|_\infty \le \frac{c_3 \cdot (\frac{r}{s})^{\log_2 d_0}}{|B(y,r)|} \|f\|_1.$$

Hence

$$\|f_s\|_2 \le \|f_s\|_\infty^{1/2} \cdot \|f_s\|_1^{1/2} \le \frac{c_3}{\sqrt{|B(y,r)|}} \, (\frac{r}{s})^{\frac{1}{2} \log_2 d_0} \, \|f\|_1$$

□

**Remark 4.2.3** *In Euclidean space* $\mathbf{R}^n$*, the doubling constant* $d_0 = 2^n$*. Then for* $\forall f \in C_0^\infty(B(y,r))$*,*

$$\|f_s\|_2 \le \frac{c_3}{\sqrt{|B(y,r)|}} \, (\frac{r}{s})^{\frac{n}{2}} \, \|f\|_1 = (\frac{c_3}{s^{n/2}}) \, \|f\|_1.$$

□

**Lemma 4.2.2** *Assume the doubling property (4.2.15) and weak* $L^2$ *Poincaré inequality (4.2.17) on* $M$*. Then there exists a constant* $\delta > 0$ *depending only on the doubling constant* $d_0$ *and weak Poincaré constant* $P_2$ *such that for* $\forall f \in C_0^\infty(M)$ *and* $s > 0$*,*

$$\|f - f_s\|_2 \le \delta \, s \, \|\nabla f\|_2.$$

PROOF. First we notice there exists a collection of balls $\{B(x_j, s/2), j \in J\}$ such that

$$\begin{aligned} &B(x_i, s/2) \cap B(x_j, s/2) = \emptyset, \text{ if } i \neq j, \\ &M = \bigcup_{j \in J} 2B_j, \text{ where } 2B_j = B(x_j, s). \end{aligned}$$

We will also use $kB_j$ to denote the ball $B(x_j, ks/2)$ for other $k > 0$. The doubling property implies the overlapping number

$$N_p \equiv \#\{j \in J \,|\, p \in 8B_j = B(x_j, 4s)\} \tag{4.2.21}$$

for any $p \in M$ is uniformly bounded from above by a constant depending only on the doubling constant $d_0$. In fact, for all the balls $B(x_j, 4s)$ in the collection, which contains $p$, it holds $B(x_j, s/2) \subset B(p, 4.5s)$. Since the balls $B(x_j, s/2)$ do not overlap, we have

$$\begin{aligned} |B(p, 8s)| &\geq \sum_{\{j:\, p \in B(x_j, 4s)\}} |B(x_j, \frac{s}{2})| \\ &\geq \sum_{\{j:\, p \in B(x_j, 4s)\}} c_4 \, |B(p, 8s)| = c_4 \, |B(p, 8s)| \, N_p. \end{aligned}$$

Here $c_4$ is a power of $d_0$. Thus $N_p$ is uniformly bounded from above.

We compute

$$\begin{aligned} \|f - f_s\|_2^2 &\leq \sum_{j \in J} \int_{2B_j} |f(x) - f_s(x)|^2 \, d\mu(x) \\ &= \sum_{j \in J} \int_{2B_j} |f(x) - f_{4B_j} + f_{4B_j} - f_s(x)|^2 \, d\mu(x) \\ &\leq 2 \sum_{j \in J} \int_{2B_j} |f(x) - f_{4B_j}|^2 \, d\mu(x) \\ &\quad + 2 \sum_{j \in J} \int_{2B_j} |f_{4B_j} - f_s(x)|^2 \, d\mu(x) \\ &\equiv (a) + (b) \end{aligned}$$

where $f_{4B_j}$ is the average of $f$ in the ball $4B_j = B(x_j, 2s)$.

We estimate

$$
\begin{aligned}
(a) &= 2\sum_{j\in J}\int_{2B_j} |f(x) - f_{4B_j}|^2\, d\mu(x) \\
&\le 2\sum_{j\in J}\int_{4B_j} |f(x) - f_{4B_j}|^2\, d\mu(x) \\
&\le 8\, P_2\, s^2 \sum_{j\in J}\int_{8B_j} |\nabla f(x)|^2\, d\mu(x)
\end{aligned}
$$

by the weak Poincaré inequality (4.2.17).

Also

$$
\begin{aligned}
(b) &= 2\sum_{j\in J}\int_{2B_j} |f_{4B_j} - f_s(x)|^2\, d\mu(x) \\
&= 2\sum_{j\in J}\int_{2B_j} \left( f_{4B_j} - \frac{1}{|B(x,s)|}\int_{B(x,s)} f(z)\, d\mu(z)\right)^2 d\mu(x) \\
&= 2\sum_{j\in J}\int_{2B_j} \frac{1}{|B(x,s)|^2}\left(\int_{B(x,s)} [f_{4B_j} - f(z)]\, d\mu(z)\right)^2 d\mu(x).
\end{aligned}
$$

Applying Hölder inequality to the integral inside, we deduce

$$
\begin{aligned}
(b) &\le 2\sum_{j\in J}\int_{2B_j} \frac{1}{|B(x,s)|^2}\left(\int_{B(x,s)} 1^2\, d\mu(z)\right) \\
&\qquad \times \left(\int_{B(x,s)} [f_{4B_j} - f(z)]^2\, d\mu(z)\right) d\mu(x) \\
&\le 2\sum_{j\in J}\int_{4B_j} \frac{1}{|B(x,s)|}\left(\int_{4B_j} [f(z) - f_{4B_j}]^2\, d\mu(z)\right) d\mu(x) \\
&\le 2\sum_{j\in J}\int_{4B_j} \frac{P_2\,(2s)^2}{|B(x,s)|}\left(\int_{8B_j} |\nabla f(z)|^2\, d\mu(z)\right) d\mu(x)
\end{aligned}
$$

where we have used (4.2.17) again. Further, since $x \in B(x_j, 2s)$, we can use inequality (4.2.19) to show,

$$\begin{aligned}(b) &\leq 8\,P_2\,s^2 \sum_{j\in J} \left( \int_{8B_j} |\nabla f(z)|^2\, d\mu(z) \right) \int_{B(x_j,2s)} \frac{1}{|B(x,s)|}\, d\mu(x) \\ &\leq 8\,P_2\,s^2 \sum_{j\in J} \left( \int_{8B_j} |\nabla f(z)|^2\, d\mu(z) \right) \int_{B(x_j,2s)} \frac{d_0^2\,(\frac{2s}{s})^{\log_2\, d_0}}{|B(x_j,2s)|}\, d\mu(x) \\ &\leq 8\,d_0^3\,P_2\,s^2 \sum_{j\in J} \int_{8B_j} |\nabla f(z)|^2\, d\mu(z).\end{aligned}$$

Combining parts (a)–(b) together, we have

$$\begin{aligned}\|f - f_s\|_2^2 &\leq 8(1+d_0^3)\,P_2\,s^2 \sum_{j\in J} \int_{8B_j} |\nabla f(z)|^2\, d\mu(z) \\ &\leq 8(1+d_0^3)\,P_2\,s^2\,N_p \int_M |\nabla f(z)|^2\, d\mu(z).\end{aligned}$$

Here $N_p$ is the overlapping number in (4.2.21), which has been shown to depend only on the doubling number $d_0$.

Denote $\delta = [8(1+d_0^3)P_2N_p]^{1/2}$, then

$$\|f - f_s\|_2 \leq \delta\, s\, \|\nabla f\|_2.$$

□

Now we are ready to give a proof of the theorem.

PROOF. (of the Nash inequality)

From the previous two lemmas,

$$\begin{aligned}\|f\|_2 &\leq \|f - f_s\|_2 + \|f_s\|_2 \\ &\leq \delta\, s\, \|\nabla f\|_2 + \frac{c_3}{\sqrt{|B(x,r)|}}\,(\frac{r}{s})^{\frac{1}{2}\log_2\, d_0}\, \|f\|_1\end{aligned}$$

when $0 < s \leq r$. Therefore for all $s > 0$, we have

$$\|f\|_2 \leq \delta\, s\, (\|\nabla f\|_2 + r^{-1}\|f\|_2) + \frac{c_3}{\sqrt{|B(x,r)|}}\,(\frac{r}{s})^{\frac{1}{2}\log_2\, d_0}\, \|f\|_1.$$

Notice the left side $\|f\|_2$ is independent of $s$, we minimize the right side over $s > 0$, to deduce:

$$\|f\|_2^{2+(4/\nu)} \leq C(d_0, P_2)|B(x,r)|^{-2/\nu} r^2 (\|\nabla f\|_2^2 + r^{-2}\|f\|_2^2)\ \|f\|_1^{4/\nu}.$$

Here $\nu = \log_2 d_0$. This is the desired Nash inequality.

PROOF. (of the Sobolev inequality.) Consider $G = G(y,t,z)$, the Dirichlet heat kernel in $B(x,r)$, which is defined in (4.2.3).

Following the proof on the full space case verbatim (Theorem 4.2.1), we know that the above Nash inequality induces an upper bound for $G$. The only difference is to replace the dimension $n$ there by the constant $\nu$ in the Nash inequality.

$$G(x,t;y) \le C \frac{A^{\nu/2}}{t^{\nu/2}} e^{A^{-1}Bt}$$

where

$$A = C(d_0, P_2)|B(x,r)|^{-2/\nu} r^2, \quad B = C(d_0, P_2)|B(x,r)|^{-2/\nu}.$$

If $\nu > 2$, then, Davies' argument in the same theorem ((III) to (I)) again shows the Sobolev inequality: for all $v \in C^1(M)$,

$$\left( \int_M |v|^{2\nu/(\nu-2)} d\mu \right)^{(\nu-2)/\nu} \le S(d_0, P_2)|B(x,r)|^{-2/\nu} r^2 \times \int_M (|\nabla v|^2 + r^{-2} v^2) d\mu$$

where $S(d_0, P_2)$ is a positive constant depending only on $d_0$ and $P_2$.

If $\nu \le 2$, then for any $p > 2$, we can enlarge $d_0$ so that $\ln d_0 = p$. Hence there is $a_1$ depending on $p$ and $P_2$ such that

$$G(x,t;y) \le a_1 \frac{A^{p/2}}{t^{p/2}} e^{A^{-1}Bt}.$$

By the same argument as above, we have

$$\left( \int_M |v|^{2p/(p-2)} d\mu \right)^{(p-2)/p} \le C(p, d_0, P_2)|B(x,r)|^{-2/\nu} r^2 \times \int_M (|\nabla v|^2 + r^{-2} v^2) d\mu.$$

This finishes the proof of Theorem 4.2.2. □

## 4.3 Sobolev inequalities and isoperimetric inequalities

Of all bounded smooth domains in the plane $\mathbf{R^2}$ with fixed area say $\pi$, the unit disk has the least perimeter, which is $2\pi$. This fact has been known since ancient times. Another way of stating this fact is: for all bounded smooth domains $\Omega \subset \mathbf{R^2}$,

$$\frac{|\partial\Omega|}{|\Omega|^{1/2}} \geq 2\sqrt{\pi}.$$

The equality holds if and only if $\Omega$ is a disk. This inequality is called the isoperimetric inequality in $\mathbf{R^2}$. The study of isoperimetric inequalities in various settings has been an active subject of research. We refer the interested reader to [Cha] for a nice treatment of this subject.

In this section, we prove one basic theorem which claims that an isoperimetric inequality is equivalent to a $L^1$ Sobolev inequality on a Riemann manifold.

Given a Riemann manifold $M$, we first define its isoperimetric constant and $L^1$ Sobolev constant.

**Definition 4.3.1** *The isoperimetric constant of $M$ is*

$$I \equiv \inf_{\Omega} \frac{|\partial\Omega|}{|\Omega|^{(n-1)/n}},$$

*where $\Omega$ varies over bounded, proper domains in $M$ with $C^1$ boundary.*

*The $L^1$ Sobolev constant of $M$ is*

$$S \equiv \inf_{u\neq 0} \frac{\|\nabla u\|_1}{\|u\|_{n/(n-1)}}$$

*where $u$ varies over $C_c^\infty(M)$, the class of smooth functions with compact support.*

The following theorem was proven by Federer-Fleming and Maźya independently.

**Theorem 4.3.1** *The isoperimetric constant and $L^1$ Sobolev constant are equal, i.e. $I = S$.*

Proof. Let $\Omega$ be a bounded, proper $C^1$ domain in $M$. For sufficiently small $\epsilon$ we introduce the function

$$u_\epsilon(x) = \begin{cases} 1, x \in \Omega, \\ 1 - \epsilon^{-1} d(x, \partial\Omega), \quad x \in \Omega^c, \quad d(x,\partial\Omega) < \epsilon, \\ 0, \qquad x \in \Omega^c, \quad d(x,\partial\Omega) \geq \epsilon. \end{cases}$$

It is clear that $u_\epsilon$ is a Lipschitz function and that

$$\lim_{\epsilon\to 0}\|\nabla u_\epsilon\|_1 = |\partial\Omega|,$$

$$\lim_{\epsilon\to 0}\|u_\epsilon\|_{n/(n-1)} = |\Omega|^{(n-1)/n}.$$

Hence

$$I = inf_\Omega \frac{|\partial\Omega|}{|\Omega|^{(n-1)/n}} = inf_\Omega \lim_{\epsilon\to 0}\frac{\|\nabla u_\epsilon\|_1}{\|u_\epsilon\|_{n/(n-1)}} \geq S.$$

Note $u_\epsilon$ may not be an eligible function in the definition of $L^1$ Sobolev constant since it is not $C^\infty$ in general. However one can approximate $u_\epsilon$ by a sequence of smooth functions to make the argument rigorous.

Next we need to prove $I \leq S$, i.e.

$$I\ \|u\|_{n/(n-1)} \leq \|\nabla u\|_1 \tag{4.3.1}$$

for any $u \in C_c^\infty(M)$. For a given $u \in C_c^\infty(M)$ and $t \geq 0$ we consider the sets

$$\Omega(t) = \{x \mid |u(x)| > t\}, \qquad L(t) = \partial\Omega(t).$$

By the coarea formula (cf. Theorem 2.7.1 [Zi] e.g.),

$$\int_M |\nabla u| d\mu = \int_0^\infty |L(t)|dt \geq I\int^\infty |\Omega(t)|^{(n-1)/n}dt.$$

Also

$$\int_M |u|^{n/(n-1)}d\mu = \frac{n}{n-1}\int_0^\infty t^{1/(n-1)}|\Omega(t)|dt.$$

So the proof of (4.3.1) is reduced to proving

$$\left(\frac{n}{n-1}\int_0^\infty t^{1/(n-1)}|\Omega(t)|dt\right)^{(n-1)/n} \leq \int_0^\infty |\Omega(t)|^{(n-1)/n}dt. \tag{4.3.2}$$

To prove this inequality we consider the functions

$$F(s) \equiv \left(\frac{n}{n-1}\int_0^s t^{1/(n-1)}|\Omega(t)|dt\right)^{(n-1)/n},$$

$$G(s) \equiv \int_0^s |\Omega(t)|^{(n-1)/n}dt.$$

By direct calculation and the fact that $|\Omega(t)|$ is a nonincreasing function of $t$, we have

$$F'(s) \leq G'(s), \qquad s \geq 0.$$

This and $F(0) = G(0) = 0$ show that $F(\infty) \le G(\infty)$. Hence we have proven (4.3.2) and consequently $I \le S$. $\square$

Note that the $L^1$ Sobolev inequality

$$\|u\|_{n/(n-1)} \le c_1 \|\nabla u\|_1$$

easily implies the $L^2$ Sobolev inequality

$$\|u\|_{2n/(n-2)} \le c_2 \|\nabla u\|_2$$

for all $u \in C_c^\infty$. Therefore, if the isoperimetric constant of $M$ is positive, then the above $L^2$ Sobolev inequality holds on $M$. But in general the isoperimetric constant can be 0.

**Exercise 4.3.1** *Prove that the $L^1$ Sobolev inequality implies the $L^2$ Sobolev inequality.*

Next we discuss the relation between a Sobolev inequality and the Faber-Krahn inequality. The later can also be regarded as a type of isoperimetric inequality. This time the quantities involved are the volume and the principal eigenvalues of a domain.

Let $\Omega \subset M$ be a precompact domain with smooth boundary. The principal eigenvalue is

$$\lambda(\Omega) = \inf_{0 \ne u \in C_0^\infty(\Omega)} \frac{\int_\Omega |\nabla u|^2 d\mu}{\int_\Omega u^2 d\mu}. \tag{4.3.3}$$

**Theorem 4.3.2** *Let $M$ be a Riemann manifold of dimension $n \ge 3$. The $L^2$ Sobolev inequality: for all $u \in C_c^\infty(M)$*

$$\|u\|_{2n/(n-2)} \le s_2 \|\nabla u\|_2$$

*is equivalent to the Faber-Krahn inequality:*

$$\lambda(\Omega) \ge k|\Omega|^{-2/n}$$

*for all precompact domain $\Omega$ with smooth boundary. Here $s_2$ and $k$ are positive constants.*

PROOF. One direction of proof is very simple. Suppose the $L^2$ Sobolev inequality holds. Let $u$ be a normalized eigenfunction of $\lambda(\Omega)$. Then

$$\Delta u + \lambda(\Omega) u = 0, \qquad \|u\|_{L^2(\Omega)} = 1.$$

By assigning 0 value outside of $\Omega$, we can regard $u$ as a weakly differentiable function with compact support in $M$. Applying the $L^2$ Sobolev inequality and using integration by parts, we have

$$\|u\|_{2n/(n-2)}^2 \le s_2^2\|\nabla u\|_2^2 = s_2^2\|u\Delta u\|_1 = s_2^2\lambda(\Omega)\|u\|_2^2 = s_2^2\lambda(\Omega).$$

The Hölder inequality implies

$$1 = \|u\|_2^2 \le \|u\|_{2n/(n-2)}^2 |\Omega|^{2/n}.$$

Combining the last two inequalities, we deduce

$$\lambda(\Omega) \ge s_2^{-2}|\Omega|^{-2/n}$$

which is the Faber-Krahn inequality with $k = s_2^{-2}$.

Next we prove that the Faber-Krahn inequality implies the $L^2$ Sobolev inequality. Actually we will first prove an upper bound for the heat kernel $G = G(x,t,y)$ on the whole manifold $M$. The $L^2$ Sobolev inequality then follows as a result of Theorem 4.2.1.

Fixing $y$, let $u = u(x,t) = G(x,t,y)$. Consider the integral

$$I(t) \equiv \int_M u^2(x,t)d\mu.$$

Using integration by parts, we have

$$I'(t) = 2\int_M uu_t d\mu = -2\int_M |\nabla u|^2(x,t)d\mu. \qquad (4.3.4)$$

For any positive constant $s$, we know that

$$u^2 \le (u-s)_+^2 + 2su.$$

Therefore

$$I(t) \le \int_M (u-s)_+^2 d\mu + \int_M 2su d\mu.$$

For fixed $s, t > 0$, consider the domains

$$D(s,t) \equiv \{x \mid x \in M, u(x,t) > s\}$$

and its principal eigenvalue

$$\lambda(D(s,t)) = \inf_{0 \ne v \in C_0^\infty(D(s,t))} \frac{\|\nabla v\|_2^2}{\|v\|_2^2}.$$

By taking $v = (u - s)_+$, we deduce

$$I(t) \leq \int_M |\nabla(u-s)_+|^2 d\mu \ \lambda(D(s,t))^{-1} + 2s$$

which implies

$$I(t) \leq \int_M |\nabla u|^2 d\mu \ \lambda(D(s,t))^{-1} + 2s. \tag{4.3.5}$$

Here we have used the property that $\int_M u d\mu = 1$ which also shows

$$|D(s,t)| \leq s^{-1}.$$

The Faber-Krahn inequality then tells us

$$\lambda(D(s,t)) \geq k|D(s,t)|^{-2/n} \geq ks^{2/n}.$$

If $D(s,t)$ is not precompact, we can use a sequence of precompact domains to approximate it. This and (4.3.5) imply that

$$I(t) \leq \int_M |\nabla u|^2 d\mu \ k^{-1} s^{-2/n} + 2s.$$

Minimizing the right-hand side, we deduce

$$I(t) \leq c(n) k^{-n/(n+2)} \left( \int_M |\nabla u|^2 d\mu \right)^{n/(n+2)}.$$

Using this and (4.3.4) we arrive at the inequality

$$I'(t) \leq -c(n)k \ I(t)^{(n+2)/n}.$$

Integrating from $t/2$ to $t$, we know that

$$I(t) \leq \frac{c(n,k)}{t^{n/2}}, \quad t > 0.$$

The reproducing property of the heat kernel then shows, as in the proof of the Nash inequality in Theorem 4.2.1, that

$$G(x,t,y) \leq \frac{c(n,k)}{t^{n/2}}, \quad t > 0.$$

Here the value of $c(n,k)$ may have changed. As mentioned earlier the $L^2$ Sobolev inequality now follows as a result of Theorem 4.2.1. □

## 4.4 Parabolic Harnack inequality

We will prove the following Harnack inequality for solutions of the heat equation on certain manifolds satisfying volume doubling property and weak Poincaré inequality. Our presentation is modeled on that in [Sal]. The result is a generalization of Moser's [Mo1] Harnack inequality for second order parabolic equations of divergence form and with bounded coefficients in $\mathbf{R^n}$. For simplicity we will just consider the heat equation.

**Theorem 4.4.1** *(Harnack inequality) Let $M$ be a connected complete noncompact Riemann $n$-manifold. Then the doubling property (4.2.15) and weak $L^2$ Poincaré inequality (4.2.17) together is equivalent to the following Harnack inequality:*

*Let $u$ be a positive solution the heat equation in $Q = B(x_0, r) \times [t_0 - r^2, t_0]$, then $u$ satisfies*

$$\sup_{Q_-} u \le C_H \inf_{Q_+} u.$$

*Here $Q_- = B(x, \delta r) \times [t_0 - \eta r^2, t_0 - \rho r^2]$, $Q_+ = B(x, \delta r) \times [t_0 - \epsilon r^2, t_0]$, $0 < \epsilon < \rho < \eta < 1$, $0 < \delta < 1$, and $C_H$ is a positive constant depending only on $\epsilon, \eta, \delta, \rho$, and on the controlling constants $d_0$ and $P_2$ in the doubling and the weak $L^2$ Poincaré inequality.*

**Remark 4.4.1** *The gap in time direction between $Q_-$ and $Q_+$ is necessary. The constant $C_H$ can become infinity when $\rho$ approaches $\epsilon$.*

*Proof of the theorem, $(D) + (WP)$ implies Harnack inequality.*

Here $(D)$ stands for the doubling property and $(WP)$ stands for weak $L^2$ Poincaré inequality.

The proof takes several steps.

*Step 1.* We show that (D) and (WP) imply a mean value inequality for solutions of the heat equation.

From last section, we know that (D) and (WP) yield a Sobolev inequality. From here the Moser's iteration gives us a $L^2$ mean value inequality. The details are given in the next paragraph.

Let $u$ be a positive solution to the heat equation $\Delta u - \partial_t u = 0$ in the region

$$Q_{\sigma r}(x,t) \equiv \{(y,s) \mid y \in M, t-(\sigma r)^2 \le s \le t,\ d(y,x) \le \sigma r\}.$$

Here $r > 0, 2 \ge \sigma \ge 1$. Given any $p \ge 1$, it is clear that

$$\Delta u^p - \partial_t u^p \ge 0. \tag{4.4.1}$$

Let $\phi : [0, \infty) \to [0, 1]$ be a smooth function such that $|\phi'| \le 2/((\sigma-1)r)$, $\phi' \le 0$, $\phi(l) = 1$ when $0 \le l \le r$, $\phi(l) = 0$ when $l \ge \sigma r$. Let $\eta : [0, \infty) \to [0, 1]$ be a smooth function such that $|\eta'| \le 2/((\sigma-1)r)^2$, $\eta' \ge 0$, $\eta \ge 0$, $\eta(s) = 1$ when $t - r^2 \le s \le t$, $\eta(s) = 0$ when $s \le t - (\sigma r)^2$.

Define $\psi = \psi(y, s) = \phi(d(x, y))\eta(s)$. Writing $w = u^p$ and using $w\psi^2$ as a test function on (4.4.1), we deduce

$$\int \nabla(w\psi^2)\nabla w d\mu(y)ds \le -\int (\partial_s w)w\psi^2 d\mu(y)ds. \tag{4.4.2}$$

By direct calculation

$$\int \nabla(w\psi^2)\nabla w d\mu(y)ds = \int |\nabla(w\psi)|^2 d\mu(y)ds - \int |\nabla\psi|^2 w^2 d\mu(y)ds. \tag{4.4.3}$$

Next we estimate the right-hand side of (4.4.2).

$$-\int (\partial_s w)w\psi^2 d\mu(y)ds = \int w^2\psi\partial_s\psi d\mu(y)ds - \frac{1}{2}\int (w\psi)^2|_{s=t} d\mu(y).$$

Combing the last three inequalities, we obtain,

$$\begin{aligned} \int |\nabla(w\psi)|^2 d\mu(y)ds + \frac{1}{2}\int (w\psi)^2|_{s=t} d\mu(y) \\ \le \frac{c}{(\sigma-1)^2r^2}\int_{Q_{\sigma r(x,t)}} w^2 d\mu(y)ds. \end{aligned} \tag{4.4.4}$$

By Hölder's inequality, for any fixed $\nu > 2$,

$$\begin{aligned} \int (\psi w)^{2(1+(2/\nu)} d\mu(y) \quad &\le \quad \left(\int (\psi w)^{2\nu/(\nu-2)} d\mu(y)\right)^{(\nu-2)/\nu} \\ &\times \left(\int (\psi w)^2 d\mu(y)\right)^{2/\nu}. \end{aligned} \tag{4.4.5}$$

From Theorem 4.2.2, the following Sobolev imbedding holds: for a constant $S = S(d_0, P_2)$ and $\nu > 2$,

$$\begin{aligned} \left(\int (\psi w)^{2\nu/(\nu-2)} d\mu(y)\right)^{(\nu-2)/\nu} \le \frac{S\sigma^2r^2}{|B(x,\sigma r)|^{2/\nu}}\int [|\nabla(\psi w)|^2 \\ + r^{-2}(\psi w)^2] d\mu(y). \end{aligned}$$

Therefore we have

$$\left(\int (\psi w)^{2\nu/(\nu-2)} d\mu(y)\right)^{(\nu-2)/\nu} \leq \frac{S\sigma^2 r^2}{|B(x,\sigma r)|^{2/\nu}} \int [|\nabla(\psi w)|^2 + r^{-2}(\psi w)^2] d\mu(y). \quad (4.4.6)$$

for $s \in [t-(\sigma r)^2, t]$ and $\psi w = \psi w(y,s)$. Substituting (4.4.5) and (4.4.6) to (4.4.4), we arrive at the estimate

$$\int_{Q_r(x,t)} w^{2\theta} d\mu(y) ds \leq \frac{S r^2}{|B(x,\sigma r)|^{2/\nu}} \left(\frac{1}{(\sigma-1)^2 r^2} \int_{Q_{\sigma r}(x,t)} w^2 d\mu(y) ds\right)^{\theta}, \quad (4.4.7)$$

with $\theta = 1 + (2/\nu)$. This inequality is often referred to as the reverse Hölder inequality.

Now we apply (4.4.7) with the parameters $\sigma_0 = 2, \sigma_i = 2 - \Sigma_{j=1}^i 2^{-j}$ and $p_i = \theta^i$. Carrying out Moser's iteration and using the volume doubling property, we derive a $L^2$ mean value inequality: for some constant $C = C(S, d_0)$,

$$\sup_{Q_{r/2}(x,t)} u^2 \leq \frac{C(S,d_0)}{r^2 |B(x,r)|} \int_{Q_r(x,t)} u^2 d\mu(y) ds.$$

From a generic trick in [LS] e.g., the $L^2$ mean value inequality implies $L^p$ mean value inequality for any $p > 0$, i.e. for some constant $C = C(S, d_0, p)$

$$\sup_{Q_{r/2}(x,t)} u^p \leq \frac{C(S,d_0,p)}{r^2 |B(x,r)|} \int_{Q_r(x,t)} u^p d\mu(y) ds. \quad (4.4.8)$$

*Step 2.* upper bound for weak $L^1$ norms of $\ln u^{-1}$ and $\ln u$.

Let $u$ be a positive super-solution to the heat equation in the domain $B(x_0, r) \times [t - r^2, t]$, i.e. $\Delta u - \partial_t u \leq 0$. For two numbers $\delta, \rho \in (0,1)$, we write

$$R_+ = B(x_0, \delta r) \times [t_0 - \rho r^2, t_0], \qquad R_- = B(x_0, \delta r) \times [t_0 - r^2, t_0 - \rho r^2].$$

We will show the following:

There exists a positive constant $c_0 = c_0(\delta, \rho, d_0, P_2)$ and a constant $a$, depending on $u$ and given in (4.4.14) below, such that, for all $\lambda > 0$,

$$\begin{aligned} |\{(x,t) \in R_+ \mid \ln u^{-1} > \lambda + a\}| &\le c_0 \lambda^{-1} |B(x_0,r)| r^2, \\ |\{(x,t) \in R_- \mid \ln u > \lambda - a\}| &\le c_0 \lambda^{-1} |B(x_0,r)| r^2. \end{aligned} \tag{4.4.9}$$

We will just give a detailed proof of the first inequality since the second one is similar. Write $w = -\ln u$. Since $u$ is a super-solution, direct calculation shows

$$\Delta w - \partial_t w - |\nabla w|^2 \ge 0. \tag{4.4.10}$$

Consider the function $\lambda : [0,1] \to [0,1]$ defined by

$$\lambda = 1 \quad \text{on} \quad [0,\delta], \;\; \lambda(s) = \frac{1-s}{1-\delta}, \;\; s \in [\delta, 1].$$

Take $\phi = \phi(x) = \lambda(d(x_0,x)/r)$ and use $\phi^2$ as a test function in (4.4.10), we have

$$\begin{aligned} &\frac{d}{dt}\int w\phi^2 d\mu(x) \\ &\qquad \le \int (\Delta w)\phi^2 d\mu(x) - \int |\nabla w|^2 \phi^2 d\mu(x) \\ &\qquad = -2\int \phi \nabla w \nabla \phi d\mu(x) - \int |\nabla w|^2 \phi^2 d\mu(x) \\ &\qquad \le -\frac{1}{2}\int |\nabla w|^2 \phi^2 d\mu(x) + 2\int |\nabla \phi|^2 d\mu(x) \end{aligned}$$

Hence

$$\frac{d}{dt}\int w\phi^2 d\mu(x) + \frac{1}{2}\int |\nabla w|^2 \phi^2 d\mu(x) \le \frac{2}{[(1-\delta)r]^2} \, |B(x_0,r)|.$$

Now we apply, on the above inequality, the following weighted Poincaré inequality

$$\int |w - \bar{w}_\phi|^2 \phi^2 d\mu(x) \le C_\delta(d_0, P_2) r^2 \int |\nabla w|^2 \phi^2 d\mu(x), \tag{4.4.11}$$

where

$$\bar{w}_\phi = \int w\phi^2 d\mu(x) \Big/ \int \phi^2 d\mu(x).$$

The proof of (4.4.11) will be explained as part of Proposition 4.4.1 at the end of this section. We obtain

$$\frac{d}{dt}\int w\phi^2 d\mu(x) + (C_\delta r^2)^{-1}\int |w - \bar{w}_\phi|^2 \phi^2 d\mu(x) \le \frac{2}{[(1-\delta)r]^2} \, |B(x_0,r)|.$$

Since $\int \phi^2 d\mu(x)$ and $|B(x_0, r)|$ are comparable for a fixed $\delta \in (0,1)$, the above implies

$$\frac{d}{dt}\bar{w}_\phi + (C_\delta r^2 |B(x_0,r)|)^{-1} \int |w - \bar{w}_\phi|^2 \phi^2 d\mu(x) \le \frac{C}{[(1-\delta)r]^2}.$$

For simplicity we write

$$V_1 = C_\delta r^2 |B(x_0,r)|,\ V_2 = \frac{C}{[(1-\delta)r]^2},$$

so the above inequality becomes

$$\frac{d}{dt}\bar{w}_\phi + V_1^{-1} \int |w - \bar{w}_\phi|^2 \phi^2 d\mu(x) \le V_2. \tag{4.4.12}$$

Fixing $t_1 = t_0 - \rho r^2$, define

$$w_1(x,t) = w(x,t) - V_2(t - t_1)$$

$$\bar{w}_{\phi,1} = \bar{w}_\phi - V_2(t - t_1).$$

Then they satisfy, via (4.4.12), the inequality

$$\frac{d}{dt}\bar{w}_{\phi,1} + V_1^{-1} \int |w_1 - \bar{w}_{\phi,1}|^2 \phi^2 d\mu(x) \le 0. \tag{4.4.13}$$

Here

$$\bar{w}_{\phi,1} = \int w_1 \phi^2 d\mu(x) / \int \phi^2 d\mu(x)$$

is the $\phi^2$- weighted average of $w_1(\cdot, t)$.

Set

$$a = \bar{w}_{\phi,1}(t_1) \tag{4.4.14}$$

For a given $\lambda > 0$ and $t \in [t_0 - r^2, t_0]$, we identify the regions,

$$D_t^+(\lambda) = \{x \in B(x_0, \delta r) \mid w_1(x,t) > a + \lambda\},$$

$$D_t^-(\lambda) = \{x \in B(x_0, \delta r) \mid w_1(x,t) < a - \lambda\}.$$

We consider two cases. For the first bound in (4.4.9), consider:

*Case 1.* $t > t_1 = t_0 - \rho r^2$ and $x \in D_t^+(\lambda)$.

Then

$$w_1(x,t) - \bar{w}_{\phi,1}(t) \ge a + \lambda - \bar{w}_{\phi,1}(t) \ge \lambda. \tag{4.4.15}$$

Here we just used the definition $a = \bar{w}_{\phi,1}(t_1)$ and the fact that $\bar{w}_{\phi,1}(t)$ is a decreasing function of $t$, by virtue of (4.4.13). Next we reduce the integral in (4.4.13) by integrating only on $D_t^+(\lambda)$. This gives

$$\frac{d}{dt}\bar{w}_{\phi,1}(t) + V_1^{-1}|\lambda + a - \bar{w}_{\phi,1}(t)|^2 \; |D_t^+(\lambda)| \leq 0.$$

Writing

$$f(t) = \bar{w}_{\phi,1}(t) - (\lambda + a),$$

we then turn the above inequality into

$$f'(t) + V_1^{-1}|D_t^+(\lambda)|f^2 \leq 0.$$

As we have seen earlier in applying the Nash inequality, this kind of ordinary differential inequality with a square nonlinearity usually produces useful information. Indeed, solving from $t_1$ to $t_0$, we arrive at

$$V_1(\frac{1}{f(t_0)} - \frac{1}{f(t_1)}) \geq \int_{t_1}^{t_0} |D_t^+(\lambda)|dt.$$

Let us recall that

$$f(t_1) = \bar{w}_{\phi,1}(t_1) - (\lambda + a) = -\lambda,$$

and that $f(t) \leq 0$ when $t \geq t_1$. Therefore the above implies

$$|\{(x,t) \in R_+ \mid w_1(x,t) > \lambda + a\}| \leq V_1\lambda^{-1}.$$

Recalling that

$$w_1 = w - V_2(t - t_1) = -\ln u - V_2(t - t_1),$$

we now deduce, the first bound in (4.4.9)

$$|\{(x,t) \in R_+ \mid \; \ln u^{-1} > \lambda + a\}| \leq c_0\lambda^{-1}|B(x_0, r)|r^2.$$

Here we just used the fact that $V_2(t - t_1) \leq C\rho/(1-\delta)^2$ which is a constant by definition. For the second bound in (4.4.9), consider:

*Case 2.* $t \leq t_1 = t_0 - \rho r^2$ and $x \in D_t^-(\lambda)$.
In this case, we have, just like Case 1,

$$w_1(x,t) - \bar{w}_{\phi,1}(t) \leq a - \lambda - \bar{w}_{\phi,1}(t) \leq -\lambda. \tag{4.4.16}$$

The second inequality in (4.4.9) follows in the same manner. This completes Step 2.

*Step 3.* In this step, we prove the following interpolation result concerning $L^p$ norms of a function.

**Lemma 4.4.1** *Let $R_\sigma$, $\sigma \in (0,1]$ be measurable subset of $M \times \mathbf{R}$ such that $R_{\sigma'} \subset R_\sigma$ if $\sigma' \leq \sigma$. Let $m, k$, and $\delta \in [\frac{1}{2}, 1]$, $p_1 < p_0 \leq \infty$ be positive constants.*

*Suppose $f$ is a positive measurable function satisfying the two assumptions.*

*1. Reverse Hölder inequality.*

$$\|f\|_{p_0, R_{\sigma'}} \leq K \left[(\sigma - \sigma')^{-m} |R_1|^{-1}\right]^{(1/p)-(1/p_0)} \ \|f\|_{p, R_\sigma}$$

*for all $\sigma, \sigma', p$ such that $\frac{1}{2} \leq \delta \leq \sigma' \leq \sigma \leq 1, \quad 0 < p \leq p_1 < p_0$;*

*2. Upper bound for weak $L^1$ norm.*

$$|\{(x,t) \in R_1 \mid \ln f > \lambda\}| \leq K |R_1| \lambda^{-1},$$

*for all $\lambda > 0$.*

*Then there exists a positive constant $\xi$ depending only on $m, \delta$ and the lower bound of $(1/p_1) - (1/p_0)$ such that*

$$\|f\|_{p_0, R_\delta} \leq |R_1|^{1/p_0} e^{\xi(1+K^3)}.$$

Proof. Without loss of generality we take $|R_1| = 1$. Set, for $\sigma \in [\delta, 1]$,

$$\psi = \psi(\sigma) = \ln(\|f\|_{p_0, R_\sigma}). \tag{4.4.17}$$

Fixing $\sigma$, we split $R_\sigma$ into the sets where $\ln f > \psi(\sigma)/2$ or else. By Hölder's inequality, we have

$$\|f\|_{p,R_\sigma} \leq \|f\|_{p_0, R_\sigma} \ |\{(x,t) \in R_\sigma \mid \ln f > \psi(\sigma)/2\}|^{(1/p)-(1/p_0)} + e^{\psi(\sigma)/2}.$$

By Assumption 2, this shows

$$\|f\|_{p, R_\sigma} \leq e^{\psi(\sigma)} \left(\frac{2K}{\psi(\sigma)}\right)^{(1/p)-(1/p_0)} + e^{\psi(\sigma)/2}. \tag{4.4.18}$$

If $\|f\|_{p_0, R_\sigma} \leq e^{2K}$, then there is nothing to prove. So we assume

$$\|f\|_{p_0, R_\sigma} > e^{2K}, \quad i.e. \quad \psi(\sigma) > 2K.$$

In this case, we can choose $p$ (less than $p_0$) such that

$$e^{\psi(\sigma)} \left(\frac{2K}{\psi(\sigma)}\right)^{(1/p)-(1/p_0)} = e^{\psi(\sigma)/2}. \tag{4.4.19}$$

Therefore (4.4.18) becomes

$$\|f\|_{p, R_\sigma} \leq 2 e^{\psi(\sigma)/2}. \tag{4.4.20}$$

Now, from this and Assumption 1, the reverse Hölder inequality, we have

$$\begin{aligned}\psi(\sigma') &= \ln \|f\|_{p_0, R_{\sigma'}} \\ &\le \ln \left( [K(\sigma-\sigma')^{-m}]^{(1/p)-(1/p_0)} \ \|f\|_{p, R_\sigma} \right) \\ &\le \ln \left( 2[K(\sigma-\sigma')^{-m}]^{(1/p)-(1/p_0)} \ e^{\psi(\sigma)/2} \right).\end{aligned}$$

That is

$$\psi(\sigma') \le \frac{1}{2}\psi(\sigma) + \left(\frac{1}{p} - \frac{1}{p_0}\right) \ln[K(\sigma-\sigma')^{-m}] + \ln 2$$

for all $\delta \le \sigma' < \sigma \le 1$. Solving (4.4.19), we know that

$$\frac{1}{p} - \frac{1}{p_0} = \frac{\psi(\sigma)}{2\ln(\psi(\sigma)/(2K)))}.$$

Hence

$$\psi(\sigma') \le \frac{1}{2}\psi(\sigma) + \frac{\psi(\sigma)}{2\ln(\psi(\sigma)/(2K)))} \ln[K(\sigma-\sigma')^{-m}] + \ln 2$$

for all $\delta \le \sigma' < \sigma \le 1$.

Now, if

$$\psi(\sigma) \ge 2K^3(\sigma-\sigma')^{-2m},$$

then the above shows

$$\psi(\sigma') \le \frac{3}{4}\psi(\sigma) + 2.$$

Therefore, we always have, since $\sigma - \sigma' < 1$,

$$\psi(\sigma') \le \frac{3}{4}\psi(\sigma) + 2(K^3+1)(\sigma-\sigma')^{-2m}. \tag{4.4.21}$$

By easy iteration, there exists $c > 0$ such that

$$\ln \|f\|_{p_0, R_\delta} \equiv \psi(\delta) \le c(1-\delta)^{-2m}(1+K^3).$$

This ends the proof of the lemma. □

*Step 4.* We prove the Harnack inequality in this step.

Recall that $u$ is a positive solution in $Q = B(x_0, r) \times [t_0 - r^2, t_0]$. We want to bound the maximum of $u$ in $Q_- = B(x_0, \delta r) \times [t_0 - \eta r^2, t_0 - \rho r^2]$ by its infimum in $Q_+ = B(x_0, \delta r) \times [t_0 - \epsilon r^2, t_0]$. To this end, let us construct two families of parabolic cubes contained in $Q$ such that one family is an expansion of $Q_+$ while the other is an expansion of $Q_-$. More precisely, we let, for $\sigma \in [\delta, 1]$,

$$Q_{+,\sigma} = B(x_0, \sigma r) \times [t_0 - l_1(\sigma)\epsilon r^2, t_0],$$

$$Q_{-,\sigma} = B(x_0, \sigma r) \times [t_0 - l_2(\sigma)\eta r^2, t_0 - \rho r^2].$$

Here $l_1$ is a linear function such that $l_1(\delta) = 1$ and $l_1(1)\epsilon = \rho$; and $l_2$ is a linear function such that $l_2(\delta) = 1$ and $l_2(1)\eta = 1$.

Let $a$ be the constant in (4.4.9) in Step 2. Then from (4.4.9), we know that the function $f = e^a u$ satisfies the weak $L^1$ bound in $Q_{-,\sigma}$. It also satisfies the reverse Hölder inequality (4.4.7). Therefore, Lemma 4.4.1 shows that, for some $p_0 > 0$,

$$e^a \|u\|_{p_0, Q_{-,(1+\delta)/2}} \leq [r^2 |B(x_0, r)|]^{1/p_0} e^{\xi(1+K^3)}.$$

Similarly, one can apply Lemma 4.4.1 for the function $e^{-a}u^{-1}$ in $Q_{+,\sigma}$, with $p_0$ chosen as infinity. One concludes

$$e^{-a} \sup_{Q_+} u^{-1} \leq e^{\xi(1+K^3)}.$$

Next, combining these two inequalities with the mean value inequality in Step 1 (applied on the cube $Q_{-,(1+\delta)/2}$), we deduce

$$\sup_{Q_-} u \leq C e^{2\xi(1+K^3)} \inf_{Q_+} u.$$

So we have proven the Harnack inequality except for the weighted Poincaré inequality (4.4.11) which will be done in Proposition 4.4.1 after the proof of the theorem.

*Proof of the theorem, Harnack inequality implies* $(D) + (WP)$, i.e. doubling condition and weak $L^2$ Poincaré inequality.

First we show that the Harnack inequality implies the doubling condition of geodesic balls. Fixing $x \in M$, let $G = G(x, t, y)$ be the heat kernel, i.e. the fundamental solution of the heat equation on $M$. Picking $r > 0$ and $y \in B(x, r)$, the Harnack inequality applied on cubes of size $r$ shows

$$G(x, r^2, x) \leq C G(x, 2r^2, y).$$

Integrating this over $B(x, r)$ with respect to $y$, we obtain

$$|B(x, r)| G(x, r^2, x) \leq C \int G(x, 2r^2, y) d\mu(y) = C.$$

i.e.

$$G(x, r^2, x) \le \frac{C}{|B(x,r)|}. \tag{4.4.22}$$

Actually, a parallel lower bound for $G$ also holds. Here is a proof. Consider the function

$$u = u(z,s) = \int_{B(x,r)} G(z,s,y) d\mu(y).$$

This is a solution to the heat equation and $u(z,0) = 1$ for all $z \in B(x,r)$. So we can extend $u$ even when $s < 0$ by setting $u(z,s) = 1$, $z \in B(x,r)$, $s < 0$. Now this extended function $u$ is a positive solution to the heat equation in $B(x,r) \times (-\infty, \infty)$. Applying the Harnack inequality twice on cubes of size comparable to $r$ and suitable vertex, we find

$$\begin{aligned} 1 = u(x, -r^2/4) &\le C u(x, r^2/2) = C \int_{B(x,r)} G(x, r^2/2, y) d\mu(y) \\ &\le C^2 \int_{B(x,r)} G(x, r^2, x) d\mu(y) = C^2 |B(x,r)| \, G(x, r^2, x). \end{aligned}$$

This shows, together with (4.4.22), the so-called on-diagonal lower and upper bound for the heat kernel:

$$\frac{1}{C^2 |B(x,r)|} \le G(x, r^2, x) \le \frac{C}{|B(x,r)|}. \tag{4.4.23}$$

Here $C$ is the constant in the Harnack inequality. By this and the Harnack inequality again

$$\frac{1}{C^2 |B(x,r)|} \le G(x, r^2, x) \le C G(x, 4r^2, x) \le \frac{C^2}{|B(x,2r)|}.$$

Hence

$$|B(x,2r)| \le C^4 |B(x,r)|.$$

This is the doubling condition.

Finally we will prove the weak Poincaré inequality by the Harnack inequality. We will use the method in [KS] as presented in [Sal2].

Pick a geodesic ball $B(x,r)$. Let $P = P(y,t,z)$ be the heat kernel with Neumann boundary condition on $B(x,r)$. By Harnack inequality, as in the proof of (4.4.23), we know that

$$P(y, r^2, z) \ge \frac{C}{|B(x,r)|}$$

for all $y, z$ in the ball of half the size: $B(x, r/2)$. Now we pick a smooth function $f$ in $B(x, r)$. Define

$$u(y,t) = (P * f)(y,t) = \int_{B(x,r)} P(y,t,z)f(z)d\mu(z),$$

which is a solution to the heat equation in $B(x, r)$ with Neumann boundary condition. By the lower bound for $P$, we have

$$[P * (f - u(y,r^2))^2](y,r^2) \geq \frac{C}{|B(x,r)|} \int_{B(x,r/2)} |f(z) - u(y,r^2)|^2 d\mu(z)$$

Let $f_{B(x,r/2)}$ be the average of $f$ in $B(x, r/2)$, then it immediately implies

$$[P * (f - u(y,r^2))^2](y,r^2) \geq \frac{C}{|B(x,r)|} \int_{B(x,r/2)} |f(z) - f_{B(x,r/2)}|^2 d\mu(z).$$

Integrating over $B(x, r/2)$ and using the doubling property of geodesic balls just proven, we deduce

$$\begin{aligned} &\int_{B(x,r/2)} [P * (f - u(y,r^2))^2](y,r^2) d\mu(y) \\ &\qquad \geq C \int_{B(x,r/2)} |f(z) - f_{B(x,r/2)}|^2 d\mu(z). \end{aligned} \tag{4.4.24}$$

Next we bound the left-hand side of the above inequality.

$$\begin{aligned} &\int_{B(x,r/2)} [P * (f - u(y,r^2))^2](y,r^2) d\mu(y) \\ &\qquad = \int_{B(x,r/2)} \int_{B(x,r)} P(y,r^2,z)(f(z) - u(y,r^2))^2 d\mu(z) d\mu(y) \\ &\qquad \leq \int_{B(x,r)} \int_{B(x,r)} P(y,r^2,z)(f^2(z) - 2f(z)u(y,r^2) \\ &\qquad + u^2(y,r^2)) d\mu(z) d\mu(y). \end{aligned}$$

Using the fact that $\int_{B(x,r)} P(y, r^2, z)d\mu(z) = 1$ and $\int_{B(x,r)} P(y, r^2, z)f(z)d\mu(z) = u(y, r^2)$, we can convert the above inequality into

$$\begin{aligned}
&\int_{B(x,r/2)} [P*(f-u(y,r^2))^2](y,r^2)d\mu(y) \\
&\qquad \le \int_{B(x,r)} f^2(z)d\mu(z) - \int_{B(x,r)} u^2(z,r^2)d\mu(z) \\
&\qquad = -\int_0^{r^2} \partial_s \int_{B(x,r)} u^2(z,s)d\mu(z)ds \\
&\qquad = 2\int_0^{r^2} \int_{B(x,r)} |\nabla u(z,s)|^2 d\mu(z)ds \qquad \text{via integration by parts.}
\end{aligned}$$

Note that

$$\begin{aligned}
&\partial_s \int_{B(x,r)} |\nabla u(z,s)|^2 d\mu(z) \\
&\quad = 2\int_{B(x,r)} \nabla u(z,s)\nabla\Delta u(z,s)d\mu(z) = -2\int_{B(x,r)} |\Delta u(z,s)|^2 d\mu(z) \le 0.
\end{aligned}$$

Therefore

$$\int_{B(x,r/2)} [P*(f-u(y,r^2))^2](y,r^2)d\mu(y) \le 2r^2 \int_{B(x,r)} |\nabla f(z)|^2 d\mu(z).$$

Substituting this to (4.4.24), we arrive at

$$\int_{B(x,r/2)} |f(z) - f_{B(x,r/2)}|^2 d\mu(z) \le Cr^2 \int_{B(x,r)} |\nabla f(z)|^2 d\mu(z),$$

which is the desired weak Poincaré inequality. □

The next proposition shows that doubling condition and weak $L^2$ Poincaré inequality imply certain weighted $L^2$ Poincaré inequality, which is crucial to the proof of the parabolic Harnack inequality in the last theorem.

For simplicity, we choose a weight function which is radial relative to a fixed point. For more general weights see Theorem 5.3.4 in [Sal].

**Proposition 4.4.1** *Let* $\lambda : [0,1] \to [0,1]$ *be defined by*

$$\lambda = 1 \quad on \quad [0,\delta],\ \lambda(s) = \left(\frac{1-s}{1-\delta}\right)^m,\ s \in [\delta, 1].$$

*Here* $m > 0$ *is a constant. For* $x_0, x \in M$ *and* $r > 0$*, define* $\phi = \phi(x) = \lambda(d(x_0,x)/r)$*. Suppose the doubling condition (4.2.15) and the following weak* $L^2$ *Poincaré inequality with parameter* $\kappa$ *hold in* $M$*:*

*there exist constants* $\kappa > 1$ *and* $P_\kappa > 0$ *such that for any* $f \in C^\infty(B(x,kr))$,

$$\int_{B(x,r)} |f(y) - f_{B(x,r)}|^2 \, d\mu(y) \le P_\kappa \, r^2 \int_{B(x,\kappa r)} |\nabla f(y)|^2 \, d\mu(y),$$
$$x \in M, r > 0. \qquad (4.4.25)$$

*Here*

$$f_{B(x,r)} = |B(x,r)|^{-1} \int_{B(x,r)} f(y) d\mu(y).$$

*Then for any* $f \in C^\infty(B(x_0,r))$, *there exists* $C = C(\delta, m, d_0, P_\kappa)$ *such that*

$$\int |f - \bar{f}_\phi|^2 \phi d\mu \le Cr^2 \int |\nabla f|^2 \phi d\mu,$$

*where*

$$\bar{f}_\phi = \int f\phi d\mu / \int \phi d\mu.$$

Proof. The idea of the proof, according to D. Jerison [Je], is to use a Whitney type covering and to decompose the integrals into the sum of integrals over the balls in the covering. We will just present the key part and leave many details as exercises. A complete proof can be found in Section 5.3 of [Sal]. We divide the proof into several steps.

*Step 1.* In this step, we show that if the weak $L^2$ Poincaré inequality (4.4.25) with parameter $\kappa$ holds then it actually holds when $\kappa$ is replaced by any constant $\tau > 1$. Consequently we will always take $\kappa = 2$ in (4.4.25) for the rest of the proof, i.e. we will always assume (4.2.17) holds.

The statement when $\tau \ge \kappa$ is trivially true. The case when $\tau \in (1, \kappa)$ is left as an exercise.

**Exercise 4.4.1** *Prove the following result:*

*Suppose the doubling condition (4.2.15) and weak* $L^2$ *Poincaré inequality (4.4.25) with parameter* $\kappa$ *hold in* $M$. *Then for any* $\tau \in (1, \kappa]$, *there exist positive constant* $c = c(\tau, d_0, P_\kappa)$ *such that*

$$\int_{B(x,r)} |f(y) - f_{B(x,r)}(x)|^2 \, d\mu(y) \le c \, r^2 \int_{B(x,\tau r)} |\nabla f(y)|^2 \, d\mu(y)$$

*for all* $f \in C^\infty(B(x, \tau r))$, $r > 0$.

The idea of the proof is to use Vitali type covering. See Lemma 5.3.1 in [Sal].

*Step 2.* Whitney type covering.

Fixing the ball $E \equiv B(x_0, r)$, one can show that there exists a collection $\mathbf{F}$ of balls $B$ with the following properties:

(1) The balls in $\mathbf{F}$ are disjoint.

(2) $E \subset \cup_{B \in \mathbf{F}} 2B$. Here and later, for a positive number $\lambda$, the notation $\lambda B$ stands for the ball at the same center as $B$ but with $\lambda$ times the radius of $B$.

(3) For any ball $B \in \mathbf{F}$, the radius of $B$, called $r(B)$, satisfies

$$r(B) = 10^{-3} d(B, \partial E), \qquad 10^3 B \subset E.$$

(4) There exists a positive constant $K$ depending only on $d_0$ such that

$$\sup_{x \in E} Cardinal\{B \in \mathbf{F} \,|\, x \in 100B\} \le K.$$

**Exercise 4.4.2** *Prove that the above Whitney type covering exists.*

*Step 3.* The concepts of central balls and subcollection $F(B)$.

Let $\mathbf{F}$ be a covering of $E$, which satisfies properties (1)–(4) in Step 2. Then there exists a ball $B^0 \in \mathbf{F}$ such that $2B^0$ contains the point $x_0$, the center of $E$. We call such a ball $B^0$ a central ball of $\mathbf{F}$ and use $x_B$ to denote the center of $B^0$.

Given $B \in \mathbf{F}$ with center $x_B$, let $\gamma_B$ be a minimum geodesic connecting $x_0$ and $x_B$. Then we can find a finite subcollection of $\mathbf{F}$:

$$F(B) = \{B^0, \ldots, B^{l(B)}\}$$

such that $B^{l(B)} = B$ and

$$2\bar{B}^i \cap 2\bar{B}^{i+1} \ne empty, \quad 2B^i \cap \gamma_B \ne empty, \quad i = 0, \ldots, l(B).$$

**Exercise 4.4.3** *Prove that the following properties hold:*

*(i) For any $B \in \mathbf{F}$,*

$$d(\gamma_B, \partial E) \ge \frac{1}{2} d(B, \partial E) = 500 r(B).$$

*(ii) For any $B \in \mathbf{F}$ and any two consecutive balls $B^i, B^{i+1}$ in $F(B)$,*

$$1.01^{-1} r(B^i) \le r(B^{i+1}) \le 1.01 r(B^i),$$

$$B^{i+1} \subset 4B^i,$$

$$|4B^i \cap 4B^{i+1}| \ge c \max\{|B^i|, |B^{i+1}|\}.$$

*(iii) For any $B \in \mathbf{F}$ and ball $A \in F(B)$, $B \subset 10^4 A$.*

All these properties are obviously true in $\mathbf{R^n}$. The point is that they hold in the general setting of metric spaces where only the volume doubling condition is assumed. Detailed proofs can be found in Lemmas 5.3.6, 5.3.7 and 5.3.8 of [Sal].

Using this exercise we can prove the following claim which controls the difference of the average of $f$ over consecutive balls in $F(B)$. Here we use the notation $f_B$ to denote the average of $f$ in the ball $B$.

*Claim: Under the assumptions of the proposition, there exists a positive constant $C$ such that for any ball $B \in \mathbf{F}$ and any consecutive balls $B^i, B^{i+1} \in F(B)$,*

$$|f_{4B^i} - f_{4B^{i+1}}| \le C \frac{r(B^i)}{|B^i|^{1/2}} \left( \int_{32B^i} |\nabla f|^2 d\mu \right)^{1/2}.$$

The proof goes like this.

$$\begin{aligned}
&|4B^i \cap 4B^{i+1}|^{1/2} |f_{4B^i} - f_{4B^{i+1}}| \\
&\quad = \left( \int_{4B^i \cap 4B^{i+1}} |f_{4B^i} - f_{4B^{i+1}}|^2 d\mu \right)^{1/2} \\
&\quad \le \left( \int_{4B^i \cap 4B^{i+1}} |f - f_{4B^i}|^2 d\mu \right)^{1/2} + \left( \int_{4B^i \cap 4B^{i+1}} |f - f_{4B^{i+1}}|^2 d\mu \right)^{1/2} \\
&\quad \le \left( \int_{4B^i} |f - f_{4B^i}|^2 d\mu \right)^{1/2} + \left( \int_{4B^{i+1}} |f - f_{4B^{i+1}}|^2 d\mu \right)^{1/2} \\
&\quad \le C r(B^i) \left( \int_{8B^i} |\nabla f|^2 d\mu \right)^{1/2} + C r(B^{i+1}) \left( \int_{8B^{i+1}} |\nabla f|^2 d\mu \right)^{1/2}.
\end{aligned}$$

Here the last step is due to the weak Poincaré inequality (4.2.17). The claim follows by (ii) of the previous exercise.

*Step 4.* Let $\mathbf{F}$ be a Whitney type covering for $E = B(x_0, r)$, which was constructed in Step 2. Since $E \subset \cup_{B \in \mathbf{F}} 2B$ and $supp\, \phi \subset E$ by construction, we can use Minkowski inequality to deduce

$$\begin{aligned}
\int |f - f_{4B^0}|^2 \phi d\mu &\le \Sigma_{B \in \mathbf{F}} \int_{2B} |f - f_{4B^0}|^2 \phi d\mu \\
&\le 2^2 \Sigma_{B \in \mathbf{F}} \int_{4B} (|f - f_{4B}|^2 + |f_{4B} - f_{4B^0}|^2) \phi d\mu.
\end{aligned}$$

Here and later in the proof, $B^0$ stands for a central ball of the covering. Therefore

$$\int |f - f_{4B^0}|^2 \phi d\mu \le 2^2 \Sigma_{B\in\mathbf{F}} \int_{4B} |f - f_{4B}|^2 \phi d\mu + 2^2 \Sigma_{B\in\mathbf{F}} |f_{4B} - f_{4B^0}|^2 \phi(4B) \equiv T_1 + T_2 \tag{4.4.26}$$

where and later

$$\phi(S) \equiv \int_S \phi d\mu, \quad S \subset M.$$

We can bound $T_1$ in the following manner. By property (3) in Step 2, the weight $\phi$ satisfies

$$\sup_{x\in B} \phi \le C \inf_{x\in B} \phi, \quad B \in \mathbf{F}.$$

Hence the weak Poincaré inequality (4.2.17) implies

$$\int_{4B} |f - f_{4B}|^2 \phi d\mu \le C P_2 r(4B)^2 \int_{8B} |\nabla f|^2 \phi d\mu.$$

Note, by construction, that $8B \subset E$ and the overlapping number for $\{8B \mid B \in \mathbf{F}\}$ is bounded. Thus

$$T_1 \le C P_2 \Sigma_{B\in\mathbf{F}} r(4B)^2 \int_{8B} |\nabla f|^2 \phi d\mu \le C P_2 r^2 \int_E |\nabla f|^2 \phi d\mu. \tag{4.4.27}$$

It remains to bound $T_2$. By the doubling condition (4.2.15) and the definition of $\phi$, there exists $C_1 = C_1(\phi, d_0)$ such that

$$T_2 = 2^2 \Sigma_{B\in\mathbf{F}} |f_{4B} - f_{4B^0}|^2 \phi(4B) \le C_1 \Sigma_{B\in\mathbf{F}} \int |f_{4B} - f_{4B^0}|^2 \frac{\phi(B)}{|B|} \chi_B d\mu. \tag{4.4.28}$$

Fixing $B \in \mathbf{F}$, let $F(B) = \{B^0, \ldots, B^{l(B)}\}$ be the subcollection defined in the previous step with $B^0$ being a central ball and $B^{l(B)} = B$. Then

$$|f_{4B} - f_{4B^0}| \left(\frac{\phi(B)}{|B|}\right)^{1/2} \le \Sigma_{i=0}^{l(B)-1} |f_{4B^i} - f_{4B^{i+1}}| \left(\frac{\phi(B)}{|B|}\right)^{1/2}.$$

By the claim in the previous step and the easily verified fact that

$$\frac{\phi(B)}{|B|} \le C_2 \phi(x), \quad x \in 32B^i, \quad i = 0, \ldots, l(B),$$

we deduce

$$|f_{4B} - f_{4B^0}| \left(\frac{\phi(B)}{|B|}\right)^{1/2} \le C_3 \Sigma_{i=0}^{l(B)-1} \frac{r(B^i)}{|B^i|^{1/2}} \left(\int_{32B^i} |\nabla f|^2 \phi d\mu\right)^{1/2}.$$

From property (iii) in Step 3, the ball $B$ is contained in $10^4 B^i$ for any $B^i \in F(B)$. Therefore

$$|f_{4B} - f_{4B^0}| \left(\frac{\phi(B)}{|B|}\right)^{1/2} \chi_B \le C_3 \Sigma_{A\in\mathbf{F}} \frac{r(A)}{|A|^{1/2}} \left(\int_{32A} |\nabla f|^2 \phi d\mu\right)^{1/2} \times \chi_{10^4 A}\, \chi_B$$

which implies, since the balls $B$ in $\mathbf{F}$ are disjoint,

$$\begin{aligned}&\Sigma_{B\in\mathbf{F}} |f_{4B} - f_{4B^0}|^2 \frac{\phi(B)}{|B|} \chi_B \\ &\qquad \le C_3 \left[\Sigma_{A\in\mathbf{F}} \frac{r(A)}{|A|^{1/2}} \left(\int_{32A} |\nabla f|^2 \phi d\mu\right)^{1/2} \chi_{10^4 A}\right]^2.\end{aligned}$$

Plugging this into (4.4.28), we know that

$$\begin{aligned} T_2 &\le C_1 \int \Sigma_{B\in\mathbf{F}} |f_{4B} - f_{4B^0}|^2 \frac{\phi(B)}{|B|} \chi_B d\mu \\ &\le C_4 \int \left[\Sigma_{A\in\mathbf{F}} \frac{r(A)}{|A|^{1/2}} \left(\int_{32A} |\nabla f|^2 \phi d\mu\right)^{1/2} \chi_{10^4 A}\right]^2 d\mu \\ &\equiv C_4 \int \left[\Sigma_{A\in\mathbf{F}} J_A\, \chi_{10^4 A}\right]^2 d\mu \end{aligned}$$

where, for simplicity, we have used the notation

$$J_A \equiv \frac{r(A)}{|A|^{1/2}} \left(\int_{32A} |\nabla f|^2 \phi d\mu\right)^{1/2}. \tag{4.4.29}$$

Now we need to remove the factor $10^4$ in $\chi_{10^4 A}$.

Note

$$\begin{aligned} T_2^{1/2} &\le C_4^{1/2} \sup_{\|\rho\|_2=1} \int \Sigma_{A\in\mathbf{F}} J_A\, \chi_{10^4 A} |\rho| d\mu = C_4^{1/2} \sup_{\|\rho\|_2=1} \Sigma_{A\in\mathbf{F}} J_A \\ &\qquad \times \int_{10^4 A} |\rho| d\mu \\ &= C_4^{1/2} \sup_{\|\rho\|_2=1} \Sigma_{A\in\mathbf{F}} J_A |10^4 A| \frac{1}{|10^4 A|} \int_{10^4 A} |\rho| d\mu \\ &\le C_5 \sup_{\|\rho\|_2=1} \Sigma_{A\in\mathbf{F}} J_A |A| \frac{1}{|10^4 A|} \int_{10^4 A} |\rho| d\mu \end{aligned}$$

where we just used the doubling condition again. Observe, for each $x \in A$,

$$\frac{1}{|10^4 A|}\int_{10^4 A} |\rho| d\mu \le \frac{1}{|10^4 A|}\int_{B(x, 10^5 r(A))} |\rho| d\mu \le C M\rho(x)$$

where $M\rho$ is the standard maximal function of $\rho$, i.e.

$$M\rho(x) = \sup_{r>0} \frac{1}{|B(x,r)|}\int_{B(x,r)} |\rho(y)| d\mu(y).$$

Therefore

$$\frac{1}{|10^4 A|}\int_{10^4 A} |\rho| d\mu \le \frac{C}{|A|}\int_A M\rho(x) d\mu,$$

which shows

$$\begin{aligned} T_2^{1/2} &\le C C_5 \sup_{\|\rho\|_2=1} \Sigma_{A\in\mathbf{F}} J_A \int_A M\rho(x) d\mu \\ &= C C_5 \sup_{\|\rho\|_2=1} \int \Sigma_{A\in\mathbf{F}} J_A \chi_A M\rho(x) d\mu \\ &\le C C_5 \sup_{\|\rho\|_2=1} \|\Sigma_{A\in\mathbf{F}} J_A \chi_A\|_2 \, \|M\rho\|_2. \end{aligned}$$

Hence

$$\begin{aligned} T_2 &\le C_6 \sup_{\|\rho\|_2=1} \Sigma_{A\in\mathbf{F}} \|J_A \chi_A\|_2^2 \, \|\rho\|_2^2 \\ &= C_6 \Sigma_{A\in\mathbf{F}} J_A^2 |A|. \end{aligned} \tag{4.4.30}$$

Here we have used the fact that $A \in \mathbf{F}$ are disjoint and the following well-known property for maximal functions:

*Let $M$ be a complete metric space satisfying the volume doubling condition: $|B(x,2r)| \le d_0 |B(x,r)|$ for all $x \in M$ and $r > 0$. Then for all $f \in C_0^\infty(M)$, there exists a positive constant $c = c(d_0, p)$ such that*

$$\|Mf\|_p \le c\|f\|_p, \quad 1 < p \le \infty.$$

**Exercise 4.4.4** *Prove the above statement.*

By (4.4.29) and (4.4.30), we arrive at

$$T_2 \le C_6 r^2 \Sigma_{A\in\mathbf{F}} \int_{32A} |\nabla f|^2 \phi d\mu \le C_7 r^2 \int_E |\nabla f|^2 \phi d\mu.$$

Here we have used properties (3) and (4) in Step 2.

Combining the last inequality with (4.4.27) and (4.4.26), we know that

$$\int |f - f_{4B^0}|^2 \phi d\mu \le C_8 r^2 \int_E |\nabla f|^2 \phi d\mu.$$

Finally

$$\int |f - \bar{f}_\phi|^2 \phi d\mu \le \int |f - f_{4B^0}|^2 \phi d\mu \le C_8 r^2 \int |\nabla f|^2 \phi d\mu.$$

which proves the proposition. □

**Remark 4.4.2** *There is also $L^p$ version of this result. The proof is identical.*

**Exercise 4.4.5** *State and prove a $L^p$ version of the proposition.*

## 4.5 Maximum principle for parabolic equations

The maximum principle plays an important role in the study of elliptic and parabolic equations. Let us first prove a basic maximum principle for the heat equation on noncompact manifolds. It first appeared in Karp-Li [KaLi] and Grigor'yan [Gr2].

**Theorem 4.5.1** *Let $u$ be a smooth subsolution to the heat equation on $\mathbf{M} \times [0, T)$, i.e. $\Delta u - \partial_t u \ge 0$. Here $\mathbf{M}$ is a noncompact Riemann manifold without boundary and $T > 0$. Suppose*

$$\int_0^T \int_{\mathbf{M}} e^{-\alpha d^2(x,0)} u^2(x,t) d\mu(x) dt < \infty$$

*for some $\alpha > 0$. Here $0 \in \mathbf{M}$ and $d(x, 0)$ is the Riemann distance between $x$ and $0$. Then $u \le 0$ on $\mathbf{M} \times [0, T)$ provided that $u(x, 0) \le 0$.*

Proof. Define

$$h(x,t) = -\frac{d^2(x,0)}{4(2\tau - t)}$$

for some $\tau > T$. It is easy to check that

$$|\nabla h|^2 + \partial_t h = 0, \qquad a.e.$$

Let $\phi_s(\cdot)$ be a cut-off function such that $0 \le \phi_s \le 1$; $\phi_s(x) = 1$, $x \in B(0, s)$, *supp* $\phi \subset B(0, s+1)$; and $|\nabla \phi_s| \le 2$.

Using $\phi_s^2 e^h u_+$ as a test function on $\Delta u - \partial_t u \geq 0$, we obtain

$$\int_0^T \int_{\mathbf{M}} e^h (\phi_s^2 \ |\nabla u_+|^2 + 2 < \nabla \phi_s, \nabla u_+ > (\phi_s u_+) \\ + (\phi_s^2 u_+) < \nabla h, \nabla u_+ >) d\mu dt \\ + \frac{1}{2} \int_0^T \int_{\mathbf{M}} \phi_s^2 e^h \partial_t u_+^2 d\mu dt \leq 0.$$

Using Cauchy-Schwarz inequality on the second and third term of the above inequality, we deduce

$$\int_0^T \int_{\mathbf{M}} e^h (-2|\nabla \phi_s^2| u_+^2 - \frac{1}{2} \phi_s^2 u_+^2 \ |\nabla h|^2) d\mu dt \\ - \frac{1}{2} \int_0^T \int_{\mathbf{M}} \phi_s^2 e^h (u_+^2) \partial_t h d\mu dt + \frac{1}{2} \int_{\mathbf{M}} \phi_s^2 e^h u_+^2 |_0^T \leq 0.$$

This shows

$$\int_{\mathbf{M}} \phi_s^2 e^h u_+^2 (x, T) d\mu \leq 4 \int_0^T \int_{\mathbf{M}} e^h u_+^2 |\nabla \phi_s|^2 d\mu dt.$$

If $T$ is sufficiently small we can choose $2\tau - T$ sufficiently small so that the above integrations are finite. Note that the integral on the right-hand side takes place on $[B(0, s+1) - B(0, s)] \times [0, T]$. Letting $s \to \infty$, we conclude that $u_+ = 0$. For an arbitrary $T$ we just repeat the process. □

**Remark 4.5.1** *The growth condition on solutions is necessary in general. A classical counterexample due to Tychonov is the following.*

*Let*

$$u(x, t) = \Sigma_{i=0}^{\infty} f^{(i)}(t) \frac{x^{2i}}{(2i)!},$$

*where* $x \in \mathbf{R}$, $t \geq 0$, *and*

$$f(t) = \begin{cases} e^{-t^{-2}}, & t > 0, \\ 0, & t = 0. \end{cases}$$

*Then, it is easy to check that* $u$ *is a nontrivial solution to the heat equation in* $\mathbf{R} \times (\mathbf{0}, \infty)$. *However* $u(x, 0) = 0$.

**Exercise 4.5.1** *Verify the above* $u$ *is a nontrivial solution of the heat equation.*

The next result is a maximum principle for tensor that is often referred to as Hamilton's weak maximum principle. Stronger versions of this result are also found by Hamilton. Some of these will be discussed in Section 5.2 below.

Let $\mathbf{M}$ be a compact, $n$ dimensional manifold, and $g = g(t)$ be a family of smooth metrics on $\mathbf{M} \times [0, T]$. Let $V$ be a vector bundle over $\mathbf{M}$ with a time independent metric $h = h_{\alpha\beta}$, and a connection $\nabla(t) = \{\Gamma^\alpha_{i\beta}\}$ which is compatible with $h_{\alpha\beta}$. This means $\nabla_X h = 0$ for any tangent vector $X$ on $\mathbf{M}$. Let $\sigma$ be a $C^\infty$ section of $V$ over $\mathbf{M}$. Define the Laplacian by

$$\Delta\sigma = g^{ij}(x,t)\nabla_i\nabla_j\sigma$$

where $\nabla_i\nabla_j\sigma \equiv \nabla^2_{ij}\sigma$, the second covariant derivative of $\sigma$. This Laplacian can be regarded as a version of the rough Laplacian on $(2,0)$ tensors associated with $g(t)$ and the connection $\nabla(t)$. The connection and Laplacian can be extended in the usual way to act on any smooth tensor. We still use the same notation for their extensions.

**Theorem 4.5.2** *(Hamilton's weak maximum principle for tensor)* *Let $M_{\alpha\beta}$ be a family of smooth, symmetric bilinear forms on $V$ and*

$$N_{\alpha\beta} = N_{\alpha\beta}(J, h)$$

*be a polynomial of $J \equiv (M_{\eta\xi})$, formed by contracting elements of $J$ with the metric $h = (h_{\eta\xi})$. Here $\eta$ and $\xi$ inside $J$, $h$ are dummy indices.*

*Assume, for any $x \in \mathbf{M}$ and $v \in V_x$, the fiber of the bundle $V$ at $x$, there holds $N_{\alpha\beta}v^\alpha v^\beta \ge 0$ whenever $M_{\alpha\beta}v^\alpha = 0$. Here $M_{\alpha\beta}$ is any smooth bilinear form on $V$.*

*Suppose $M_{\alpha\beta}$ evolves by the equation*

$$\partial_t M_{\alpha\beta} = \Delta M_{\alpha\beta} + u^i\nabla_i M_{\alpha\beta} + N_{\alpha\beta}, \quad \text{on} \quad \mathbf{M} \times [0, T]. \tag{4.5.1}$$

*Here $\Delta$ is the rough Laplacian on tensor associated with $g(t)$ and the connection $\nabla(t)$, and $u^i$ is a bounded vector field.*

*Then $(M_{\alpha\beta}) \ge 0$ for all $t \in (0, T]$ if $(M_{\alpha\beta}) \ge 0$ at $t = 0$.*

Proof. Let $\epsilon > 0$ and $A > 0$ be two constants to be chosen later. Consider the bilinear form:

$$\tilde{M}_{\alpha\beta} = M_{\alpha\beta} + \epsilon e^{tA} h_{\alpha\beta}.$$

Then $(\tilde{M}_{\alpha\beta}) \ge 0$ at $t=0$, due to the assumption that $(M_{\alpha\beta}) \ge 0$ at $t=0$. We claim that for a fixed large $A$ and all sufficiently small $\epsilon$, it holds $(\tilde{M}_{\alpha\beta}) > 0$ for all $t \in (0,T]$.

Suppose the claim is not true. Then there exists arbitrarily small $\epsilon > 0$, some point $x_0 \in \mathbf{M}$ and some unit vector $v \in V_{x_0}$ such that

$$\tilde{M}_{\alpha\beta}(x_0,t_0)v^\alpha = 0.$$

Here unit vector means relative to the metric $h$. We also chose $t_0$ to be the first time this can happen. Under the metric $g(t_0)$, we use parallel translation along geodesics emanating from $x_0$ to extend $v$ to a smooth vector field in a neighborhood of $x_0$. We still denote this vector field by $v$. Write

$$F(x,t) = \tilde{M}_{\alpha\beta}(x,t)v^\alpha v^\beta = M_{\alpha\beta}v^\alpha v^\beta + \epsilon e^{At} h_{\alpha\beta} v^\alpha v^\beta = M_{\alpha\beta}v^\alpha v^\beta + \epsilon e^{At}.$$

By our choice, $F(x_0,t_0)=0$ and $F(x,t) \ge 0$ when $t < t_0$ and $F(x,t_0) \ge 0$, $x \in M$. Hence

$$\partial_t F \le 0, \quad \Delta F \ge 0, \quad \text{at} \quad (x_0,t_0).$$

Here $\Delta$ is the Laplace-Beltrami operator on $\mathbf{M}$, relative to the metric $g$.

Therefore, at the space time point $(x_0,t_0)$,

$$\begin{aligned} 0 \ge \partial_t F &= \partial_t (M_{\alpha\beta}v^\alpha v^\beta + \epsilon e^{At}) \\ &= \Delta(\tilde{M}_{\alpha\beta}v^\alpha v^\beta) + u^i \nabla_i (\tilde{M}_{\alpha\beta}v^\alpha v^\beta) + N_{\alpha\beta}(J,h)v^\alpha v^\beta + \epsilon A e^{At_0} \\ &\ge N_{\alpha\beta}(J,h)v^\alpha v^\beta + \epsilon A e^{At_0} \end{aligned} \tag{4.5.2}$$

Here we have used the fact that $v$ is parallel and $\nabla_i(\tilde{M}_{\alpha\beta}v^\alpha v^\beta) = \nabla_i F = 0$ at $(x_0,t_0)$.

Notice that, for $\tilde{J} \equiv J + \epsilon e^{At} h = (M_{\alpha\beta}) + \epsilon e^{At}(h_{\alpha\beta})$,

$$\begin{aligned} N_{\alpha\beta}(J,h)v^a v^\beta &= N_{\alpha\beta}(\tilde{J} - \epsilon e^{At}h, h)v^a v^\beta \\ &\ge N_{\alpha\beta}(\tilde{J},h)v^a v^\beta - C\Sigma_{k=1}^m \epsilon^k e^{kAt}, \end{aligned}$$

where $m$ is the highest order of the polynomials in $(N_{\alpha\beta})$. Since $\tilde{M}_{\alpha\beta}v^\alpha = 0$ at $(x_0,t_0)$, we have, by assumption

$$N_{\alpha\beta}(\tilde{J},h)v^a v^\beta \ge 0,$$

which implies, at $(x_0, t_0)$ again,

$$N_{\alpha\beta}(J, h) v^a v^\beta \geq -C\Sigma_{k=1}^m \epsilon^k e^{kAt}.$$

Here the constant $C$ depends on the bound of $M_{\alpha\beta}$. Substituting this inequality to (4.5.2), we come to the inequality

$$C\Sigma_{k=1}^m \epsilon^k e^{kAt_0} \geq \epsilon A e^{At_0}.$$

But this is impossible for a fixed large $A$ and all sufficiently small $\epsilon$. Hence we have proven the claim. Letting $\epsilon \to 0$, we finish the proof of the theorem. □

**Exercise 4.5.2** *State and prove a noncompact version of Theorem 4.5.2.*

## 4.6 Gradient estimates for the heat equation

In this section we present several gradient estimates for the heat equation $\Delta u - \partial_t u = 0$ on manifolds. These are taken from the works of Li-Yau [LY], R. Hamilton [Ha5] and Souplet-Zhang [SZ]. These gradient estimates can be regarded as differential versions of the Harnack inequality. Therefore they are also closely related to Sobolev inequalities. Unlike the Harnack inequality in the previous Section 4.4, these gradient estimates follow from the maximum principle. The general idea is to show that certain quantities involving the gradient of a positive solution to the heat equation satisfy another linear or nonlinear parabolic equation for which the maximum principle is applicable. This is the case, for example, if the nonlinear term has the right sign. The gradient estimate by Li-Yau has been extended by Hamilton to a matrix form and to cover the scalar curvature under Ricci flow. Perelman's gradient estimate for the conjugate heat equation [P1] Section 9 is also similar in spirit.

In 1986 Li and Yau [LY] proved, among other things, the following famous estimate.

**Theorem** *(Li-Yau [LY]). Let* $\mathbf{M}$ *be a complete manifold with dimension* $n \geq 2$, *Ricci*$(\mathbf{M}) \geq -K$, $K \geq 0$. *Suppose* $u$ *is any positive solution to the heat equation*

$$\Delta u - \partial_t u = 0,$$

*in* $B(x_0, R) \times [t_0 - T, t_0] \subset \mathbf{M} \times [t_0 - T, t_0]$*. Then, for any* $\alpha \in (0, 1)$*, there exists a constant* $c = c(n, \alpha)$ *such that*

$$\alpha \frac{|\nabla u|^2}{u^2} - \frac{u_t}{u} \le \frac{c}{R^2} + \frac{c}{T} + cK,$$

*in* $B(x_0, R/2) \times [t_0 - T/2, t_0]$*.*

*Moreover, if* $\mathbf{M}$ *has nonnegative Ricci curvature and* $u$ *is a positive solution of the heat equation in* $\mathbf{M} \times (0, T]$*, then, at* $(x, t) \in \mathbf{M} \times (0, T]$*,*

$$\frac{|\nabla u|^2}{u^2} - \frac{u_t}{u} \le \frac{c_n}{t}.$$

We mention that in the paper by Aronson and Bénilan [AB], a similar estimate for the porous medium equation and the heat equation in $\mathbf{R^n}$ has also appeared.

Let us observe that, even in the case of nonnegative Ricci curvature, the first local estimate does not match the second global estimate completely, due to the presence of the parameter $\alpha < 1$. Rather than presenting the original proof of [LY], we provide a slightly different local Li-Yau estimate. Here we show that, modulo a lower order term, $\alpha$ can be taken as 1 and hence the global and local estimate indeed agree. The short proof is based on a modification of an idea in [Ha5] and the cut-off method in [LY].

**Theorem 4.6.1** *( [Z1]) Let* $B(x_0, R)$ *be a geodesic ball in a Riemann manifold* $\mathbf{M}$ *with dimension* $n \ge 2$ *such that* $Ricci|_{B(x_0,R)} \ge -K$*,* $K \ge 0$*. Suppose* $u$ *is any positive solution to the heat equation in* $Q \equiv B(x_0, R) \times [t_0 - T, t_0]$*. Then*

$$\frac{|\nabla u|^2}{u^2} - \frac{u_t}{u} \le \frac{c_n}{R^2} + \frac{c_n}{T} + c_n K + c_n \sqrt{K} \sup_Q \frac{|\nabla u|}{u},$$

*in* $B(x_0, R/2) \times [t_0 - T/2, t_0]$*. Here* $c_n$ *depends only on the dimension* $n$*.*

Proof. By direct computation (see [Ha5]), we have, in a local coordinate,

$$(\Delta - \partial_t)(\frac{|\nabla u|^2}{u}) = \frac{2}{u}\left|\partial_i \partial_j u - \frac{\partial_i u \partial_j u}{u}\right|^2 + 2R_{ij}\frac{\partial_i u \partial_j u}{u}.$$

In view of the inequality

$$\left|\partial_i\partial_j u - \frac{\partial_i u \partial_j u}{u}\right|^2 \geq \frac{1}{n}\left(\Delta u - \frac{|\nabla u|^2}{u}\right)^2,$$

the above implies

$$(\Delta - \partial_t)(\frac{|\nabla u|^2}{u}) \geq \frac{2}{nu}\left(\Delta u - \frac{|\nabla u|^2}{u}\right)^2 + 2R_{ij}\frac{\partial_i u \partial_j u}{u}.$$

Since $\Delta u$ is also a solution to the heat equation, it follows that

$$(\Delta - \partial_t)(-\Delta u + \frac{|\nabla u|^2}{u}) \geq \frac{2}{nu}\left(\Delta u - \frac{|\nabla u|^2}{u}\right)^2 - 2K\frac{|\nabla u|^2}{u}.$$

Let us write

$$H = -\Delta u + \frac{|\nabla u|^2}{u} = \frac{|\nabla u|^2}{u} - u_t.$$

Then $H$ satisfies

$$(\Delta - \partial_t)H \geq \frac{2}{nu}H^2 - 2K\frac{|\nabla u|^2}{u}. \tag{4.6.1}$$

Now, define

$$Y = H/u = \frac{|\nabla u|^2}{u^2} - \frac{\partial_t u}{u} = -\Delta \ln u.$$

From the above inequality for $H$, we calculate

$$(\Delta - \partial_t)Y + 2\frac{\nabla u}{u}\nabla Y \geq \frac{2}{n}Y^2 - 2K\frac{|\nabla u|^2}{u^2}. \tag{4.6.2}$$

Now we can use the Li-Yau idea of cut-off functions to derive the desired bound. The only place that may cause difficulty is that $Y$ may change sign. However it turns out that it does not hurt. Here is the detail. Let $\psi = \psi(x,t)$ be a smooth cut-off function supported in $Q_{R,T} \equiv B(x_0, R) \times [t_0 - T, t_0]$, satisfying the following properties

(1) $\psi = \psi(d(x,x_0),t) \equiv \psi(r,t)$; $\psi(x,t) = 1$ in $Q_{R/2,T/4}$, $0 \leq \psi \leq 1$.
(2) $\psi$ is decreasing as a radial function in the spatial variables.
(3) $\frac{|\partial_r \psi|}{\psi^a} \leq \frac{C_a}{R}$, $\frac{|\partial_r^2 \psi|}{\psi^a} \leq \frac{C_a}{R^2}$ when $0 < a < 1$.
(4) $\frac{|\partial_t \psi|}{\psi^{1/2}} \leq \frac{C}{T}$.

Then, from (4.6.2) and a straightforward calculation, one has

$$\begin{aligned}
&\Delta(\psi Y) - (\psi Y)_t - 2\frac{\nabla\psi}{\psi}\cdot\nabla(\psi Y) + 2\frac{\nabla u}{u}\cdot\nabla(\psi Y) - 2\nabla\psi\cdot\frac{\nabla u}{u}Y \\
&\qquad + 2\psi K\frac{|\nabla u|^2}{u^2} \\
&\quad \geq \frac{2}{n}\psi Y^2 + (\Delta\psi)Y - 2\frac{|\nabla\psi|^2}{\psi}Y - \psi_t Y \\
&\quad = \frac{2}{n}\psi Y^2 - 2\frac{|\nabla\psi|^2}{\psi}Y + (\partial_r^2\psi + (n-1)\frac{\partial_r\psi}{r} + \partial_r\psi\partial_r\log\sqrt{g})Y - \psi_t Y.
\end{aligned} \tag{4.6.3}$$

Suppose that at $(y, s)$, the function $\psi Y$ reaches a maximum. If the value is nonpositive, there is nothing to prove. So we assume the maximum value is positive. Then (4.6.3) shows

$$\begin{aligned}
&2\psi K\frac{|\nabla u|^2}{u^2} + 2\frac{|\nabla\psi|^2}{\psi}Y + 2|\nabla\psi|\frac{|\nabla u|}{u}Y \\
&\quad \geq \frac{2}{n}\psi Y^2 + (\partial_r^2\psi + (n-1)\frac{\partial_r\psi}{r} + \partial_r\psi\partial_r\log\sqrt{g})Y - \psi_t Y.
\end{aligned}$$

In the above, the only term we need extra care of is

$$\partial_r\psi\partial_r\log\sqrt{g}Y.$$

Note that $-C/R \leq \partial_r\psi/\psi^a \leq 0$, $\partial_r\log\sqrt{g} \leq \sqrt{K}$ and $Y(y, s) > 0$. Therefore, at $(y, s)$,

$$\begin{aligned}
&2\psi K\frac{|\nabla u|^2}{u^2} + 2\frac{|\nabla\psi|^2}{\psi}Y + 2|\nabla\psi|\frac{|\nabla u|}{u}Y \\
&\quad \geq \frac{2}{n}\psi Y^2 + (\partial_r^2\psi + (n-1)\frac{\partial_r\psi}{r})Y - C\sqrt{K}\psi^a Y/R - \psi_t Y.
\end{aligned}$$

In the terms involving the first-order term of $Y$, we write $Y = \sqrt{1/\psi}\sqrt{\psi}Y$ and use Young's inequality. This shows, at $(y, s)$, that

$$\psi Y^2 = \psi\left(\frac{|\nabla u|^2}{u^2} - \frac{u_t}{u}\right)^2 \leq (\frac{c_n}{R^4} + \frac{c_n}{T^2} + c_n K^2) + c_n K\frac{|\nabla u|^2}{u^2}.$$

Since $\psi = 1$ on $Q_{R/2,T/2}$, we have, in $Q_{R/2,T/2}$,

$$\frac{|\nabla u|^2}{u^2} - \frac{u_t}{u} \leq \frac{c_n}{R^2} + \frac{c_n}{T} + c_n K + c_n\sqrt{K}\sup_{Q_{R,T}}\frac{|\nabla u|}{u}.$$

□

This kind of parabolic gradient estimate is rooted in an elliptic type gradient estimate.

**Theorem** *(Cheng-Yau [CY]). Let* $\mathbf{M}$ *be a complete manifold with dimension* $n \geq 2$, *Ricci*$(\mathbf{M}) \geq -K$, $K \geq 0$. *Suppose* $u$ *is any positive harmonic function in a geodesic ball* $B(x_0, R) \subset \mathbf{M}$. *There holds*

$$\frac{|\nabla u|}{u} \leq \frac{c_n}{R} + c_n \sqrt{K}, \tag{4.6.4}$$

*in* $B(x_0, R/2)$, *where* $c_n$ *depends only on the dimension* $n$.

Clearly the above Li-Yau estimate reduces to the Cheng-Yau estimate when $u$ is independent of time. On the other hand, for a time dependent solution of the heat equation, it is well known that the Cheng-Yau type elliptic gradient estimate cannot hold in general. This can be seen from the simple example where $u(x,t) = e^{-|x|^2/4t}/(4\pi t)^{n/2}$ being the fundamental solution of the heat equation in $\mathbf{R}^n$. The parabolic Harnack inequality also exhibits the same phenomenon in that the temperature at a given point in space time is controlled from the above by the temperature at a later time.

However R. Hamilton proved the following theorem, which is an elliptic type estimate for bounded solutions.

**Theorem** *(Hamilton [Ha5]) Let* $\mathbf{M}$ *be a compact manifold without boundary and with Ricci*$(\mathbf{M}) \geq -k$, $k \geq 0$ *and* $u$ *be a smooth positive solution of the heat equation with* $u \leq M$ *for all* $(x,t) \in \mathbf{M} \times (0, \infty)$. *Then*

$$\frac{|\nabla u|^2}{u^2} \leq (\frac{1}{t} + 2k) \ln \frac{M}{u}. \tag{4.6.5}$$

For a proof, see Theorem 6.5.1 where a version of this inequality in the context of Ricci flow is proven.

Just like the Cheng-Yau and Li-Yau estimates, it would be highly desirable to have a noncompact or localized version of Hamilton's estimate. However, the example in Remark 4.6.1 below shows that the suspected noncompact version of Hamilton's estimate is *false* even for $\mathbf{R}^n$. This situation contrasts sharply with the Cheng-Yau and Li-Yau inequality for which the local and noncompact versions are readily available.

However, in the next theorem, we show, for noncompact manifolds the elliptic Cheng-Yau estimate actually holds for the heat equation, after inserting a necessary logarithmic correction term. This correction term is slightly bigger than that in Hamilton's theorem (in the power of the log term). But the estimate holds for noncompact manifolds and it also has a localized version as the Cheng-Yau estimate. This result seems unexpected since it enables the comparison of temperature distribution instantaneously, without any lag in time, even for noncompact manifolds, regardless of the boundary behavior (see Remark 4.6.1). In some cases, our estimate (see (4.6.7) below) even holds for *any* positive solutions, bounded or not. This result seems new even in $\mathbf{R}^n$ or compact manifolds.

Here is the statement of the theorem.

**Theorem 4.6.2** *( [SZ]) Let* $\mathbf{M}$ *be a Riemann manifold of dimension* $n \ge 2$ *such that*

$$Ricci(\mathbf{M}) \ge -k, \qquad k \ge 0.$$

*Suppose* $u$ *is any positive solution to the heat equation in* $Q_{R,T} \equiv B(x_0, R) \times [t_0 - T, t_0] \subset \mathbf{M} \times (-\infty, \infty)$*. Suppose also* $u \le M$ *in* $Q_{R,T}$*. Then there exists a dimensional constant* $c$ *such that*

$$\frac{|\nabla u(x,t)|}{u(x,t)} \le c(\frac{1}{R} + \frac{1}{T^{1/2}} + \sqrt{k})\big(1 + \ln \frac{M}{u(x,t)}\big) \tag{4.6.6}$$

*in* $Q_{R/2,T/2}$.

*Moreover, if* $\mathbf{M}$ *has nonnegative Ricci curvature and* $u$ *is any positive solution of the heat equation on* $\mathbf{M} \times (0, \infty)$*, then there exist dimensional constants* $c_1, c_2$ *such that*

$$\frac{|\nabla u(x,t)|}{u(x,t)} \le c_1 \frac{1}{t^{1/2}} \big(c_2 + \ln \frac{u(x,2t)}{u(x,t)}\big) \tag{4.6.7}$$

*for all* $x \in \mathbf{M}$ *and* $t > 0$.

An immediate application of the theorem is the following time-dependent Liouville theorem, generalizing Yau's celebrated Liouville theorem for positive harmonic functions, which states that any positive harmonic function on a noncompact manifold with nonnegative Ricci curvature is a constant. One tends to expect that Liouville theorem would still hold for positive ancient solutions to the heat equation. However the following simple example shows that this expectation is false. Let $u = e^{x+t}$ for $x \in \mathbf{R}^1$. Clearly, $u$ is a positive ancient solution for the heat equation in $\mathbf{R}^1$ and it is not a constant. Nevertheless, our next theorem shows that under certain growth conditions, the Liouville theorem continues to hold for positive ancient solutions of the heat equation. Moreover, our growth condition in the spatial direction is sharp by the above example.

**Theorem 4.6.3** *( [SZ]) Let* $\mathbf{M}$ *be a complete, noncompact manifold with nonnegative Ricci curvature. Then the following conclusions hold.*

*(a). Let* $u$ *be a positive ancient solution to the heat equation, i.e. the solution is defined in* $\mathbf{M} \times (-\infty, T)$ *for some* $T \geq 0$*. Suppose that* $u(x,t) = e^{o(d(x)+\sqrt{|t|})}$ *near infinity. Then* $u$ *is a constant.*

*(b). Let* $u$ *be an ancient solution to the heat equation such that* $u(x,t) = o\big([d(x) + \sqrt{|t|}]\big)$ *near infinity. Then* $u$ *is a constant.*

Note that the growth condition on the second statement of the above theorem is also sharp in the spatial direction, due to the example $u = x$.

**Remark 4.6.1** *Here we give an example showing that the above theorem is sharp for incomplete or noncompact manifolds. This is surprising since it shows that Hamilton's estimate for the compact case is actually false for incomplete or noncompact manifolds in general. Recently in [Ko], Hamilton's estimate is generalized to bounded solutions of the heat equation on certain noncompact manifolds.*

*For* $a > 0$ *consider* $u = e^{ax+a^2t}$*. Clearly* $u$ *is a positive solution of the heat equation in* $Q = [1,3] \times [1,2] \subset \mathbf{R} \times (-\infty, \infty)$*. Also* $\nabla u(2,2)/u(2,2) = a$ *and* $M = \sup_Q u = e^{3a+2a^2}$*. Hence* $\log(M/u(2,2)) = a$*. Therefore at* $(x,t) = (2,2)$ *the left-hand side and right-hand side of (4.6.6) with* $R = T = 1$ *are* $a$ *and* $c(1+a)$ *respectively. Obviously they are equivalent when* $a$ *is large.*

The general idea for proof of Theorem 4.6.2 follows that of the paper [LY] where the maximum principle and cut-off functions are applied for an equation satisfied by $|\nabla \log u|^2$. Here $u$ is a positive solution to the heat equation. However the quantity we are using differs significantly from that of [LY]. What we are using is $|\nabla \log(M - \log u)|^2$, where $M$ is an upper bound for $u$. This quantity is also quite different from the one used in [Ha5], i.e. $|\nabla u|^2/u$. In fact a routine localization of Hamilton's method does not work here since, as pointed out earlier, that would yield a wrong result for the noncompact case.

**Proof of Theorem 4.6.2**

Suppose $u$ is a solution to the heat equation in the statement of the theorem in the parabolic cube $Q_{R,T} = B(x_0, R) \times [t_0 - T, t_0]$. It is clear that the gradient estimate in the theorem is invariant under the scaling $u \to u/M$. Therefore, we can and do assume that $0 < u \le 1$.

Write

$$f = \ln u, \qquad w \equiv |\nabla \ln(1-f)|^2 = \frac{|\nabla f|^2}{(1-f)^2}. \tag{4.6.8}$$

Since $u$ is a solution to the heat equation, simple calculation shows that

$$\Delta f + |\nabla f|^2 - f_t = 0. \tag{4.6.9}$$

We will derive an equation for $w$. First notice that

$$\begin{aligned} w_t &= \frac{2\nabla f(\nabla f)_t}{(1-f)^2} + \frac{2|\nabla f|^2 f_t}{(1-f)^3} \\ &= \frac{2\nabla f \nabla(\Delta f + |\nabla f|^2)}{(1-f)^2} + \frac{2|\nabla f|^2(\Delta f + |\nabla f|^2)}{(1-f)^3} \end{aligned}$$

In local orthonormal system, this can be written as

$$w_t = \frac{2f_j f_{iij} + 4 f_i f_j f_{ij}}{(1-f)^2} + 2\frac{f_i^2 f_{jj} + |\nabla f|^4}{(1-f)^3}. \tag{4.6.10}$$

Here and below, we have adopted the convention $u_i^2 = |\nabla u|^2$ and $u_{ii} = \Delta u$.

Next

$$\nabla w = \Big(\frac{f_i^2}{(1-f)^2}\Big)_j = \frac{2f_i f_{ij}}{(1-f)^2} + 2\frac{f_i^2 f_j}{(1-f)^3}. \tag{4.6.11}$$

It follows that

$$\begin{aligned} \Delta w &= \Big(\frac{f_i^2}{(1-f)^2}\Big)_{jj} \\ &= \frac{2f_{ij}^2}{(1-f)^2} + \frac{2f_i f_{ijj}}{(1-f)^2} + \frac{4 f_i f_{ij} f_j}{(1-f)^3} \\ &\quad + \frac{4 f_i f_{ij} f_j}{(1-f)^3} + 2\frac{f_i^2 f_{jj}}{(1-f)^3} + 6\frac{f_i^2 f_j^2}{(1-f)^4}. \end{aligned} \tag{4.6.12}$$

By (4.6.12) and (4.6.10),

$$\begin{aligned}
&\Delta w - w_t \\
&= \frac{2f_{ij}^2}{(1-f)^2} + 2\frac{f_i f_{ijj} - f_j f_{iij}}{(1-f)^2} \\
&\quad + 6\frac{|\nabla f|^4}{(1-f)^4} + 8\frac{f_i f_{ij} f_j}{(1-f)^3} + 2\frac{f_i^2 f_{jj}}{(1-f)^3} \\
&\quad - 4\frac{f_i f_{ij} f_j}{(1-f)^2} - 2\frac{f_i^2 f_{jj}}{(1-f)^3} - 2\frac{|\nabla f|^4}{(1-f)^3}.
\end{aligned}$$

The 5th and 7th terms on the right-hand side of this identity cancel each other. Also, by Bochner's identity

$$f_i f_{ijj} - f_j f_{iij} = f_j(f_{jii} - f_{iij}) = R_{ij} f_i f_j \geq -k|\nabla f|^2,$$

where $R_{ij}$ is the Ricci curvature. Therefore

$$\begin{aligned}
\Delta w - w_t \geq \frac{2f_{ij}^2}{(1-f)^2} + 6\frac{|\nabla f|^4}{(1-f)^4} + 8\frac{f_i f_{ij} f_j}{(1-f)^3} \\
- 4\frac{f_i f_{ij} f_j}{(1-f)^2} - 2\frac{|\nabla f|^4}{(1-f)^3} - \frac{2k|\nabla f|^2}{(1-f)^2}.
\end{aligned} \tag{4.6.13}$$

Notice from (4.6.11) that

$$\nabla f \nabla w = \frac{2f_i f_{ij} f_j}{(1-f)^2} + 2\frac{f_i^2 f_j^2}{(1-f)^3}.$$

Hence

$$0 = 4\frac{f_i f_{ij} f_j}{(1-f)^2} - 2\nabla f \nabla w + 4\frac{|\nabla f|^4}{(1-f)^3}, \tag{4.6.14}$$

$$0 = -4\frac{f_i f_{ij} f_j}{(1-f)^3} + [2\nabla f \nabla w - 4\frac{|\nabla f|^4}{(1-f)^3}]\frac{1}{1-f}. \tag{4.6.15}$$

Adding (4.6.13) with (4.6.14) and (4.6.15), we deduce

$$\begin{aligned}
&\Delta w - w_t \\
&\geq \frac{2f_{ij}^2}{(1-f)^2} + 2\frac{|\nabla f|^4}{(1-f)^4} + 4\frac{f_i f_{ij} f_j}{(1-f)^3} \\
&\quad + \frac{2}{1-f}\nabla f \nabla w - 2\nabla f \nabla w + 2\frac{|\nabla f|^4}{(1-f)^3} - \frac{2k|\nabla f|^2}{(1-f)^2}.
\end{aligned}$$

Since

$$\frac{2f_{ij}^2}{(1-f)^2} + 2\frac{|\nabla f|^4}{(1-f)^4} + 4\frac{f_i f_{ij} f_j}{(1-f)^3} \geq 0,$$

we have

$$\Delta w - w_t \geq \frac{2f}{1-f}\nabla f \nabla w + 2\frac{|\nabla f|^4}{(1-f)^3} - \frac{2k|\nabla f|^2}{(1-f)^2}.$$

Since $f \leq 0$, it follows that

$$\Delta w - w_t \geq \frac{2f}{1-f}\nabla f \nabla w + 2(1-f)\frac{|\nabla f|^4}{(1-f)^4} - \frac{2k|\nabla f|^2}{(1-f)^2},$$

i.e.

$$\Delta w - w_t \geq \frac{2f}{1-f}\nabla f \nabla w + 2(1-f)w^2 - 2kw. \tag{4.6.16}$$

From here, we will use the well-known cut-off function by Li-Yau [LY], to derive the desired bounds. We caution the reader that the calculation is not the same as in that of [LY] due to the difference of the first-order term.

Let $\psi = \psi(x,t)$ be a smooth cut-off function supported in $Q_{R,T}$, satisfying the following properties

(1). $\psi = \psi(d(x,x_0),t) \equiv \psi(r,t)$; $\psi(x,t) = 1$ in $Q_{R/2,T/4}$, $0 \leq \psi \leq 1$.

(2). $\psi$ is decreasing as a radial function in the spatial variables.

(3). $\frac{|\partial_r \psi|}{\psi^a} \leq \frac{C_a}{R}$, $\frac{|\partial_r^2 \psi|}{\psi^a} \leq \frac{C_a}{R^2}$ when $0 < a < 1$.

(4). $\frac{|\partial_t \psi|}{\psi^{1/2}} \leq \frac{C}{T}$.

Then, from (4.6.16) and a straightforward calculation, one has

$$\begin{aligned}\Delta(\psi w) &+ b\cdot\nabla(\psi w) - 2\frac{\nabla\psi}{\psi}\cdot\nabla(\psi w) - (\psi w)_t \\ &\geq 2\psi(1-f)w^2 + (b\cdot\nabla\psi)w - 2\frac{|\nabla\psi|^2}{\psi}w + (\Delta\psi)w - \psi_t w - 2kw\psi,\end{aligned} \tag{4.6.17}$$

where we have written

$$b = -\frac{2f}{1-f}\nabla f.$$

Suppose the maximum of $\psi w$ is reached at $(x_1,t_1)$. By [LY], we can assume, without loss of generality that $x_1$ is not in the cut-locus of $x_0$. Then at this point, one has, $\Delta(\psi w) \leq 0$, $(\psi w)_t \geq 0$ and $\nabla(\psi w) = 0$.

Therefore

$$2\psi(1-f)w^2(x_1,t_1) \le -[\ (b\cdot\nabla\psi)w - 2\frac{|\nabla\psi|^2}{\psi}w + (\Delta\psi)w - \psi_t w + 2kw\psi](x_1,t_1). \tag{4.6.18}$$

We need to find an upper bound for each term of the right-hand side of (4.6.18).

$$\begin{aligned}|(b\cdot\nabla\psi)w| &\le \frac{2|f|}{1-f}|\nabla f|w|\nabla\psi| \le 2w^{3/2}|f|\ |\nabla\psi|\\ &= 2[\psi(1-f)w^2]^{3/4}\ \frac{f|\nabla\psi|}{[\psi(1-f)]^{3/4}}\\ &\le \psi(1-f)w^2 + c\frac{(f|\nabla\psi|)^4}{[\psi(1-f)]^3}.\end{aligned}$$

This implies

$$|(b\cdot\nabla\psi)w| \le (1-f)\psi w^2 + c\frac{f^4}{R^4(1-f)^3}. \tag{4.6.19}$$

For the second term on the right-hand side of (4.6.18), we proceed as follows

$$\begin{aligned}\frac{|\nabla\psi|^2}{\psi}w &= \psi^{1/2}w\frac{|\nabla\psi|^2}{\psi^{3/2}}\\ &\le \frac{1}{8}\psi w^2 + c\big(\frac{|\nabla\psi|^2}{\psi^{3/2}}\big)^2 \le \frac{1}{8}\psi w^2 + c\frac{1}{R^4}.\end{aligned} \tag{4.6.20}$$

Furthermore, by the properties of $\psi$ and the assumption of on the Ricci curvature, one has

$$\begin{aligned}-(\Delta\psi)w &= -(\partial_r^2\psi + (n-1)\frac{\partial_r\psi}{r} + \partial_r\psi\partial_r\ln\sqrt{g})w\\ &\le (|\partial_r^2\psi| + 2(n-1)\frac{|\partial_r\psi|}{R} + \sqrt{k}|\partial_r\psi|)w\\ &\le \psi^{1/2}w\frac{|\partial_r^2\psi|}{\psi^{1/2}} + \psi^{1/2}w2(n-1)\frac{|\partial_r\psi|}{R\psi^{1/2}} + \psi^{1/2}w\frac{\sqrt{k}|\partial_r\psi|}{\psi^{1/2}}\\ &\le \frac{1}{8}\psi w^2 + c\big([\frac{|\partial_r^2\psi|}{\psi^{1/2}}]^2 + [\frac{|\partial_r\psi|}{R\psi^{1/2}}]^2 + [\frac{\sqrt{k}|\partial_r\psi|}{\psi^{1/2}}]^2\big).\end{aligned}$$

Therefore

$$-(\Delta\psi)w \le \frac{1}{8}\psi w^2 + c\frac{1}{R^4} + ck\frac{1}{R^2}. \tag{4.6.21}$$

Now we estimate $|\psi_t|\ w$.

$$\begin{aligned}|\psi_t|\ w &= \psi^{1/2} w \frac{|\psi_t|}{\psi^{1/2}} \\ &\le \frac{1}{8}\big(\psi^{1/2} w\big)^2 + c\big(\frac{|\psi_t|}{\psi^{1/2}}\big)^2.\end{aligned}$$

This shows

$$|\psi_t| w \le \frac{1}{8}\psi w^2 + c\frac{1}{T^2}. \tag{4.6.22}$$

Finally, for the last term, we have

$$2kw\psi \le \frac{1}{8}\psi w^2 + ck^2. \tag{4.6.23}$$

Substituting (4.6.19)–(4.6.23) to the right-hand side of (4.6.18), we deduce,

$$2(1-f)\psi w^2 \le (1-f)\psi w^2 + c\frac{f^4}{R^4(1-f)^3} + \frac{1}{2}\psi w^2 + \frac{c}{R^4} + \frac{c}{T^2} + \frac{ck}{R^2} + ck^2.$$

Recall that $f \le 0$, therefore the above implies

$$\psi w^2(x_1, t_1) \le c\frac{f^4}{R^4(1-f)^4} + \frac{1}{2}\psi w^2(x_1, t_1) + \frac{c}{R^4} + \frac{c}{T^2} + ck^2.$$

Since $\frac{f^4}{(1-f)^4} \le 1$, the above shows, for all $(x, t)$ in $Q_{R,T}$,

$$\begin{aligned}\psi^2(x,t)w^2(x,t) &\le \psi^2(x_1,t_1)w^2(x_1,t_1) \\ &\le \psi(x_1,t_1)w^2(x_1,t_1) \\ &\le c\frac{c}{R^4} + \frac{c}{T^2} + ck^2.\end{aligned}$$

Notice that $\psi(x,t) = 1$ in $Q_{R/2,T/4}$ and $w = |\nabla f|^2/(1-f)^2$. We finally have

$$\frac{|\nabla f(x,t)|}{1 - f(x,t)} \le \frac{c}{R} + \frac{c}{\sqrt{T}} + c\sqrt{k}.$$

We have completed the proof of the first statement in the theorem since $f = \ln(u/M)$ with $M$ scaled to 1.

To prove the second statement, we apply the first one on the cube $Q_{\sqrt{t},t/2} = B(x, \sqrt{t}) \times [t/2, t]$. By Li-Yau's inequality [LY], we know that

$$M = \sup_{Q_{\sqrt{t},t/2}} u \le cu(x, 2t).$$

Now the second inequality follows from the first. □

The proof of 4.6.3 follows easily from Theorem 4.6.2.

**Exercise 4.6.1** *Prove Theorem 4.6.3.*

# Chapter 5

# Basics of Ricci flow

## 5.1 Local existence, uniqueness and basic identities

In this section we introduce the concept of Hamilton's Ricci flow together with a number of important properties. Due to the vastness of the subject, we can only touch upon the bare essentials without proof. We refer the interested reader to the original papers of [Ha1] and the books [CK], [CLN] and [Cetc] for further details.

We begin by introducing the definitions and notations, which will be used in this and later chapters. $\mathbf{M}$ or $M$ denotes a complete compact, or noncompact Riemann manifold, unless stated otherwise; $g$ (or $g_{ij}$), $R_{ij}$ (or $Ric$) will be the metric and Ricci curvature respectively; $R$ is for the scalar curvature; $\nabla$, $\Delta$ are the corresponding gradient and Laplace-Beltrami operator; $c$, $C$ with or without index denote generic positive constant that may change from line to line. If the metric $g(t)$ evolves with time, then $d(x, y, t)$ or $d(x, y, g(t))$ will denote the corresponding distance function; $dg(x, t)$ or $dg(t)$, or $d\mu(g(t))$ denote the volume element under $g(t)$. We will use $B(x, r, t)$ or $B(x, r, g(t))$ to denote the geodesic ball centered at $x$ with radius $r$ under the metric $g(t)$; $|B(x, r; t)|_s$ to denote the volume of $B(x, r; t)$ under the metric $g(s)$. We will still use $\nabla$, $\Delta$ to denote the corresponding gradient and Laplace-Beltrami operator for $g(t)$, without mentioning the time $t$, when no confusion arises. Let $p \in \mathbf{M}$ and $t$ be a time in the Ricci flow, we use $(\mathbf{M}, p)$ or $(\mathbf{M}, p, t)$ to denote the manifold marked by $p$ or $(p, t)$.

**Definition 5.1.1** *Let* $\mathbf{M}$ *be a Riemann manifold and* $g(t) = g_{ij}(t)$ *be a family of Riemann metric depending on time* $t$ *in some interval*

$[T_0, T) \subset \mathbf{R}$. *Let* $R_{ij}$ *be the corresponding Ricci curvature. If*

$$\frac{\partial g_{ij}}{\partial t} = -2R_{ij}, \tag{5.1.1}$$

*then one says that* $(\mathbf{M}, g_{ij}(t))$ *is a Ricci flow.*

This system of equations, introduced by Richard Hamilton in 1982 [Ha1], is a quasilinear, degenerate second order parabolic system. Hamilton proved, for a compact 3 dimensional manifold, that the Ricci flow develops singularity in finite time in a uniform manner if the initial metric $g(0)$ has positive Ricci curvature. By carefully studying the formation of singularity, he proved that such a manifold with positive Ricci curvature is diffeomorphic to the standard 3 sphere $S^3$.

We point out that $g_{ij}$ and $R_{ij}$ are regarded as the local expression of the metric and Ricci curvature under a background local coordinates. For example, one can choose local coordinates on $\mathbf{M}$ which is independent of time or the metrics. One can also pick local coordinates which depend on time and metrics in a way that the equations for curvatures associated with Ricci flow become simpler. Sometimes we also use the compressed notation $\partial_t g = -2Ric$ for Ricci flow, which just means (5.1.1).

As a warm up run, we discuss a simple example of Ricci flow on an Einstein manifold. In general, the Ricci flow is so complicated that there are only a few explicit examples in the form of Ricci solitons (c.f. Definition 5.4.2).

Let $\mathbf{M}$ be a Riemann manifold equipped with a family of metric such that

$$R_{ij}(x, 0) = \lambda g_{ij}(x, 0),$$
$$g_{ij}(x, t) = \rho^2(t) g_{ij}(x, 0).$$

Here $\lambda$ is a constant and $\rho > 0$ is a function to be determined by the Ricci flow. Since the Ricci tensor $R_{ij}$ (coefficients relative to the local basis $dx^i \otimes dx^j$) does not change under scaling, we have

$$R_{ij}(x, t) = R_{ij}(x, 0) = \lambda g_{ij}(x, 0).$$

Therefore the Ricci flow equation becomes

$$\frac{\partial \rho^2(t) g_{ij}(x, 0)}{\partial t} = -2\lambda g_{ij}(x, 0).$$

Its solution is

$$\rho^2 = 1 - 2\lambda t.$$

So the explicit formula for the evolving metric is

$$g_{ij}(x,t) = (1-2\lambda t)g_{ij}(x,0).$$

If $\lambda > 0$, the initial Ricci curvature is positive, the Ricci flow becomes singular when $t = 1/(2\lambda)$. If $\lambda < 0$, the Ricci flow exists for all time.

The Ricci flow is a second order weakly parabolic system for the metric $g_{ij}$. Initially Hamilton [Ha1] proved local existence and uniqueness of the Ricci flow on compact manifolds with the help of the Nash-Moser implicit function theorem. Then DeTurck [DT] gave a much simpler proof by converting the Ricci flow into a strictly parabolic system via a family of time dependent diffeomorphisms. Then the standard theory for strictly parabolic systems imply the existence and uniqueness of the converted system and the original Ricci flow. In the noncompact case, W. X. Shi [Shi] proved local existence of the Ricci flow for manifolds with bounded curvature operator. Under the same condition, the uniqueness is proven by Chen and Zhu [ChZ2]. Independently, Lu and Tian [LT] proved uniqueness for the standard solution of Ricci flow which is a noncompact flow on $\mathbf{R}^3$ with radially symmetric metrics. These results are summarized by the three theorems below.

**Theorem 5.1.1** *(R. Hamilton [Ha1]) Let* $(\mathbf{M}, g^0_{ij}(x))$ *be a compact Riemann manifold. There exists a constant* $T > 0$ *such that the initial value problem of the Ricci flow*

$$\begin{cases} \partial_t g_{ij}(x,t) = -2R_{ij}(x,t), \\ g_{ij}(x,0) = g^0_{ij}(x) \end{cases} \tag{5.1.2}$$

*has a unique smooth solution* $g_{ij}(x,t)$ *on* $\mathbf{M} \times [0,T)$.

Proof. Step 1. Construct a strictly parabolic system of modified Ricci flow.

Following DeTurck, consider a parabolic system which under, $\{x^1, \ldots, x^n\}$, a time independent local orthonormal coordinates, can be written as

$$\begin{cases} \partial_t g_{ij}(x,t) = -2Ric_{ij}(x,t) + [L_{V(t)}g]_{ij}(x,t) \\ g_{ij}(x,0) = g^0_{ij}(x). \end{cases} \tag{5.1.3}$$

Terms in the above system are defined as follows.

(i) $L_{V(t)}g$ is the Lie derivative of $g = g(t) = g_{ij}(\cdot, t)$ with respect to the vector field

$$V = V(t) = V^k(x,t)\partial_k,\ \partial_k = \partial/\partial x^k; \qquad V^k = g^{pq}(\Gamma^k_{pq} - \Gamma^k_{pq}(0)); \tag{5.1.4}$$

(ii) $R_{ij}$ and $\Gamma^k_{pq}$ are respectively the Ricci curvature and Christoffel symbol of $g = g(t)$ in (5.1.3).

(iii) $\Gamma^k_{pq}(0)$ is the Christoffel symbol of the initial metric $g^0$.

We now show that (5.1.3) is a strictly parabolic system of the metric.

Using Proposition 3.1.3, we know, for vector fields $X$ and $Y$, there hold

$$\partial_t g(t)(X,Y) = -2Ric_{g(t)}(X,Y) + g(t)(\nabla_X V(t), Y) + g(t)(X, \nabla_Y V(t)).$$

Hence in the local system, we have

$$\partial_t g_{ij} = -2R_{ij} + \nabla_i V_j + \nabla_j V_i. \tag{5.1.5}$$

Here, $\nabla_i V_j$ means the $j$-th component of the covariant derivative of the dual one form of $V$ with respect to $g$, i.e. the $j$-th component of $\nabla_{\partial/\partial x^i} V^*$ where $V^* = V_i dx^i$ and $V_i = g_{ik}V^k$.

The principal terms on the right-hand side of (5.1.5) are those containing the second order derivatives of the metric $g$. Let us write these in the local coordinates. Using the local formula (3.2.16),

$$R_{ij} = \partial_k \Gamma^k_{ij} - \partial_i \Gamma^k_{kj} + \Gamma^k_{kp}\Gamma^p_{ij} - \Gamma^k_{ip}\Gamma^p_{kj}.$$

From (3.1.3),

$$\Gamma^i_{jk} = \frac{1}{2} g^{il}(\partial_j g_{kl} + \partial_k g_{lj} - \partial_l g_{jk}).$$

Hence

$$\begin{aligned} R_{ij} &= -\frac{1}{2}\partial_i(g^{kl}\partial_j g_{kl}) + \frac{1}{2}\partial_k\left[g^{kl}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})\right] \\ &\qquad + \text{lower order terms} \\ &= -\frac{1}{2}g^{kl}\left(\partial_i\partial_j g_{kl} - \partial_k\partial_i g_{jl} - \partial_k\partial_j g_{il} + \partial_k\partial_l g_{ij}\right) \\ &\qquad + \text{lower order terms.} \end{aligned} \tag{5.1.6}$$

The lower order terms here and later in the proof are those containing at most the first derivative terms involving $g$. Using (5.1.4), we see that

$$\nabla_j V_i = g_{ik} g^{pq} \partial_j \Gamma^k_{pq} + \text{lower order terms.}$$

Consequently

$$\begin{aligned}\nabla_j V_i &= \frac{1}{2} g_{ik} g^{pq} \partial_j \left[ g^{kl} (\partial_q g_{pl} + \partial_p g_{ql} - \partial_l g_{pq}) \right] + \text{lower order terms} \\ &= \frac{1}{2} g^{pq} (\partial_j \partial_q g_{pi} + \partial_j \partial_p g_{qi} - \partial_j \partial_i g_{pq}) + \text{lower order terms.}\end{aligned}$$

Renaming the indices $p, q$ as $k, l$ and switching $i$ and $j$, we deduce

$$\begin{aligned}\nabla_i V_j + \nabla_j V_i = &\frac{1}{2} g^{kl} (\partial_j \partial_l g_{ki} + \partial_j \partial_k g_{li} - \partial_j \partial_i g_{kl}) \\ &+ \frac{1}{2} g^{kl} (\partial_i \partial_l g_{kj} + \partial_i \partial_k g_{lj} - \partial_i \partial_j g_{kl}) \\ &+ \text{lower order terms.}\end{aligned}$$

Combining this with (5.1.6), we arrive at

$$-2R_{ij} + \nabla_i V_j + \nabla_j V_i = g^{kl} \partial_k \partial_l g_{ij} + \text{lower order terms.}$$

Therefore (5.1.5) can be written as

$$\partial_t g_{ij} = g^{kl} \partial_k \partial_l g_{ij} + \text{lower order terms} = \Delta g_{ij} + \text{lower order terms.} \tag{5.1.7}$$

Equation (5.1.7) and hence (5.1.3) is a strictly parabolic, semi-linear system. The standard parabolic theory shows that (5.1.3) has a smooth solution at least for a short time.

Step 2. Show that the solutions of the modified system gives rise to a solution of the original Ricci flow via a family of diffeomorphisms.

Let $g = g(t)$ be a smooth solution of (5.1.3), define a family of diffeomorphisms $\phi_t$ via the equation

$$\frac{d\phi_t}{dt} = \hat{V}(\phi_t, t) \equiv (\phi_{t*} V)(\phi_t, t), \quad \phi_0 = I \tag{5.1.8}$$

where $V$ is the vector field defined in (5.1.4). We remark, for a smooth function $f$ on $\mathbf{M}$ and a point $p \in \mathbf{M}$,

$$\hat{V}(\phi_t(p), t)(f) = (\phi_{t*} V)(\phi_t(p), t)(f) = V(p, t)(f \circ \phi_t).$$

We show that the metrics

$$\hat{g}(t) \equiv (\phi_t^*)^{-1}(g(t)) \tag{5.1.9}$$

is a solution of the original Ricci flow equation. Since $g(t) = \phi_t^* \hat{g}(t)$, we compute

$$\begin{aligned} \partial_t g(t) &= \phi_t^* \partial_t \hat{g}(t) + \partial_t [\phi_t^* \hat{g}(l)]\big|_{l=t} \\ &= \phi_t^* \partial_t \hat{g}(t) + \phi_t^* (L_{\hat{V}(t)} \hat{g}(t)) \quad \text{(Proposition} \quad 3.1.4(ii)), \\ &= \phi_t^* \partial_t \hat{g}(t) + L_{(\phi_{t*})^{-1}\hat{V}(t)} \phi_t^* (\hat{g}(t)) \quad \text{(Remark} \quad 3.1.8), \\ &= \phi_t^* \partial_t \hat{g}(t) + L_{V(t)}(g(t)). \end{aligned} \tag{5.1.10}$$

On the other hand, by (5.1.3)

$$\partial_t g(t) = -2Ric_{g(t)} + L_{V(t)} g(t) = -2\phi_t^* (Ric_{\hat{g}(t)}) + L_{V(t)} g(t).$$

Therefore

$$\partial_t \hat{g}(t) = -2Ric_{\hat{g}(t)},$$

which proves the short time existence of the original Ricci flow (5.1.2).

Step 3. Proof of uniqueness.

First we observe that the above process is reversible. Namely, if $\hat{g} = \hat{g}(t)$ is a solution to (5.1.2), then one can construct a family of diffeomorphisms $\phi_t$ by (5.1.8) such that $g(t) \equiv \phi_t^* \hat{g}(t)$ is a solution to (5.1.3).

To prove this statement, we just need to show that (5.1.8) under the metric $g(t)$ has smooth solutions in short time. It is convenient to work in local system. Let $x = \{x^1, \ldots, x^n\}$ be the time independent orthonormal system as in Step 1. We suppose $\phi_t$, which is unknown so far, is given by

$$\phi_t(x) = (y^1(x,t), \ldots, y^n(x,t)).$$

Let $\hat{\Gamma}_{jl}^k$ be the Christoffel symbol of $\hat{g}(t)$, which is already given. Then, by standard calculation, the Christoffel symbol of $g(t) = \phi_t^* \hat{g}(t)$, denoted by $\Gamma_{jl}^k$ is:

$$\Gamma_{jl}^k = \frac{\partial y^\alpha}{\partial x^j} \frac{\partial y^\beta}{\partial x^l} \frac{\partial x^k}{\partial y^\gamma} \hat{\Gamma}_{\alpha\beta}^\gamma + \frac{\partial x^k}{\partial y^\alpha} \frac{\partial^2 y^\alpha}{\partial x^j \partial x^l},$$

Substituting this to (5.1.4), we see that

$$V = V^k \frac{\partial}{\partial x^k} = g^{jl} \left[ \frac{\partial y^\alpha}{\partial x^j} \frac{\partial y^\beta}{\partial x^l} \frac{\partial x^k}{\partial y^\gamma} \hat{\Gamma}_{\alpha\beta}^\gamma + \frac{\partial x^k}{\partial y^\alpha} \frac{\partial^2 y^\alpha}{\partial x^j \partial x^l} - \Gamma_{jl}^k(0) \right] \frac{\partial}{\partial x^k}.$$

Recall, for any smooth function $f$, $(\phi_{t*}V)(f) = V(f \circ \phi_t)$. Therefore

$$\begin{aligned}(\phi_{t*}V)(f) &= V^k \frac{\partial(f \circ \phi_t)}{\partial x^k} = V^k \frac{\partial f}{\partial y^\eta}\frac{\partial y^\eta}{\partial x^k} \\ &= g^{jl}\left[\frac{\partial y^\alpha}{\partial x^j}\frac{\partial y^\beta}{\partial x^l}\frac{\partial x^k}{\partial y^\gamma}\hat{\Gamma}^\gamma_{\alpha\beta} + \frac{\partial x^k}{\partial y^\alpha}\frac{\partial^2 y^\alpha}{\partial x^j \partial x^l} - \Gamma^k_{jl}(0)\right]\frac{\partial f}{\partial y^\eta}\frac{\partial y^\eta}{\partial x^k} \\ &= g^{jl}\left[\frac{\partial y^\alpha}{\partial x^j}\frac{\partial y^\beta}{\partial x^l}\hat{\Gamma}^\eta_{\alpha\beta} + \frac{\partial^2 y^\eta}{\partial x^j \partial x^l} - \Gamma^k_{jl}(0)\frac{\partial y^\eta}{\partial x^k}\right]\frac{\partial f}{\partial y^\eta}.\end{aligned}$$

By this then we can write (5.1.8) as

$$\begin{cases}\partial_t y^\eta = g^{jl}\left(\frac{\partial^2 y^\eta}{\partial x^j \partial x^l} + \hat{\Gamma}^\eta_{\gamma\beta}\frac{\partial y^\beta}{\partial x^j}\frac{\partial y^\gamma}{\partial x^l} - \Gamma^k_{jl}(0)\frac{\partial y^\eta}{\partial x^k}\right) \\ y^\eta(x,0) = x^\eta.\end{cases}$$

Note that $g_{jl} = \hat{g}_{\alpha\beta}\frac{\partial y^\alpha}{\partial x^j}\frac{\partial y^\beta}{\partial x^l}$ and $(g^{jl}) = (g_{jl})^{-1}$. Hence (5.1.8) is a quasilinear and strictly parabolic system which has a unique smooth solution $\phi_t = (y^1(x,t), \ldots, y^n(x,t))$ for short time. By reversing (5.1.10), we know that $g(t) = (\phi_t)^*\hat{g}(t)$ is a solution of the modified Ricci flow (5.1.3).

Suppose $\hat{g}^{(1)}(t)$ and $\hat{g}^{(2)}(t)$ are two solutions to the original Ricci flow (5.1.2). Let $\phi_t^{(1)}$ and $\phi_t^{(2)}$ be the two solutions of (5.1.8) corresponding to $\hat{g}^{(1)}(t)$ and $\hat{g}^{(2)}(t)$ respectively. Then the metrics $g^{(1)}(t) = (\phi_t^{(1)})^*\hat{g}^{(1)}(t)$ and $g^{(2)}(t) = (\phi_t^{(2)})^*\hat{g}^{(2)}(t)$ are two solutions to the modified Ricci flow (5.1.3), which share the same initial metric $g^0$. Since the equation in (5.1.3) is strictly parabolic, it can only have one solution with the same initial value. Therefore $g^{(1)}(t) = g^{(2)}(t)$ which together with (5.1.8) imply $\hat{g}^{(1)}(t) = \hat{g}^{(2)}(t)$. □

The metric $g^0_{ij}$ in the above theorem is called the initial metric. Sometimes it is convenient to scale the initial metric to a normalized one which we now define.

**Definition 5.1.2** *(normalized metric and normalized manifold) A metric in Riemann manifold is called a normalized metric if $|Rm| \le 1$ everywhere and the volume of every unit ball is at least half of the volume of the Euclidean unit ball.*

*A Riemann manifold equipped with a normalized metric is called a normalized manifold.*

Obviously one can always multiply the metric of a compact manifold by a large number to make the manifold with the scaled metric a normalized one.

The next two theorems generalize the above existence and uniqueness results to Ricci flow on certain noncompact manifolds.

**Theorem 5.1.2** *(W. X. Shi) [Shi] Let $(\mathbf{M}, g_{ij}(x))$ be a complete noncompact Riemann manifold with bounded curvature tensor. There exists a constant $T > 0$ such that the initial value problem of the Ricci flow*

$$\begin{cases} \partial_t g_{ij}(x,t) = -2R_{ij}(x,t), \\ g_{ij}(x,0) = g_{ij}(x) \end{cases}$$

*has a smooth solution $g_{ij}(x,t)$ on $\mathbf{M} \times [0,T)$, with uniformly bounded curvature tensor.*

**Theorem 5.1.3** *(Chen and Zhu) [ChZ1] The solution in Theorem 5.1.2 is unique.*

The proofs of these two theorems are quite long due to the technical issue of controlling behaviors of geometric quantities such as curvatures and injectivity radius near infinity. The reader is referred the original papers for the nuts and bolts of the proofs.

In the next proposition we will collect a number of formulas describing the evolution of geometric quantities along the Ricci flow. All these are due to R. Hamilton.

**Proposition 5.1.1** *(evolution of geometric quantities: scalar curvature, volume, arclength, connection, curvature tensor, Ricci curvature, Laplace operator)*

*Suppose $(\mathbf{M}, g(t))$ is a Ricci flow. Then the following conclusions are true.*

*(1) Let $R(x,t)$ be the scalar curvature with respect to $g_{ij}(x,t)$, then*

$$\Delta R - \partial_t R + 2|Ric|^2 = 0. \tag{5.1.11}$$

*(2) Let $d\mu(x,t)$ be the volume element with respect to $g_{ij}(x,t)$, then*

$$\partial_t d\mu(x,t) = -R(x,t) d\mu(x,t). \tag{5.1.12}$$

*(3) Let $x_0, x_1 \in \mathbf{M}$ and $d(x_0,x_1,t)$ be the distance between $x_0$, $x_1$ under the metric $g(t)$. Suppose $d$ is a smooth function of $t$, then*

$$\frac{d}{dt} d(x_0,x_1,t) = -inf \int_0^{d(x_0,x_1,t)} Ric(X(s),X(s)) ds.$$

*Here the "inf" is taken over all minimizing geodesic $c = c(s)$ which connects $x_0$, $x_1$, parameterized by arclength, under the metric $g(t)$. Also $X(s) = c'(s)$.*
*(4) Let $\Gamma^k_{ij}$ be the Christoffel symbol of $g_{ij}(x,t)$ under a time independent local coordinate $\{x^1, \dots, x^n\}$, then*

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

*(5) Let $R_{ijkl}$ be the curvature tensor of $g_{ij}(t)$ under a time independent local coordinate $\{x^1, \dots, x^n\}$, then*

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

*Here*

$$B_{ijkl} = -g^{pr}g^{qs}R_{piqj}R_{rksl}, \tag{5.1.13}$$

*$R_{qi}$ is the Ricci curvature and $\Delta$ is the rough Laplacian with respect to $g_{ij}(t)$ (c.f. Definition 3.7.1), i.e.*

$$\Delta R_{ijkl} = g^{pq}\nabla_p \nabla_q R_{ijkl} \equiv g^{pq}\nabla^2_{p,q} R_{ijkl}.$$

*(6) Let $R_{ij}$ be the Ricci curvature of $g_{ij}(t)$ under a time independent local coordinate $\{x^1, \dots, x^n\}$, then*

$$\partial_t R_{ij} = \Delta R_{ij} + 2g^{pl}g^{qm}R_{pijq}R_{lm} - 2g^{pq}R_{pi}R_{qj}.$$

*(7) For any smooth function $u = u(x)$ on $\mathbf{M}$,*

$$\left(\frac{\partial \Delta_{g(t)}}{\partial t}\right) u = 2 < Ric, Hess\, u > .$$

PROOF. All these can be derived from the Ricci flow equations in local coordinates in a straightforward, but laboring manner.

Proof of (1). See the proof of (6).

Proof of (2). In a time independent local orthonormal system $\{x^1, \dots, x^n\}$,

$$d\mu = \sqrt{g}dx^1 \wedge \dots \wedge dx^n \equiv \sqrt{det(g_{ij}(t))}dx^1 \wedge \dots \wedge dx^n.$$

Write $G = (g_{ij}(t))$. The formula follows immediately from

$$\frac{d}{dt} detG = detG\, tr(G^{-1}\frac{d}{dt}G).$$

Proof of (3). (Sketch. For details, see [CK] e.g.) Let $c = c(s)$ be a curve connecting $x$ and $x_1$, parameterized by $s \in [0, 1]$. Differentiate the length formula $L(c) = \int_0^1 \sqrt{g(c'(s), c'(s))} ds$. Then minimize among all such curves.

Proof of (4). For a fixed time $t_0$ and point $p \in \mathbf{M}$, let $\{\frac{\partial}{\partial x^i}\}$ be a local orthonormal system (c.f. Definition 3.4.6) under the metric $g(\cdot, t_0)$. Using this system, by the standard formula (3.1.3), we can write $\Gamma^k_{ij}(x, t)$, for $(x, t)$ in a small neighborhood of $(p, t_0)$, in the form of

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

Note that the operators $\partial_i = \frac{\partial}{\partial x^i}$ etc. are independent of time. Hence differentiating the above with respect to time gives

$$\begin{aligned}\partial_t \Gamma^k_{ij}(x, t) = &\frac{1}{2} \partial_t g^{kl}(x, t)(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})(x, t)\\ &+ \frac{1}{2} g^{kl}(x, t)(\partial_i \partial_t g_{jl} + \partial_j \partial_t g_{il} - \partial_l \partial_t g_{ij})(x, t).\end{aligned}$$

The Ricci flow equations turn this into

$$\begin{aligned}\partial_t \Gamma^k_{ij}(x, t) = &R^{kl}(x, t)(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})(x, t) - g^{kl}(x, t)(\partial_i R_{jl}\\ &+ \partial_j R_{il} - \partial_l R_{ij})(x, t).\end{aligned}$$

Now we take $x = p$ and $t = t_0$. At $p$, the center of the local orthonormal system, the quantities $\partial_l g_{ij}$, $\Gamma^k_{ij}$ are zero (see the remark after Definition 3.4.6). Hence, at $(p, t_0)$, by the formula in Definition 3.1.13 v),

$$\nabla_i R_{jl} = \partial_i R_{jl} - \Gamma^k_{ij} R_{kl} - \Gamma^k_{il} R_{jk} = \partial_i R_{jl}.$$

Therefore, at $(p, t_0)$,

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

as stated in (4).

Proof of (5). Again we assume that the local coordinate is an orthonormal one centered at a fixed point, where all computation below take place. At this point, $\Gamma^i_{jk} = 0$, hence the partial derivative $\partial_i = \frac{\partial}{\partial x^i}$

is the same as the covariant derivative $\nabla_i$ for any tensor. See Proposition 3.1.1.

Recall (see the third remark after Definition 3.2.2)

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

Therefore

$$\partial_t R^h_{ijk} = \partial_i \partial_t \Gamma^h_{jk} - \partial_j \partial_t \Gamma^h_{ik},$$

$$\begin{aligned}\partial_t R_{ijkl} &= \partial_t(g_{hl} R^h_{ijk}) = g_{hl}\partial_t R^h_{ijk} + \partial_t(g_{hl}) R^h_{ijk} \\ &= g_{hl}[\partial_i \partial_t \Gamma^h_{jk} - \partial_j \partial_t \Gamma^h_{ik}] - 2R_{hl} R^h_{ijk}.\end{aligned}$$

By (4), the above becomes

$$\begin{aligned}\partial_t R_{ijkl} = g_{hl}\big[&\partial_i\big( - g^{hm}(\nabla_j R_{km} + \nabla_k R_{jm} - \nabla_m R_{jk})\big) \\ &- \partial_j\big( - g^{hm}(\nabla_i R_{km} + \nabla_k R_{im} - \nabla_m R_{ik})\big)\big] - 2R_{hl} R^h_{ijk} \\ = -&\partial_i \nabla_j R_{kl} - \partial_i \nabla_k R_{jl} + \partial_i \nabla_l R_{jk} + \partial_j \nabla_i R_{kl} + \partial_j \nabla_k R_{il} \\ &- \partial_j \nabla_l R_{ik} - 2R_{hl} R^h_{ijk} \\ \equiv -&R_{kl,ji} - R_{jl,ki} + R_{jk,li} + R_{kl,ij} + R_{il,kj} - R_{ik,lj} - 2R_{hl} R^h_{ijk}.\end{aligned}$$

Here, to simplify computation, we have used the notation $R_{kl,ji}$ to denote the second covariant derivative $\nabla_i \nabla_j R_{kl}$, etc.

Paring the first and fourth term on the right-hand side and using the Ricci identity in Proposition 3.2.1, we arrive at

$$\begin{aligned}\partial_t R_{ijkl} &= -R_{jl,ki} + R_{jk,li} + R_{il,kj} - R_{ik,lj} \\ &\quad + R^q_{ijk} R_{ql} + R^q_{ijl} R_{kq} - 2R_{hl} R^h_{ijk} \\ &= -R_{jl,ki} + R_{jk,li} + R_{il,kj} - R_{ik,lj} \\ &\quad + R_{ijkp} g^{pq} R_{ql} + R_{ijlp} g^{pq} R_{kq} - 2R_{ql} g^{qp} R_{ijkp}.\end{aligned}$$

By the antisymmetry of $Rm$, this becomes,

$$\begin{aligned}\partial_t R_{ijkl} &= -R_{jl,ki} + R_{jk,li} + R_{il,kj} - R_{ik,lj} \\ &\quad - g^{pq}(R_{ijkp} R_{ql} + R_{ijpl} R_{kq}).\end{aligned} \tag{5.1.14}$$

Next we compute $\Delta R_{ijkl}$. By the second Bianchi identity in Proposition 3.2.2

$$\begin{aligned}\Delta R_{ijkl} &= g^{pq} \nabla_p \nabla_q R_{ijkl} = g^{pq} R_{ijkl,qp} \\ &= g^{pq} R_{qjkl,ip} - g^{pq} R_{qikl,jp}.\end{aligned} \tag{5.1.15}$$

We are going to switch the order of differentiation on the right-hand side. Indeed, by Ricci identity and using $g^{pq}R_{ipqm} = R_{im}$,

$$\begin{aligned} g^{pq}R_{qjkl,ip} &= g^{pq}R_{qjkl,pi} + g^{pq}g^{mn} \\ &\quad \times (R_{ipqm}R_{njkl} + R_{ipjm}R_{qnkl} + R_{ipkm}R_{qjnl} + R_{iplm}R_{qjkn}) \\ &= g^{pq}R_{qjkl,pi} + g^{mn}R_{im}R_{njkl} + g^{pq}g^{mn}R_{ipmj}(R_{qkln} + R_{qlnk}) \\ &\qquad\qquad \text{(1st Bianchi identity)} \\ &\quad + g^{pq}g^{mn}R_{ipkm}R_{qjnl} + g^{pq}g^{mn}R_{iplm}R_{qjkn}. \end{aligned}$$

Therefore

$$g^{pq}R_{qjkl,ip} = g^{pq}R_{qjkl,pi} + g^{mn}R_{im}R_{njkl} - B_{ijkl} + B_{ijlk} - B_{ikjl} + B_{iljk}.$$

By the second Bianchi identity

$$g^{pq}R_{qjkl,p} = g^{pq}R_{klqj,p} = -g^{pq}R_{lpqj,k} - g^{pq}R_{pkqj,l} = -R_{lj,k} + R_{kj,l}.$$

Combining the last two equalities, we obtain

$$g^{pq}R_{qjkl,ip} = -R_{lj,ki} + R_{kj,li} + g^{mn}R_{im}R_{njkl} - B_{ijkl} + B_{ijlk} - B_{ikjl} + B_{iljk}.$$

Switching $i$ and $j$, we also have

$$g^{pq}R_{qikl,jp} = -R_{li,kj} + R_{ki,lj} + g^{mn}R_{jm}R_{nikl} - B_{jikl} + B_{jilk} - B_{jkil} + B_{jlik}.$$

Using the symmetries

$$B_{ijkl} = B_{klij} = B_{jilk}$$

we deduce, from (5.1.15)

$$\begin{aligned} \Delta R_{ijkl} &= g^{pq}R_{qjkl,ip} - g^{pq}R_{qikl,jp} \\ &= -R_{ki,lj} + R_{kj,li} - R_{lj,ki} + R_{li,kj} \\ &\qquad - 2(B_{ijkl} - B_{ijlk} + B_{ikjl} - B_{iljk}) \\ &\qquad - g^{pq}R_{pikl}R_{jq} + g^{pq}R_{pjkl}R_{iq}. \end{aligned}$$

The curvature evolution formula is proven by combining this and (5.1.14), and by noticing that all terms involving second order derivatives on the right-hand side cancel.

Proof of (6). Here we derive the evolution formula for the Ricci tensor $R_{ij}$. From the computation at the beginning of part (5),

$$\begin{aligned} \partial_t R_{ij} &= \partial_t R^p_{pij} = (\partial_t \Gamma^p_{ij})_{,p} - (\partial_t \Gamma^p_{pj})_{,i} \\ &= -g^{pl}(R_{jl,ip} + R_{il,jp} - R_{ij,lp}) + g^{pl}(R_{jl,pi} + R_{pl,ji} - R_{pj,li}) \\ &= \Delta R_{ij} - g^{pl}(R_{jl,ip} + R_{il,jp}) + R_{,ij}. \end{aligned} \tag{5.1.16}$$

Here we just used that $g^{pl}(R_{jl,pi} - R_{pj,li}) = 0$. From the Ricci identity again

$$\begin{aligned} g^{pl} R_{jl,ip} &= g^{pl} R_{jl,pi} + g^{pl} R^q_{ipj} R_{ql} + g^{pl} R^q_{ipl} R_{jq} \\ &= g^{pl} R_{jl,pi} + g^{pl} g^{qm} R_{ipjm} R_{ql} + g^{pl} g^{qm} R_{iplm} R_{jq} \\ &= g^{pl} R_{jl,pi} + g^{pl} g^{qm} R_{ipjm} R_{ql} + g^{pq} R_{ip} R_{jq}. \end{aligned}$$

Switching $i$ and $j$, we have

$$g^{pl} R_{il,jp} = g^{pl} R_{il,pj} + g^{pl} g^{qm} R_{jpim} R_{ql} + g^{pq} R_{jp} R_{iq}.$$

Combining the last two identities and using the twice contracted Bianchi identity again, we arrive at

$$\begin{aligned} & g^{pl} R_{jl,ip} + g^{pl} R_{il,jp} \\ & \quad = g^{pl} R_{jl,pi} + g^{pl} R_{il,pj} + 2 g^{pl} g^{qm} R_{ipjm} R_{ql} + 2 g^{pq} R_{ip} R_{jq} \\ & \quad = R_{,ij} + 2 g^{pl} g^{qm} R_{ipjm} R_{ql} + 2 g^{pq} R_{ip} R_{jq}. \end{aligned} \tag{5.1.17}$$

Here we have used the fact that $g^{pl} g^{qm} R_{ipjm} R_{ql} = g^{pl} g^{qm} R_{jpim} R_{ql}$, which is proven by rearranging indices and using the symmetry $R_{ipjm} = R_{jmip}$. Plugging (5.1.17) to (5.1.16), we conclude

$$\partial_t R_{ij} = \Delta R_{ij} - 2 g^{pl} g^{qm} R_{ipjq} R_{lm} - 2 g^{pq} R_{ip} R_{jq}.$$

This proves (6).

The evolution equation for the scalar curvature in (1) follows immediately from (6). Indeed

$$\begin{aligned} \partial_t R &= \partial_t (g^{jk} R_{jk}) = g^{jk} \partial_t R_{jk} + R_{jk} \partial_t g^{jk} = g^{jk} \partial_t R_{jk} + R_{jk} 2 R^{jk} \\ &= g^{jk} (\Delta R_{jk} + 2 g^{pr} g^{qs} R_{pjkq} R_{rs} - 2 g^{pq} R_{pj} R_{qk}) + 2 |Ric|^2 \\ &= \Delta R + 2 |Ric|^2. \end{aligned}$$

Here the last line is obtained by taking $g^{jk} = \delta^{jk}$.

Proof of (7). Using the same local coordinates as in (4), by (3.3.5),

$$\Delta u = \partial_i \left( g^{ij} \partial_j u \right) + \frac{1}{\sqrt{g}} \partial_i \sqrt{g} g^{ij} \partial_j u.$$

Here $g = g(t)$ and $\Delta = \Delta_{g(t)}$. Differentiating with respect to $t$, we have

$$\frac{d\Delta}{dt} u = \partial_i \left( \partial_t g^{ij} \partial_j u \right) + \partial_i \left( \frac{\partial_t \sqrt{g}}{\sqrt{g}} \right) g^{ij} \partial_j u + \frac{1}{\sqrt{g}} \partial_i \sqrt{g} \partial_t g^{ij} \partial_j u.$$

Since $\partial_t g^{ij} = 2R^{ij}$ and $\partial_t \sqrt{g} = -R\sqrt{g}$ by (2), we deduce

$$\frac{d\Delta}{dt}u = 2\partial_i \left(R^{ij}\partial_j u\right) - \partial_i(R)g^{ij}\partial_j u + 2\frac{1}{\sqrt{g}}\partial_i\sqrt{g}R^{ij}\partial_j u.$$

When $t = t_0$, at $p$, the center of the normal coordinates $g^{ij} = \delta^{ij}$ and $\partial_i\sqrt{g} = 0$. Therefore

$$\frac{d\Delta}{dt}u = 2\partial_i \left(R^{ij}\partial_j u\right) - \partial_i R\partial_i u = 2R^{ij}\partial_i\partial_j u + 2\partial_i R^{ij}\partial_j u - \partial_i R\partial_i u.$$

At $t_0$ and $p$, by the contracted Bianchi identity (Proposition 3.2.3)

$$2\partial_i R^{ij} = 2\partial_i(g^{ik}g^{jl}R_{kl}) = 2g^{ik}g^{jl}\partial_i R_{kl} = 2\partial_i R_{ij} = \partial_j R.$$

Hence

$$\frac{d\Delta}{dt}u = 2R^{ij}\partial_i\partial_j u = 2 < Ric, Hess u > . \quad \square$$

$\square$

The evolution equations for the curvature tensor and Ricci curvature can be significantly simplified by choosing local coordinates which move with time suitably. This method is often referred to as Uhlenbeck's trick. We introduce

**Proposition 5.1.2** *(evolving orthonormal frame) Let $\{x^1, \ldots, x^n\}$ be a time independent local coordinate. Suppose $F_a^i$, $a, i = 1, \ldots, n$, are smooth functions satisfying the equation*

$$\partial_t F_a^i(x,t) = g^{ij}(x,t)R_{jk}(x,t)F_a^k(x,t).$$

*Here $g_{ij}(t)$ is a Ricci flow. Define vector fields*

$$F_a = F_a^i \frac{\partial}{\partial x^i}, \qquad a = 1, \ldots, n.$$

*Then the following properties hold:*

*(a). Suppose $\{F_a\}$ at $t = 0$ is an orthonormal frame with respect to $g(0)$. Then $\{F_a\}$ at $t$ is an orthonormal frame with respect to $g(t)$.*

*(b). The local expression of the pull back metric*

$$h_{ab} \equiv g_{ij}F_a^i F_b^j \tag{5.1.18}$$

*is independent of time. (c). $\nabla_i F_b^j = 0$, $\nabla_i h_{ab} = 0$. Here $\nabla_i = \nabla_{\partial/\partial x^i}$.*

*(d). $\Delta R_{abcd} = g^{ij}\nabla_i\nabla_j R_{abcd} = g^{ij}F_a^k F_b^l F_c^m F_d^n \nabla_i\nabla_j R_{klmn}$.*

The proof of the Proposition is an easy exercise of differentiation in time.

**Exercise 5.1.1** *Prove this proposition.*

The frame $\{F_a\}$ is called *evolving orthonormal frame.* We will use indices $a, b, c, d, e, f$ to denote component of curvature tensors under the frame.

**Proposition 5.1.3** *In an evolving orthonormal frame $\{F_a\}$, the evolution equations of curvature and Ricci tensors are:*

$$\partial_t R_{abcd} = \Delta R_{abcd} + 2(B_{abcd} - B_{abdc} + B_{acbd} - B_{adbc}) \tag{5.1.19}$$

*where*

$$B_{abcd} = -R_{aebf}R_{cedf}. \tag{5.1.20}$$

$$\partial_t R_{ab} = \Delta R_{ab} + 2R_{cabd}R_{cd}.$$

**Remark 5.1.1** *We are not lowering or raising the indices when doing summation since the frame is orthonormal.*

Proof. We just prove the first equation. The second one is similar. Let $\{x^1, \ldots, x^n\}$ be a time independent local coordinate. By definition and using Proposition 5.1.1 (5),

$$\begin{aligned}
\partial_t R_{abcd} = \partial_t(F_a^i F_b^j F_c^k F_d^l R_{ijkl}) &= F_a^i F_b^j F_c^k F_d^l \partial_t R_{ijkl} \\
&\quad + (\partial_t F_a^i) F_b^j F_c^k F_d^l R_{ijkl} + \cdots + F_a^i F_b^j F_c^k (\partial_t F_d^l) R_{ijkl} \\
&= F_a^i F_b^j F_c^k F_d^l \Delta R_{ijkl} + 2F_a^i F_b^j F_c^k F_d^l (B_{ijkl} - B_{ijlk} - B_{iljk} + B_{ikjl}) \\
&\quad - g^{pq} F_a^i F_b^j F_c^k F_d^l (R_{pjkl}R_{qi} + R_{ipkl}R_{qj} + R_{ijpl}R_{qk} + R_{ijkp}R_{ql}) \\
&\quad + (\partial_t F_a^i) F_b^j F_c^k F_d^l R_{ijkl} + \cdots + F_a^i F_b^j F_c^k (\partial_t F_d^l) R_{ijkl}.
\end{aligned}$$

Using $g^{pq}R_{qi}F_a^i = \partial_t F_a^p$, etc., the last two lines in the above expression cancel. Also,

$$F_a^i F_b^j F_c^k F_d^l (B_{ijkl} - B_{ijlk} - B_{iljk} + B_{ikjl}) = B_{abcd} - B_{abdc} + B_{acbd} - B_{adbc}$$

by using the previous proposition. This gives us the desired formula. □

The evolution equation of the curvature tensor $Rm$ can be simplified further in a manner we describe now.

Let $V$ be the tangent bundle over $\mathbf{M}$. Let $\Lambda^2(V)$ be the vector bundle of two forms on $\mathbf{M}$, which is equipped with a fixed metric: for $\phi, \psi \in \Lambda^2(V)$,

$$< \phi, \psi > \equiv \phi_{ab}\psi_{ab}. \tag{5.1.21}$$

Here $(\phi_{ab})$ and $(\psi_{ab})$ are skew symmetric matrices, which represent $\phi$ and $\psi$ respectively, under the frame $\{F_1, \ldots, F_a, \ldots, F_n\}$. The vector bundle $\Lambda^2(V)$ can be regarded as a Lie algebra with the Lie bracket

$$[\phi, \psi]_{ab} \equiv \phi_{ac}\psi_{bc} - \psi_{ac}\phi_{bc}.$$

Let

$$\{\phi^\alpha\}, \quad \alpha = 1, \ldots, n(n-1)/2$$

be an orthonormal basis for $\Lambda^2(V)$ under the metric (5.1.21). Then there exists $C_\gamma^{\alpha\beta}$ such that

$$[\phi^\alpha, \psi^\beta] = C_\gamma^{\alpha\beta}\phi^\gamma. \tag{5.1.22}$$

We can regard the $Rm$ as a symmetric bilinear form on $\Lambda^2(V)$ defined by

$$Rm(\phi, \psi) = R_{abcd}\phi_{ab}\psi_{dc}. \tag{5.1.23}$$

Let $\phi^\alpha = \phi^\alpha_{ab}$. We can write

$$R_{abcd} \equiv M_{\alpha\beta}\phi^\alpha_{ab}\phi^\beta_{dc}. \tag{5.1.24}$$

Equivalently, we have $R_{abcd}\phi^\beta_{dc} = M_{\alpha\beta}\phi^\alpha_{ab}$. Hence the curvature tensor can also be treated as a symmetric operator on $\Lambda^2(V) \equiv \Lambda^2(T^*(\mathbf{M}))$.

**Definition 5.1.3** *The operator $M_{\alpha\beta} : \Lambda^2(V) \to \Lambda^2(V)$, defined by*

$$R_{abcd}\phi^\beta_{dc} = M_{\alpha\beta}\phi^\alpha_{ab}$$

*is called the curvature operator.*

According to Hamilton [Ha2], there is

**Proposition 5.1.4** *Let $R_{abcd} \equiv M_{\alpha\beta}\phi^\alpha_{ab}\phi^\beta_{dc}$. Then under the Ricci flow, $M_{\alpha\beta}$ obeys the equation*

$$\partial_t M_{\alpha\beta} = \Delta M_{\alpha\beta} + M^2_{\alpha\beta} + M^\sharp_{\alpha\beta} \tag{5.1.25}$$

*where* $M^2_{\alpha\beta} = M_{\alpha\gamma}M_{\beta\gamma}$ *is the operator square; and*

$$M^\sharp_{\alpha\beta} = (C^{\gamma\eta}_\alpha C^{\delta\theta}_\beta M_{\gamma\delta}M_{\eta\theta})$$

*is the Lie algebra square.*

PROOF. The starting point is the first equation in the statement of Proposition 5.1.3, i.e.

$$\begin{aligned}\partial_t R_{abcd} &= \Delta R_{abcd} + 2(B_{abcd} - B_{abdc}) + 2(B_{acbd} - B_{adbc}) \\ &\equiv \Delta R_{abcd} + I + II.\end{aligned} \tag{5.1.26}$$

Using the first Bianchi identity and (5.1.20), we have

$$\begin{aligned}-[B_{abcd} - B_{abdc}] = R_{aebf}R_{cedf} - R_{aebf}R_{decf} \quad &= R_{aebf}(-R_{cefd} - R_{cfde}) \\ &= R_{aebf}R_{cdef}.\end{aligned}$$

Note also

$$\begin{aligned}R_{aebf}R_{cdef} &= (-R_{abfe} - R_{afeb})R_{cdef} = R_{abef}R_{cdef} - R_{afeb}R_{cdef} \\ &= R_{abef}R_{cdef} - R_{afbe}R_{cdfe} = R_{abef}R_{cdef} - R_{aebf}R_{cdef}.\end{aligned}$$

Therefore

$$R_{aebf}R_{cdef} = \frac{1}{2}R_{abef}R_{cdef}$$

and

$$I = 2(B_{abcd} - B_{abdc}) = -R_{abef}R_{cdef} = M_{\alpha\gamma}M_{\beta\gamma}\phi^\alpha_{ab}\phi^\beta_{dc}. \tag{5.1.27}$$

Next we compute, using (5.1.22) and switching some dummy variables,

$$\begin{aligned}-[B_{acbd} - B_{adbc}] &= R_{aecf}R_{bedf} - R_{aedf}R_{becf} \\ &= M_{\gamma\delta}\phi^\gamma_{ae}\phi^\delta_{cf}M_{\eta\theta}\phi^\eta_{be}\phi^\theta_{df} - M_{\gamma\delta}\phi^\gamma_{ae}\phi^\delta_{df}M_{\eta\theta}\phi^\eta_{be}\phi^\theta_{cf} \\ &= M_{\gamma\delta}(\phi^\eta_{ae}\phi^\gamma_{be} + C^{\gamma\eta}_\alpha\phi^\alpha_{ab})\phi^\delta_{cf}M_{\eta\theta}\phi^\theta_{df} \quad \text{(after summing } e) \\ &\qquad - M_{\eta\theta}\phi^\eta_{ae}\phi^\theta_{df}M_{\gamma\delta}\phi^\gamma_{be}\phi^\delta_{cf} \quad \text{(after switching } \gamma,\eta;\delta,\theta) \\ &= M_{\gamma\delta}C^{\gamma\eta}_\alpha\phi^\alpha_{ab}\phi^\delta_{cf}M_{\eta\theta}\phi^\theta_{df} \\ &= M_{\gamma\delta}M_{\eta\theta}C^{\gamma\eta}_\alpha\phi^\alpha_{ab}\phi^\delta_{cf}\phi^\theta_{df}.\end{aligned}$$

Using (5.1.22) again we have

$$
\begin{aligned}
&M_{\gamma\delta}M_{\eta\theta}C_{\alpha}^{\gamma\eta}\phi_{ab}^{\alpha}\phi_{cf}^{\delta}\phi_{df}^{\theta}\\
&= M_{\gamma\delta}M_{\eta\theta}C_{\alpha}^{\gamma\eta}\phi_{ab}^{\alpha}(\phi_{cf}^{\theta}\phi_{df}^{\delta} + C_{\beta}^{\delta\theta}\phi_{cd}^{\beta})\\
&= M_{\gamma\theta}M_{\eta\delta}C_{\alpha}^{\gamma\eta}\phi_{ab}^{\alpha}\phi_{cf}^{\delta}\phi_{df}^{\theta} + M_{\gamma\delta}M_{\eta\theta}C_{\alpha}^{\gamma\eta}\phi_{ab}^{\alpha}C_{\beta}^{\delta\theta}\phi_{cd}^{\beta}\\
&\qquad\qquad\qquad\qquad\text{(switching } \delta,\ \theta \text{ in the 1st term)}\\
&= -M_{\gamma\delta}M_{\eta\theta}C_{\alpha}^{\gamma\eta}\phi_{ab}^{\alpha}\phi_{cf}^{\delta}\phi_{df}^{\theta} + C_{\alpha}^{\gamma\eta}C_{\beta}^{\delta\theta}M_{\gamma\delta}M_{\eta\theta}\phi_{ab}^{\alpha}\phi_{cd}^{\beta}.\\
&\qquad\qquad\qquad\qquad\text{(switching } \gamma,\ \eta \text{ in the 1st term)}
\end{aligned}
$$

Thus

$$-[B_{acbd} - B_{adbc}] = M_{\gamma\delta}M_{\eta\theta}C_{\alpha}^{\gamma\eta}\phi_{ab}^{\alpha}\phi_{cf}^{\delta}\phi_{df}^{\theta} = \frac{1}{2}C_{\alpha}^{\gamma\eta}C_{\beta}^{\delta\theta}M_{\gamma\delta}M_{\eta\theta}\phi_{ab}^{\alpha}\phi_{cd}^{\beta}$$

and

$$II = 2(B_{acbd} - B_{adbc}) = C_{\alpha}^{\gamma\eta}C_{\beta}^{\delta\theta}M_{\gamma\delta}M_{\eta\theta}\phi_{ab}^{\alpha}\phi_{dc}^{\beta}. \tag{5.1.28}$$

By this, (5.1.27) and (5.1.26), we arrive at the stated equation (5.1.25). □

In the 3-dimension case the proposition takes a very nice form.

**Corollary 5.1.1** *If the manifold has dimension* 3*, then*

$$\partial_t M_{\alpha\beta} = \Delta M_{\alpha\beta} + M_{\alpha\beta}^2 + M_{\alpha\beta}^{\sharp}$$

*where $M^{\sharp}$ is the adjoint matrix of $(M_{\alpha\beta})$.*

**Exercise 5.1.2** *Prove the corollary. Hint: Calculate $C_{ab}^{\alpha}$ in an orthonormal system for* 3 *by* 3 *skew symmetric matrices and use the fact that Lie bracket is the cross product.*

Obviously this corollary makes the evolution equation for the curvature tensor look more tangible. We will utilize this corollary in the next section, when we present the Hamilton-Ivey pinching theorem. A similar but more complicated equation for the curvature tensor holds in higher dimensional cases.

The next result, according to Perelman [P1], contains two differential inequalities about the distance function under Ricci flow. The first one can be regarded as a time dependent version of the classical Laplace comparison theorem. Interestingly, it requires less assumption than the classical one, due to cancellation induced by the Ricci flow. They are very useful when one does space time localization. Before presenting the result, we fix a notation to be used. Let $(\mathbf{M}, g(t))$ be a Ricci flow, $x \in \mathbf{M}$ and $r > 0$. Then $B(x, r, t)$ denotes the geodesic ball centered at $x$ with radius $r$, with respect to the metric $g(t)$.

**Proposition 5.1.5** *Let $(\mathbf{M}, g(t))$ be a Ricci flow on an $n$ dimensional manifold $\mathbf{M}$. Let $x_0 \in \mathbf{M}$ and $t_0$ be a time during the Ricci flow.*

*(i) Suppose $Ric(\cdot, t_0) \le (n-1)K$ on $B(x_0, r_0, t_0)$ for some positive constants $K$ and $r_0$. Then the distance function $d(x, x_0, t)$ satisfies, in the weak sense, at time $t_0$ and outside $B(x_0, r_0, t_0)$, the inequality*

$$\Delta d - \partial_t d \le (n-1)\left(\frac{2}{3}K r_0 + r_0^{-1}\right).$$

*(ii) Suppose $Ric(\cdot, t_0) \le (n-1)K$ on $B(x_0, r_0, t_0) \cup B(x_1, r_0, t_0)$ for some $x_0, x_1 \in \mathbf{M}$. Then, at $t = t_0$,*

$$\partial_t d(x_0, x_1, t) \ge -2(n-1)\left(\frac{2}{3}K r_0 + r_0^{-1}\right).$$

*If the distance function is not differentiable, then the left-hand side is understood as a forward difference quotient.*

PROOF. (i) We just assume $x$ is not a cut point of $x_0$ so that $d(x, x_0, t)$ is smooth with respect to $x$. If the opposite happens, we just apply a well-known trick by Calabi to obtain the desired inequality in the weak sense.

From Proposition 3.4.6, we know that

$$\Delta d(x, x_0, t) = \Sigma_{i=1}^{n-1} \int_0^L [|X_i'|^2 + R(X, X_i, X, X_i)] ds, \qquad (5.1.29)$$

where $L = d(x, x_0, t)$. The vector fields $X, X_i$ are defined in the following way as usual. Let $c : [0, L] \to \mathbf{M}$ be a shortest geodesic connecting $x_0$ and $x$, parameterized by arclength, under the metric $g_{ij}(t_0)$. Define $X(0) = c'(0)$ and choose $e_i$, $i = 1, \ldots, n-1$ so that $\{e_1, \ldots, e_{n-1}, X(0)\}$ forms an orthonormal basis of $T_{x_0}\mathbf{M}$. We use parallel transport of the above basis to form $\{e_1(s), \ldots, e_{n-1}(s), X(s)\}$, which is an orthonormal basis for $T_{c(s)}\mathbf{M}$, $s \in [0, L]$. At last $X_i$, $i = 1, \ldots, n-1$ is the Jacobi field along $c = c(s)$ such that $X_i(0) = 0$ and $X_i(L) = e_i(L)$.

Notice the right-hand side of (5.1.29) is the sum of index forms $I(X_i, X_i)$. Since $X_i$ is a Jacobi field, the index theorem (Theorem 3.4.2) states that $I(X_i, X_i) \le I(Y_i, Y_i)$, where

$$Y_i(s) = \begin{cases} \frac{s}{r_0} e_i(s), & s \in [0, r_0], \\ e_i(s), & s \in [r_0, L]. \end{cases} .$$

Hence

$$\begin{aligned}
\Delta d(x, x_0, t_0) &\le \Sigma_{i=1}^{n-1} \int_0^L [|Y_i'|^2 + R(X, Y_i, X, Y_i)] ds \\
&= \int_0^{r_0} \frac{1}{r_0^2} (n - 1 - s^2 Ric(X, X)) ds - \int_{r_0}^L Ric(X, X) ds \\
&= -\int_0^L Ric(X, X) ds + \int_0^{r_0} \left[ \frac{n-1}{r_0^2} + \left(1 - \frac{s^2}{r_0^2}\right) Ric(X, X) \right] ds \\
&\le -\int_0^L Ric(X, X) ds + (n-1) \left( \frac{2}{3} K r_0 + r_0^{-1} \right) \\
&= \partial_t d(x, x_0, t)|_{t=t_0} + (n-1) \left( \frac{2}{3} K r_0 + r_0^{-1} \right).
\end{aligned}$$

Here we have used the fact that $\{Y_1, \ldots, Y_{n-1}, X\}$ is an orthonormal basis when $s \in [r_0, L]$ and hence $\Sigma_{i=1}^{n-1} R(X, Y_i, Y_i, X) = Ric(X, X)$ there. This proves the inequality in (i).

(ii) The proof is similar to that of (i). We need to divide the proof into three cases depending on the distance between $x_0$ and $x_1$. We assume without loss of generality that $x_1$ is not a cut point of $x_0$. Let $c$ be a minimal geodesic connecting $x_0$ and $x_1$, parametrized by arclength.

Case 1. $d(x_0, x_1, t_0) \ge 2r_0$.

This time we define the vector fields $Y_i$, $i = 1, \ldots, n-1$ by

$$Y_i(s) = \begin{cases} \frac{s}{r_0} e_i(s), & s \in [0, r_0], \\ e_i(s), & s \in [r_0, d(x_0, x_1, t_0) - r_0] \\ \frac{d(x_0, x_1, t_0) - s}{r_0} e_i(s), & s \in [d(x_0, x_1, t_0) - r_0, d(x_0, x_1, t_0)]. \end{cases} .$$

Since the geodesic is minimum, the second variation formula of distance tells us that $I(Y_i, Y_i) \ge 0$, i.e.

$$\int_0^L [|Y_i'|^2 + R(X, Y_i, X, Y_i)] ds \ge 0$$

where $X(s) = c'(s)$. This shows, together with the property that $e_i'(s) = \nabla_{c'(s)} e_i(s) = 0$,

$$\int_0^{r_0} \frac{s^2}{r_0^2} Ric(X,X)ds + \int_{r_0}^{d(x,x_1,t_0)-r_0} Ric(X,X)ds + \int_{d(x,x_1,t_0)-r_0}^{d(x,x_1,t_0)} \frac{[d(x_0,x_1,t_0)-s]^2}{r_0^2} Ric(X,X)ds \le \frac{2(n-1)}{r_0}.$$

Therefore

$$\begin{aligned}
\partial_t d(x_0,x_1,t)|_{t=t_0} &= -\int_0^{d(x_0,x_1,t_0)} Ric(X,X)ds \\
&\ge -\int_0^{r_0} \left(1-\frac{s^2}{r_0^2}\right) Ric(X,X)ds \\
&\quad - \int_{d(x_0,x_1,t_0)-r_0}^{d(x_0,x_1,t_0)} \left(1-\frac{[d(x_0,x_1,t_0)-s]^2}{r_0^2}\right) Ric(X,X)ds - \frac{2(n-1)}{r_0} \\
&\ge -2(n-1)\left(\frac{2}{3}Kr_0 + r_0^{-1}\right).
\end{aligned}$$

Case 2. $\frac{2}{\sqrt{2K/3}} \le d(x_0,x_1,t_0) \le 2r_0$.

We just need to rerun the proof of Case 1 with $r_0$ replaced by $r_1 = \frac{1}{\sqrt{2K/3}}$ which is less than $r_0$.

Case 3. $d(x_0,x_1,t_0) \le \min\{2r_0, \frac{2}{\sqrt{2K/3}}\}$.

The proof is almost trivial:

$$\begin{aligned}
\partial_t d(x_0,x_1,t)|_{t=t_0} &= -\int_0^{d(x_0,x_1,t_0)} Ric(X,X)ds \ge -(n-1)K\frac{2}{\sqrt{2K/3}} \\
&\ge -2(n-1)\left(\frac{2}{3}Kr_0 + r_0^{-1}\right).
\end{aligned}$$

□

## 5.2 Maximum principles under Ricci flow

We present in this section a number of maximum principles for tensors or curvatures under Ricci flow. These are important tools since they allow us to control qualitative behavior of the tensor in a future time, based on the information at current time.

Throughout the section, we are working under the following environment. Let $D$ be a bounded, connected domain of a complete $n$ dimensional manifold $M$, and $g = g(t)$ be a smooth Ricci flow on $D \times [0, T]$. Let $V$ be a $J$ dimensional vector bundle over $D$ with a time independent metric $h_{\alpha\beta}$ and $\Gamma(V)$ be the space of $C^\infty$ sections of $V$. We assume that there is a time dependent connection

$$\nabla(t) : \Gamma(V) \to \Gamma(V \otimes T^*M), \quad t \in [0, T]$$

which is compatible with $h = h_{\alpha\beta}$. This means

$$\nabla_X h_{\alpha\beta} = 0 \quad i.e. \tag{5.2.1}$$

for all smooth vector field $X$ on $M$ and smooth sections $\sigma, \gamma$ of $V$, there holds

$$X[h(\sigma, \gamma)] = h(\nabla_X \sigma, \gamma) + h(\sigma, \nabla_X \gamma).$$

Let $\sigma$ be a $C^\infty$ section of $V$ over $D$. Define the time dependent Laplacian by

$$\Delta(t)\sigma = g^{pq}(x, t)\nabla_p(t)\nabla_q(t)\sigma, \quad \nabla_p(t)\nabla_q(t) \equiv \nabla^2_{pq}.$$

Here $(g^{pq}) = (g_{pq})^{-1}$ and $(g_{pq})$ is the metric $g = g(t)$ in a local coordinates $\{\partial_{x^1}, \ldots, \partial_{x^n}\}$. Also $\nabla_i = \nabla_{\partial/\partial x^i}$. If no confusions arise, we will omit the time variable $t$ in $\nabla$ and $\Delta$ in this section.

In many applications, the vector bundle $V$ is chosen as the one made from 2 forms on $M$. Then the curvature tensor can be viewed as a symmetric bilinear form (functional) on $V$ (see (5.1.23).) Under the evolving orthonormal frame given in Proposition 5.1.2, the time independent metric $h_{\alpha\beta}$ is given by (5.1.18).

**Theorem 5.2.1** *(Hamilton's strong maximum principle for tensor [Ha2]) Let $M_{\alpha\beta}$ be a family of smooth, symmetric bilinear forms on $V$ evolving by the equation*

$$\partial_t M_{\alpha\beta} = \Delta M_{\alpha\beta} + N_{\alpha\beta}, \quad on \quad D \times [0, T]. \tag{5.2.2}$$

*Here $\Delta$ is the rough Laplacian on tensor with respect to the metric $g = g(t)$; $N_{\alpha\beta}$ is a polynomial of $(M_{\alpha\beta})$, formed by contracting elements of $(M_{\alpha\beta})$ with the metric $h_{\alpha\beta}$. Suppose $(N_{\alpha\beta}) \geq 0$ whenever $(M_{\alpha\beta}) \geq 0$.*

*Suppose $(M_{\alpha\beta}) \geq 0$ on $D \times [0, T]$. Then there exists $\delta > 0$ such that, on the space time domain $D \times (0, \delta)$, the rank of $(M_{\alpha\beta})$ is constant. Moreover the null space of $(M_{\alpha\beta})$ is invariant under parallel translation and invariant in time, and it also lies in the null space of $(N_{\alpha\beta})$.*

PROOF. For the sake of clarity, we divide the proof into a few steps.

*Step 1.*

Define

$$l = \sup_{x \in D} \{ \text{ rank of } M_{\alpha\beta}(x, 0)\}.$$

From linear algebra, we can find a smooth function $\rho_0 = \rho_0(x) \geq 0$, being positive some where, such that

$$\Sigma_{i=1}^{J-l+1} M_{\alpha\beta}(x, 0) v_i^\alpha v_i^\beta \geq \rho_0(x)$$

for any $J-l+1$ orthonormal vectors $v_i$ in $V_x$, $i = 1, \ldots, J-l+1$. Here and later this means $h_{\alpha\beta} v_i^\alpha v_j^\beta = \delta_{ij}$.

Here $V_x$ is the fiber of $V$ at $x$. Note that $i$ is the index for the vectors and the Greek letters $\alpha$ and $\beta$ are the indices for components of the vectors.

Next we define a function $\rho = \rho(x, t)$, which is the solution of the heat equation in $D \times [0, T]$ with Dirichlet boundary condition and with initial value $\rho_0$. By the classical strong maximum principle, it is clear that $\rho(x, t) > 0$ for all $(x, t) \in D \times (0, T]$.

*Step 2.*

In this step, we prove the following inequality: for all $(x, t) \in D \times (0, T]$

$$\Sigma_{i=1}^{J-l+1} M_{\alpha\beta}(x, t) v_i^\alpha v_i^\beta \geq \rho(x, t) \tag{5.2.3}$$

for any $J-l+1$ orthonormal vectors $v_i$ in $V_x$, $i = 1, \ldots, J-l+1$.

This inequality is clearly a consequence of the inequalities: for any $\epsilon > 0$ and all $(x, t) \in D \times (0, T]$,

$$\Sigma_{i=1}^{J-l+1} M_{\alpha\beta}(x, t) v_i^\alpha v_i^\beta + \epsilon e^t \geq \rho(x, t) \tag{5.2.4}$$

for any $J-l+1$ orthonormal vectors $v_i$, $i = 1, \ldots, J-l+1$.

We argue by contradiction. Suppose (5.2.4) is not true. Then for some $\epsilon > 0$, there is a first time $t_0 > 0$, some point $x_0 \in M$ and some orthonormal vectors $v_i$ in $V_{x_0}$, $i = 1, \ldots, J-l+1$ such that

$$\Sigma_{i=1}^{J-l+1} M_{\alpha\beta}(x_0, t_0) v_i^\alpha v_i^\beta + \epsilon e^{t_0} = \rho(x_0, t_0).$$

Under the metric $g(t_0)$, we use parallel translation along geodesics emanating from $x_0$ to extend each $v_i$ to a smooth vector field in a neighborhood of $x_0$. We still denote these vector fields by $v_i$. By the assumption on $h$, for any tangent vector $X$ on $M$,

$$\nabla_X [h(v_i, v_j)] = h(\nabla_X v_i, v_j) + h(v_i, \nabla_X v_j) = 0$$

Hence it is clear that $v_i$ are still orthonormal under the metric $h$. Write

$$F(x,t) = \Sigma_{i=1}^{J-l+1} M_{\alpha\beta} v_i^\alpha v_i^\beta + \epsilon e^t - \rho.$$

By our choice, $F(x_0,t_0) = 0$ and $F(x,t) \geq 0$ when $t < t_0$ and $F(x,t_0) \geq 0$, $x \in M$. Hence

$$\partial_t F \leq 0, \quad \Delta F \geq 0.$$

Using the equations for $M_{\alpha\beta}$ and $\rho$, we have

$$\begin{aligned}\partial_t F &= \partial_t \left( \Sigma_{i=1}^{J-l+1} M_{\alpha\beta} v_i^\alpha v_i^\beta + \epsilon e^t - \rho \right) \\ &= \Sigma_{i=1}^{J-l+1} (\Delta M_{\alpha\beta} + N_{\alpha\beta}) v_i^\alpha v_i^\beta + \epsilon e^t - \Delta\rho.\end{aligned}$$

By our assumption, $(M_{\alpha\beta}) \geq 0$ implies $(N_{\alpha\beta}) \geq 0$. Therefore

$$\begin{aligned}\partial_t F &\geq \Sigma_{i=1}^{J-l+1} (\Delta M_{\alpha\beta} v_i^\alpha v_i^\beta) + \epsilon e^t - \Delta\rho \\ &= \Delta F + \epsilon e^t = \epsilon e^t > 0.\end{aligned}$$

This contradiction shows that (5.2.4) and consequently (5.2.3) is true.

Hence the rank of $M_{\alpha\beta}(x,t)$ is at least $l$ in $D \times (0,T]$. Since rank is an integer, there exists a $\delta > 0$ such that the rank of $M_{\alpha\beta}(x,t)$ is a constant in $D \times (0,\delta]$.

*Step 3.* In this step we take $t \in (0,\delta]$ where $\delta$ is given at the end of the last step. Here we prove the statement on the null space of $M_{\alpha\beta}$.

By Step 2, the rank of $M_{\alpha\beta}$ are constants, we can find smooth sections for its null space.

Let $v \in V$ be any smooth section of $V$ in the null space of $M_{\alpha\beta}$. Then we compute

$$\begin{aligned}0 &= \partial_t (M_{\alpha\beta} v^\alpha v^\beta) = \partial_t (M_{\alpha\beta}) v^\alpha v^\beta + 2 M_{\alpha\beta} v^\alpha \partial_t v^\beta \\ &= \partial_t (M_{\alpha\beta}) v^\alpha v^\beta.\end{aligned} \tag{5.2.5}$$

Also

$$\begin{aligned}0 &= \Delta(M_{\alpha\beta} v^\alpha v^\beta) \\ &= \Delta(M_{\alpha\beta}) v^\alpha v^\beta + 4 g^{pq} \nabla_p M_{\alpha\beta}\, v^\alpha \nabla_q v^\beta \\ &\quad + 2 M_{\alpha\beta} g^{pq} \nabla_p v^\alpha \nabla_q v^\beta + 2 M_{\alpha\beta} v^\alpha \Delta v^\beta.\end{aligned}$$

The last term is 0 since $v$ is in the null space of $M_{\alpha\beta}$. Hence

$$0 = \Delta(M_{\alpha\beta}) v^\alpha v^\beta + 4 g^{pq} \nabla_p M_{\alpha\beta}\, v^\alpha \nabla_q v^\beta + 2 M_{\alpha\beta} g^{pq} \nabla_p v^\alpha \nabla_q v^\beta.$$

Notice

$$0 = \nabla_p (M_{\alpha\beta} v^\alpha) = \nabla_p (M_{\alpha\beta}) v^\alpha + M_{\alpha\beta} \nabla_p v^\alpha.$$

Combining the last two identities with (5.2.5) and using the equation (5.2.2), we arrive at

$$N_{\alpha\beta}v^\alpha v^\beta + 2M_{\alpha\beta}g^{pq}\nabla_p v^\alpha \nabla_q v^\beta = 0.$$

Taking an orthonormal coordinate at any given point where the above equality holds, we have

$$N_{\alpha\beta}v^\alpha v^\beta + 2M_{\alpha\beta}\nabla_p v^\alpha \nabla_p v^\beta = 0.$$

Since both $M_{\alpha\beta}$ and $N_{\alpha\beta}$ are nonnegative, we conclude that

$$v \in null(N_{\alpha\beta}), \qquad \nabla_p v \in null(M_{\alpha\beta}), \quad p = 1, 2, \ldots$$

Hence $null(M_{\alpha\beta}) \subset null(N_{\alpha\beta})$ and $null(M_{\alpha\beta})$ is invariant under parallel translation.

*Step 4.* We prove that $null(M_{\alpha\beta})$ is invariant in time when $t \in (0, \delta]$.

This is done if can prove $\partial_t v \in null(M_{\alpha\beta})$, where $v$ is a smooth section of $V$ in the null space of $(M_{\alpha\beta})$. We again work in an orthonormal coordinates around a point $x \in M$. By the definition of second covariant derivative, for any tangent vectors $X$ and $Y$ on $M$, it holds

$$\nabla_X \nabla_Y v \equiv \nabla^2_{XY} v = \nabla_X(\nabla_Y v) - \nabla_{\nabla_X Y} v.$$

By Step 3, from $v \in null(M_{\alpha\beta})$, we know $\nabla_X v, \nabla_Y v, \nabla_{\nabla_X Y} v \in null(M_{\alpha\beta})$, which then imply

$$\nabla_p \nabla_q v \equiv \nabla^2_{pq} v \in null(M_{\alpha\beta}).$$

Therefore

$$\Delta v = g^{pq}\nabla_p \nabla_q v \in null(M_{\alpha\beta}).$$

It follows that

$$g^{pq}\nabla_p M_{\alpha\beta}\nabla_q v^\alpha = g^{pq}\nabla_p(M_{\alpha\beta}\nabla_q v^\alpha) - M_{\alpha\beta}\Delta v^\alpha = 0.$$

This shows:

$$\begin{aligned} 0 = \Delta(M_{\alpha\beta}v^\alpha) &= \Delta(M_{\alpha\beta})v^\alpha + 2g^{pq}\nabla_p M_{\alpha\beta}\nabla_q v^\alpha + M_{\alpha\beta}\Delta v^\alpha \\ &= (\Delta M_{\alpha\beta})v^\alpha. \end{aligned}$$

Thus

$$0 = \partial_t(M_{\alpha\beta}v^\alpha) = (\Delta M_{\alpha\beta} + N_{\alpha\beta})v^\alpha + M_{\alpha\beta}\partial_t v^\alpha = M_{\alpha\beta}\partial_t v^\alpha, \ i.e.$$

$$\partial_t v^\alpha \in null(M_{\alpha\beta})$$

which proves that $null(M_{\alpha\beta})$ is invariant for $t \in (0, \delta]$. □

One quick implication of the strong maximum principle is the following important

**Theorem 5.2.2** *(Hamilton [Ha2]) Let $g = g(t)$, $t \in [0, T)$, be a Ricci flow in a complete manifold with bounded curvature at each time slice. Suppose the curvature operator $M_{\alpha\beta}$ (Definition 5.1.3) of the initial metric is nonnegative. Then, in a time interval $(0, \delta] \subset (0, T)$, the image of $M_{\alpha\beta}$ is a Lie subalgebra of $so(n)$, which has constant rank and is invariant under parallel translation and invariant in time.*

Proof. Apply the weak maximum principle Theorem 4.5.2 and Theorem 5.2.1 to equation (5.1.25). We leave the detail as an exercise. □

**Exercise 5.2.1** *Prove Theorem 5.2.2.*

**Remark 5.2.1** *The rank of $M_{\alpha\beta}$ may not be a constant for all time in $(0, T)$. See Exercise 6.63 on p248 [CLN].*

The next result, called Hamilton's advanced maximum principle, allows one to control a tensor evolving by certain nonlinear heat equation via a system of ordinary differential equations. This theorem works for general evolution equations. However we will just present the Ricci flow version for the sake of clarity.

Let $g = g(t)$ be a smooth Ricci flow on $M \times [0, T]$, where $M$ is a complete manifold with bounded curvature tensor. Let $V$ be a vector bundle over $M$ with a time independent metric $h = h_{\alpha\beta}$, and a connection $\nabla = \nabla(t) = \{\Gamma^{\alpha}_{i\beta}\}$ which is compatible with $h_{\alpha\beta}$.

Let $N : V \times [0, T] \to V$ be a fiber preserving map which is continuous in the variables $(x, t)$, i.e. $N(x, \sigma, t)$ is a time dependent vector field defined on the bundle $V$ such that

$$N(x, \sigma, t) \in V_x, \quad \text{if} \quad \sigma \in V_x.$$

Here $V_x$ is the fiber of $V$ at $x$. Moreover we assume

$$|N(x, \sigma_1, t) - N(x, \sigma_2, t)| \le C_B |\sigma_1 - \sigma_2|$$

for all $x \in M$, $t \in [0, T]$ and $|\sigma_i| \le B$, $i = 1, 2$. Here $C_B$ is a positive constant depending on $B$ only.

The PDE under consideration now is

$$\Delta \sigma(x, t) - \partial_t \sigma(x, t) + u \nabla \sigma(x, t) + N(x, \sigma(x, t), t) = 0. \tag{5.2.6}$$

Here $\Delta$ is again the rough Laplacian on tensor with respect to $g = g(t)$; $u$ is a smooth, bounded vector field on $M$ and $u\nabla\sigma(x, t)$ is defined as $u^i \nabla_i \sigma(x, t)$ in a local coordinates. We assume $\sigma = \sigma(x, t)$ is a smooth solution of the PDE in $M \times [0, T]$.

**Theorem 5.2.3** *(Hamilton's advanced maximum principle [Ha2]) Let $K$ be a closed subset of $V$ satisfying:*

*(1). $K$ is invariant under parallel translation defined by the connection $\nabla(t)$ for each $t$.*

*(2). In each fiber $V_x$, the set $K_x \equiv V_x \cap K$ is closed and convex.*

*Suppose, for any $x \in M$, any solution to the ordinary differential equation*

$$\frac{d}{dt} b_x = N(x, b_x, t) \tag{5.2.7}$$

*which starts in $K_x$ at a time $t_0 \in [0, T]$ will remain in $K_x$ for all later time. Then any solution $\sigma(x,t)$ of (5.2.6), $t \in [t_0, T]$, remains in $K$ provided that $\sigma(x, t_0) \in K$ for all $x \in M$ and $\sigma$ is uniformly bounded in $M \times [t_0, T]$, under the bundle metric $h_{ab}$.*

The proof of the theorem relies on two calculus lemmas whose proofs are left as an exercise.

**Lemma 5.2.1** *Let $f : [a, b] \subset \mathbf{R} \to \mathbf{R}$ be a Lipschitz function. Assume, for some constant $C$,*

$$\frac{d^+ f}{dt} \equiv \limsup_{h \to 0^+} \frac{f(t+h) - f(t)}{h} \le C f(t),$$
$$\text{whenever} \quad f(t) \ge 0 \quad \text{on} \quad (a, b).$$

*If also $f(a) \le 0$, then $f(t) \le 0$ on $[a, b]$.*

**Lemma 5.2.2** *Let $X$ be a complete smooth manifold and $Y$ a compact subset of $X$. Let $f(x,t)$ be a smooth function on $X \times [a, b]$ with $[a, b] \subset \mathbf{R}$. Define*

$$h(t) = \sup\{f(y, t) \mid y \in Y\}.$$

*Then $h$ is Lipschitz and*

$$\frac{d}{dt} h(t) \le \sup\{\frac{\partial}{\partial t} f(y,t) \mid y \in Y,\ f(y,t) = h(t)\}, a.e.$$

**Exercise 5.2.2** *Prove the above two lemmas.*

Now we are ready to provide

Proof. (of Theorem 5.2.3) The statement of the theorem is local in space time. Without loss of generality, we may assume $K$ is compact. Suppose there is a solution $\sigma(x,t)$ to the PDE (5.2.6) which satisfies

$\sigma(x,t_0) \in K_x$ for all $x \in M$ and which goes out of $K$ at some later time $t_2$. Since $K$ is closed, there is a time $t_1 > t_0$ such that $\sigma(x,t_1) \in K_x$ for all $x \in M$, but for any $t \in (t_1,t_2]$ there exists $x$ such that $\sigma(x,t) \in K_x^c$.

To simplify the presentation, we introduce two more notations. Given $v_1, v_2 \in V_x$, $x \in M$, we use $|v_1 - v_2|$ to denote the distance between $v_1$ and $v_2$, under the time independent metric $h$; and we use $v_1 \cdot v_2$ to denote $h(v_1,v_2)$.

We consider the function

$$f(t) = \sup_{x\in M} dist(\sigma(x,t),K_x) \equiv \sup_{x\in M}\inf_{v\in K_x} |\sigma(x,t)-v|, \qquad t\in[t_1,t_2].$$

By assumptions, we know that $f$ is continuous and

$$f(t_1)=0, \quad f(t)>0, \quad t\in(t_1,t_2]. \tag{5.2.8}$$

For any $v \in \partial K_x$, let $S_v$ be the subset of $V_x$, which is consisted of outward unit normals of the supporting hyperplanes of $K_x$ at $v$. Then one can prove that

$$f(t) = \sup_{x\in M}\sup_{v\in\partial K_x}\sup_{n\in S_v} n\cdot(\sigma(x,t)-v).$$

By Lemma 5.2.2, we have

$$\frac{d^+ f(t)}{dt} \le \sup \frac{\partial}{\partial t}[n\cdot(\sigma(x,t)-v)]. \tag{5.2.9}$$

where the sup is taken over all $x \in M, v \in \partial K_x, n \in S_v$ such that $n\cdot(\sigma(x,t)-v) = f(t)$. By (5.2.6), we compute

$$\begin{aligned}&\frac{\partial}{\partial t}[n\cdot(\sigma(x,t)-v)] = n\cdot\frac{\partial}{\partial t}\sigma(x,t)\\ &= n\cdot\Delta\sigma(x,t) + n\cdot u\nabla\sigma(x,t) + n\cdot N(x,\sigma(x,t),t).\end{aligned} \tag{5.2.10}$$

By the assumption on the ODE (5.2.7), we claim that

$$v + N(x,v,t) \in C_v K_x$$

where $C_vK_x$ is the tangent cone of $K_x$ with vertex at $v$, i.e. the smallest convex cone with vertex $v$, which contains $K_x$. Here is the proof. We pick $v \in \partial K_x$ and a time $t_0$. Let $b = b(t)$ be the solution of the ODE (5.2.7) with the initial value $b(t_0) = v$. By the definition of the convex cone $C_vK_x$ and the assumption that $b(t) \in K_x \subset C_vK_x$, we know that the ray

$$\{v + s(b(t)-v)\,|\, s\in[0,\infty)\}$$

lies in $C_v K_x$. Take $s = 1/(t - t_0)$ and let $t \to t_0^+$, we know $v + N(x, v, t_0) \in C_v K_x$. Since $t_0$ is arbitrary, the claim is true.

Therefore

$$n \cdot N(x, v, t) \le 0$$

for any $n \in S_v$ and any $t \in (t_1, t_2)$. This shows

$$\begin{aligned} &n \cdot N(x, \sigma(x,t), t) \le n \cdot N(x, \sigma(x,t), t) - n \cdot N(x, v, t) \\ &\le |N(x, \sigma(x,t), t) - N(x, v, t)| \le C|\sigma(x,t) - v| \\ &= Cf(t). \end{aligned}$$

Suppose we can show that

$$n \cdot [\Delta \sigma(x,t)] \le 0, \qquad n \cdot [u \nabla \sigma(x,t)] = 0. \tag{5.2.11}$$

Then using this, (5.2.9) and (5.2.10), we will obtain

$$\frac{d^+ f(t)}{dt} \le Cf(t), \quad t \in (t_1, t_2).$$

Lemma 5.2.1 then implies $f(t) \le 0$ when $t \in (t_1, t_2)$. This contradiction with (5.2.8) will prove the theorem.

So all we need is to prove (5.2.11). Pick a small neighborhood $O_x \subset M$, containing $x$. For any $y \in O_x$, let $v_y, n_y \in K_y$ be the parallel translation of $v, n \in K_x$, taking place along the unique minimum geodesic connecting $x$ and $y$. Recall that $\nabla$ is compatible with the metric $h$ and $K$ is invariant under parallel translation. Hence we still have $v_y \in \partial K_y$ and $n_y \in S_{v_y}$ for $K_y$ at $v_y$. By the maximality of $f(t)$, we deduce

$$n_y \cdot (\sigma(y,t) - v_y) \le f(t), \qquad y \in O_x.$$

The function $n_y \cdot (\sigma(y,t) - v_y)$ thus reaches a local maximum at $y = x$. Hence

$$\frac{\partial}{\partial y^i}[n_y \cdot (\sigma(y,t) - v_y)] = 0, \quad y = x,$$

$$\Delta[n_y \cdot (\sigma(y,t) - v_y)] \le 0, \quad y = x.$$

Here $\{y^1, \ldots, y^n\}$ is a local coordinates centered at $y$. Since $n_y$ and $v_y$ are generated by parallel translation, we know that

$$\nabla n_y = \nabla v_y = 0, \quad \Delta n_y = \Delta v_y = 0, \quad y = x.$$

Using the assumption that $h$ and $\nabla$ are compatible (c.f. (5.2.1)), we see that

$$0 = \frac{\partial}{\partial y^i}[n_y \cdot (\sigma(y,t) - v_y)] = \frac{\partial}{\partial y^i} h(n_y, (\sigma(y,t) - v_y))$$

$$= h(\nabla_{\partial/\partial y^i} n_y, (\sigma(y,t) - v_y)) + h(n_y, \nabla_{\partial/\partial y^i}(\sigma(y,t) - v_y))$$

$$= n_y \cdot \nabla_{\partial/\partial y^i}\sigma(y,t), \quad y = x.$$

Hence

$$\begin{aligned} n \cdot [u(x,t)\nabla\sigma(x,t)] &= n \cdot [u^i(x,t)\nabla_{\partial/\partial x^i}\sigma(x,t)] \\ &= u^i(x,t) n \cdot \nabla_{\partial/\partial x^i}\sigma(x,t) = 0. \end{aligned}$$

Similarly, we also see that

$$n \cdot \Delta\sigma(x,t) \leq 0.$$

So we have finished the proof of (5.2.11) and the theorem. □

An important consequence of the advanced maximum principle is the

**Theorem 5.2.4** *(Hamilton Ivey pinching theorem [Ha7] and [Iv] ) Let $(M, g(t))$, $t \in [0,T]$ be a Ricci flow on a three dimensional complete (compact or noncompact) manifold with bounded curvature for each $t \in [0,T]$. Let $\nu = \nu(x,t)$, $x \in M$, $t \in [0,T]$, be the smallest eigenvalue of the curvature operator at the point $(x,t)$. Suppose at $t = 0$, $\nu \geq -1$. If $\nu = \nu(x,t) < 0$ at a point $(x,t)$, then the scalar curvature $R = R(x,t)$ satisfies*

$$R \geq -\nu[\ln(-\nu) - 3].$$

**Remark 5.2.2** *The main point of the theorem is that the scalar curvature is much larger than the absolute value of a very negative sectional curvature. This theorem is of fundamental importance in the analysis of singularities of 3 dimensional Ricci flow.*

Proof. Since the dimension of the manifold is 3, by Corollary 5.1.1, the curvature operator $M_{\alpha\beta}$ obeys the equation

$$\partial_t M_{\alpha\beta} = \Delta M_{\alpha\beta} + M^2_{\alpha\beta} + M^\sharp_{\alpha\beta}$$

where $M^\sharp$ is the adjoint matrix of $(M_{\alpha\beta})$. At a given point, say $x \in M$, we can diagonalize $M_{\alpha\beta}$ so that

$$(M_{\alpha\beta}) = diag(\lambda, \mu, \nu)$$

where $\lambda \geq \mu \geq \nu$. Accordingly

$$(M^2_{\alpha\beta}) = diag(\lambda^2, \mu^2, \nu^2),$$

$$(M^\sharp_{\alpha\beta}) = diag(\mu\nu, \lambda\nu, \lambda\mu).$$

Hence the ordinary differential equation in companion with the above curvature equation is

$$\begin{cases} \partial_t \lambda &= \lambda^2 + \mu\nu \\ \partial_t \mu &= \mu^2 + \lambda\nu \\ \partial_t \nu &= \nu^2 + \lambda\mu. \end{cases} \tag{5.2.12}$$

Let $f$ be the one variable function on $\mathbf{R}$

$$f = f(a) = a(\ln a - 3).$$

We introduce the set $K$ which is consisted of those matrices $M_{\alpha\beta}$ whose eigenvalues $\lambda \geq \mu \geq \nu$ satisfy

$$\begin{cases} \lambda + \mu + \nu \geq -3 \\ \lambda + \mu + \nu \geq f(-\nu). \end{cases} \tag{5.2.13}$$

Since the function $f$ is continuous and convex, it is easy to see that the set $K$ is closed and convex. The equation (5.2.12) infers

$$\frac{d}{dt}(\lambda + \mu + \nu) = \lambda^2 + \mu^2 + \nu^2 + \mu\nu + \lambda\nu + \lambda\mu \geq 0.$$

Hence the first condition in (5.2.13) is preserved under the ODE (5.2.12).

When $\nu < 0$, we consider the function

$$H = \frac{\lambda + \mu + \nu}{-\nu} - \ln(-\nu).$$

By direct calculation from (5.2.12), we have

$$\nu^2 \frac{dH}{dt} = -\nu^3 - \nu(\lambda^2 + \mu^2 + \lambda\mu) + (\lambda + \mu)\lambda\mu.$$

If $\mu < 0$, then

$$\nu^2 \frac{dH}{dt} = -\nu^3 - \mu^3 + (\mu - \nu)(\lambda^2 + \mu^2 + \lambda\mu) \geq -\nu^3.$$

If $\mu \geq 0$, then

$$\nu^2 \frac{dH}{dt} = -\nu^3 - \nu\mu^2 + (\mu - \nu)(\lambda^2 + \lambda\mu) \geq -\nu^3.$$

So in any case

$$\frac{dH}{dt} \geq -\nu.$$

Thus, if $\nu \leq 0$, then

$$\frac{d}{dt}(H+3) = \frac{d}{dt}\left(\frac{\lambda+\mu+\nu}{-\nu} - \ln(-\nu) + 3\right) \geq 0.$$

By our initial condition that $0 \geq \nu \geq -1$ and that $R \geq 3\nu$ when $t = 0$, we see

$$H(0) + 3 = (-\nu)^{-1}(R - (-\nu)\ln(-\nu) - 3\nu) \geq 0.$$

Therefore, whenever $t > 0$ and $\nu < 0$, it holds.

$$\frac{\lambda+\mu+\nu}{-\nu} - \ln(-\nu) + 3 \geq 0$$

which shows that $K$ is preserved by the ODE (5.2.12). The theorem now follows from the advanced maximum principle Theorem 5.2.3. □

Theorem 5.2.3 has been extended by Chow and Lu to the case when the convex set $K$ may depend on time.

**Theorem 5.2.5** *(Chow and Lu [CL2]) Let $K(t)$, $t \in [0,T]$ be closed subsets of $V$, which satisfy:*

*(1). $K(t)$ is invariant under parallel translation defined by the connection $\nabla(t)$ for each $t \in [0,T]$.*

*(2). In each fiber $V_x$, the set $K_x(t) \equiv V_x \cap K(t)$ is closed and convex for each $t \in [0,T]$.*

*(3). The space time track $\cup_{t\in[0,T]}(\partial K(t) \times \{t\})$ is a closed subset of $V \times [0,T]$.*

*Suppose, for any $x \in M$, any solution to the ordinary differential equation*

$$\frac{d}{dt}b_x = N(x, b_x, t)$$

*which starts in $K_x(t_0)$ at a time $t_0 \in [0,T]$ will remain in $K_x(t)$ for all later time. Then any solution $\sigma(x,t)$ of (5.2.6), $t \in [t_0,T]$, remains in $K(t)$ provided that $\sigma(x,t_0) \in K(t_0)$ for all $x \in M$ and $\sigma$ is uniformly bounded in $M \times [0,T]$, under the bundle metric $h_{ab}$.*

Actually Chow and Lu proved more general results which allow $\sigma$ to be certain unbounded solutions. But this result already provides a proof of an improved Hamilton-Ivey pinching theorem. It is useful when one considers long-term behavior of Ricci flow.

**Theorem 5.2.6** *( [Ha9] Theorem 4.1) Let $(M, g(t))$, $t \in [0, T]$ be a Ricci flow on a three dimensional complete (compact or noncompact) manifold with bounded curvature for each $t \in [0, T]$. Let $\nu = \nu(x, t)$, $x \in M$, $t \in [0, T]$, be the smallest eigenvalue of the curvature operator at the point $(x, t)$. Suppose at $t = 0$, $\nu \geq -1$. If $\nu = \nu(x, t) < 0$ at a point $(x, t)$, then the scalar curvature $R = R(x, t)$ satisfies*

$$R \geq -\nu[\ln(-\nu) + \ln(1 + t) - 3].$$

PROOF. With Theorem 5.2.5 in hand the proof of this theorem is almost the same as the proof of Theorem 5.2.4. One just replace the function $f(a) = a(\ln a - 3)$ there by $f_1(a) \equiv a(\ln a + \ln(1 + t) - 3)$. $\square$

We close this section by presenting

**Theorem 5.2.7** *(Hamilton rounding theorem) Let $(M, g_0)$ be a compact three manifold with positive Ricci curvature. Then the normalized Ricci flow $\frac{d}{dt}g = -2Ric + g$, $g(0) = g_0$ has a unique smooth solution for $t \in [0, \infty)$. Moreover, as $t \to \infty$, the metric $g(t)$ converges exponentially fast in every $C^k$-norm to a $C^\infty$ metric $g_\infty$ with constant positive sectional curvature.*

*Consequently, a compact three manifold with positive Ricci curvature is diffeomorphic to the three sphere $S^3$ or its quotient of finite group.*

For the original proof one can consult [Ha1]. The book [CLN] contains an updated proof. The maximum principles play a fundamental role.

## 5.3 Qualitative properties: Gradient estimates, Harnack inequalities, compactness, $\kappa$ noncollapsing

In this section we discuss a number of qualitative properties of the Ricci flow, which are consequences of the evolution equations from the last section.

The first result is due to W. X. Shi, which says that bounds on curvature tensor imply bounds on the covariant derivatives of the curvature tensor. Such a result is a generalization of the classical Bernstein gradient estimate to the Ricci flow setting. The main idea of the proof is to construct a quantity involving the norms of curvature and its derivatives. This quantity is then shown to satisfy a scalar heat type equation to which the maximum principle is applicable.

**Theorem 5.3.1** *(Shi's derivative estimate [Shi]) There exist positive constants $C_m$, $m = 1, 2, \dots$ such that if the curvature of a complete solution to Ricci flow satisfies*

$$|Rm| \le A$$

*up to time $t$ with $0 < t \le A^{-1}$, then the $m-th$ covariant derivatives satisfy*

$$|\nabla^m Rm| \le C_m A/t^{m/2}$$

*for $0 < t \le A^{-1}$. Here the tensor norms are taken with respect to the evolving metric.*

PROOF. We will just deal with the case $m = 1$ on compact manifolds. The rest can be done by induction. Denote by $Rm$ and $\nabla Rm$ the curvature tensor and its covariant derivative tensor. According to Proposition 5.1.1,

$$\partial_t Rm = \Delta Rm + Rm * Rm \tag{5.3.1}$$

where $Rm * Rm$ is an abbreviation of sums of certain tensor products. From Proposition 3.1.1, in an orthonormal coordinate $\{x^1, \dots, x^n\}$,

$$R_{i_1 \dots i_4, k} = \partial_{x_k} R_{i_1 \dots i_4} - \Sigma_{r=1}^4 \Gamma^l_{i_r k} R_{i_1 i_{r-1} l i_{r+1} i_4}.$$

Therefore, at center of the coordinate,

$$\partial_t R_{i_1 \dots i_4, k} = \partial_{x_k} \partial_t R_{i_1 \dots i_4} - \Sigma_{r=1}^4 \partial_t \Gamma^l_{i_r k} R_{i_1 i_{r-1} l i_{r+1} i_4}$$

By Proposition 5.1.1 again,

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

Since $\partial_t \Gamma^k_{ij}$ is an linear combination of components in $\nabla Rm$, we have

$$\partial_t \nabla Rm = \Delta \nabla Rm + Rm * (\nabla Rm). \tag{5.3.2}$$

By this and (5.3.1), we arrive at the inequalities

$$\partial_t |Rm|^2 \le \Delta |Rm|^2 - 2|\nabla Rm|^2 + c|Rm|^3, \tag{5.3.3}$$

$$\partial_t |\nabla Rm|^2 \le \Delta |\nabla Rm|^2 - 2|\nabla^2 Rm|^2 + c|Rm|\,|\nabla Rm|^2. \tag{5.3.4}$$

Here $c$ is a constant depending only on dimension. Picking a constant $a > 0$ and setting

$$F = t|\nabla Rm|^2 + a|Rm|^2.$$

Then

$$\partial_t F \le \Delta F + |\nabla Rm|^2 (1 + tc|Rm| - 2a) + ca|Rm|^3.$$

Taking $a \ge c + 1$, we obtain, using $t \le A^{-1}$, $|Rm| \le A$,

$$\partial_t F \le \Delta F + c_1 A^3$$

for a dimensional constant $c_1$. The maximum principle shows

$$F(x,t) \le \sup F(\cdot, 0) + c_1 A^3 t \le (a + c_1) A^2,$$

which yields

$$|\nabla Rm|^2 (x,t) \le (a + c_1) A^2 / t. \quad \square$$

By constructing suitable cut-off functions, Shi also proved a localized gradient estimate.

**Theorem 5.3.2** *(Shi's local derivative estimate [Shi]) Given $\theta > 0, A > 0$, assume the curvature of a solution to Ricci flow $(\mathbf{M}, g)$ satisfies*

$$|Rm| \le A, \qquad on \quad U \times [0, \theta A^{-1}] \tag{5.3.5}$$

*where $U$ is an open set containing $B(p, r, 0)$, $p \in \mathbf{M}$, $r > 0$. Then there exist positive constants $C_m$, $m = 1, 2, \ldots$ depending on dimension and $\theta$, such that the following statement holds:*

*the $m-th$ covariant derivatives satisfy*

$$|\nabla^m Rm(p,t)| \le C_m A(\frac{1}{r^m} + \frac{1}{t^{m/2}} + A^{m/2})$$

*for $0 < t \le \theta A^{-1}$. Here the tensor norms are taken with respect to the evolving metric.*

PROOF. We will only prove the main case $m = 1$ and refer other details to the original paper [Shi] and [Ha7] or [CZ] p. 192. See also [Tao] for a simplified proof, where he found a simpler auxiliary function than (5.3.6). The proof is divided into three steps.

We assume, without loss of generality, that

$$r \le \theta/\sqrt{A}.$$

*Step 1.* constructing a Bernstein type auxiliary function.

This time the key quantity to consider is

$$F = \frac{b}{A^4}(BA^2 + |Rm|^2)|\nabla Rm|^2. \tag{5.3.6}$$

Here $b$ and $B$ are positive constants such that $b < c/(1+B)^2$, where $c$ depends on dimension only. By straightforward computation, we can choose $c$ small so that (5.3.3) and (5.3.4) imply,

$$\partial_t F \le \Delta F - F^2 + A^2. \tag{5.3.7}$$

*Step 2.* constructing cut off functions.

We construct a smooth, spatial cut-off function $\phi$ supported in $B(p, r, 0)$ such that, for $t \in [0, \theta A^{-1}]$,

$$\phi(x) = r, \quad \text{when} \quad d(p, x, 0) \le r/2, \ 0 \le \phi \le C_0 r$$

where $C_0$ is a positive constant. Indeed one can take

$$\phi(x) = r\lambda(d(p, x, 0)/r)$$

where $\lambda$ is a suitable one variable function.

Now define the barrier function

$$H = \frac{a^2}{\phi^2} + \frac{1}{t} + A \tag{5.3.8}$$

in the space time domain $B(p, r, 0) \times [0, \theta A^{-1}]$. Note $H$ is $\infty$ on the parabolic boundary

$$\partial B(p, r, 0) \times [0, \theta A^{-1}] \cup B(p, r, 0) \times \{0\}.$$

By continuity, there exists $S \in [0, \theta A^{-1}]$ such that

$$F \le H, \quad \text{in} \quad B(p, r, 0) \times [0, S]. \tag{5.3.9}$$

Let $T$ be the supremum of the times $S$ so that (5.3.9) holds. We claim that $T = \theta A^{-1}$ when $a$ is sufficiently large. The theorem then follows from the claim.

We show that there exists a positive constant $\sigma_1$, depending on $\theta, b, B$, but independent of $a$, such that

$$|\nabla \phi|_{g(t)} \le \sigma_1, \quad \phi|\nabla^2 \phi|_{g(t)} \le \sigma_1(1+a), \qquad t \le T. \tag{5.3.10}$$

The covariant derivatives are with respect to $g(t)$ for all $t \in [0, T]$. Using the assumed curvature bound and $r \le \theta/\sqrt{A}$, it is very easy to verify the derivative bounds in (5.3.10) at time $t = 0$. If $t > 0$, one works under the evolving orthonormal frame $\{F_a^i(x,t)\frac{\partial}{\partial x^i}\}$ given in Proposition 5.1.2. Here $\{\frac{\partial}{\partial x^i}\}$ is an orthonormal frame around $p$, with respect to $g(0)$. By Proposition 5.1.2,

$$\partial_t \nabla_a \phi = \partial_t (F_a^i \nabla_i \phi) = F_a^i \nabla_i \partial_t \phi + \nabla_i \phi R_k^i F_a^k = R_{ab} \nabla_b \phi.$$

Here $R_k^i = g^{ij} R_{jk}$. From the assumed bound on curvature (5.3.5), it follows that

$$\partial_t |\nabla \phi|^2 \le CA|\nabla \phi|^2.$$

Hence, if $|\nabla \phi| \le C_1$ at $t = 0$ for some $C_1 > 0$, then

$$|\nabla \phi|^2 \le C_1^2 e^{tCA} \le e^{C\theta} C_1^2 \equiv \sigma_1^2 \tag{5.3.11}$$

for all $t \in [0, \theta/A]$.

Next, by Proposition 5.1.1 (4) and Proposition 5.1.2, there is the identity

$$\begin{aligned}\partial_t(\nabla_{ab}^2 \phi) &= \partial_t(F_a^i F_b^j \nabla_{ij}^2 \phi) = \partial_t(F_a^i F_b^j (\frac{\partial^2 \phi}{\partial x^i \partial x^j} - \Gamma_{ij}^k \frac{\partial \phi}{\partial x^k}))\\ &= R_{ac} \nabla_{bc}^2 \phi + R_{bc} \nabla_{ac}^2 \phi - (\nabla_c R_{ab} - \nabla_a R_{bc} - \nabla_b R_{ac}) \nabla_c \phi,\end{aligned}$$

which induces

$$\partial_t |\nabla^2 \phi| \le C|Rm|\,|\nabla^2 \phi| + C|\nabla Rm|\,|\nabla \phi|.$$

This, (5.3.11) and assumption (5.3.5) imply

$$\partial_t |\nabla^2 \phi| \le CA|\nabla^2 \phi| + C\sigma_1 |\nabla Rm|, \quad t \in (0, \theta/A].$$

For $t \in [0, T]$, we assumed (5.3.9) holds, i.e. $F \le H$. Therefore

$$|\nabla Rm|^2 \le \frac{A^2}{bB}(\frac{a^2}{\phi^2} + \frac{1+\theta}{t}).$$

Hence

$$\partial_t(\phi|\nabla^2\phi|) \le CA\phi|\nabla^2\phi| + CA(\frac{r(1+\theta)}{\sqrt{t}} + \frac{a}{\sqrt{bB}})\sigma_1.$$

Integrate the above inequality from 0 to $T$. Because $T \le \theta/A$ and $r \le \theta/\sqrt{A}$, we deduce

$$\phi|\nabla^2\phi| \le C_2\sigma_1(1+a).$$

Here $C_2$ is independent of $a$. Hence the bounds (5.3.10) hold after adjusting the constants.

Step 3. applying the maximum principle.

By direct calculation,

$$\begin{aligned}\Delta H - H^2 &= a^2\frac{6|\nabla\phi|^2 - 2\phi\Delta\phi}{\phi^4} - \left(\frac{a^2}{\phi^2} + \frac{1}{t} + A\right)^2 \\ &\le \frac{a^2}{\phi^4}\left(6|\nabla\phi|^2 + 2\phi|\nabla^2\phi|\right) - \left(\frac{a^4}{\phi^4} + \frac{1}{t^2} + A^2\right).\end{aligned}$$

Note that $\partial_t H = -1/t^2$, which, together with the bounds (5.3.10) implies that

$$\Delta H - H^2 \le \frac{a^2}{\phi^4}6[\sigma_1^2 + \sigma_1(1+a)] + \partial_t H - A^2 - \frac{a^4}{\phi^4} \le \partial_t H - A^2$$

when $a$ is chosen sufficiently large. Thus

$$\partial_t H \ge \Delta H - H^2 + A^2, \quad \text{in} \quad B(p,r,0)\times[0,T].$$

From this and (5.3.7), we deduce

$$\partial_t(F-H) \le \Delta(F-H) - (F+H)(F-H), \quad \text{in} \quad B(p,r,0)\times[0,T].$$

Also note that $H$ dominates $F$ on the parabolic boundary of $B(p,r,0)\times[0,T]$. In fact $H = \infty$ there. Hence the strong maximum principle for scalar heat equation tells us $F < H$ in $B(p,r,0)\times[0,T]$. Unless $T = \theta A^{-1}$, the final time when the assumption $|Rm| \le A$ is made, this is a contradiction with the assumption that $T$ is the supremum time that $F \le H$ holds. This proves the desired bound when $m = 1$.

Higher-order bounds follow from induction. □

If there is more derivative control on the initial curvature, then one has the following improved local derivative estimates. For a proof of an even more general result, see Lu and Tian [LT], Appendix B.

**Proposition 5.3.1** *Given $\theta > 0, A > 0$, let $(\mathbf{M}, g)$ be a Ricci flow defined on the time interval $[0, \theta A^{-1}]$. Let $Rm$ be the curvature tensor and $\nabla^l Rm$ the $l$-th covariant derivative of $Rm$, $l = 1, \ldots, m$. Let $U$ be an open set containing a ball $B(p, r, 0)$, $r > 0$. Assume, for some positive constants $A$ and $B$, that*

$$|Rm| \le A, \qquad on \quad U \times [0, \theta A^{-1}],$$

$$|\nabla^l Rm(x, 0)| \le \frac{B}{r^l}$$

*for $x \in U$ and $l = 1, \ldots, m$.*

*Then exist positive constants $C_m$, $m = 1, 2, \ldots$, depending on dimension, $A$, $B$ and $\theta$, such that the following statement holds:*

*for $l = 1, 2, \ldots m$,*

$$|\nabla^l Rm(x, t)| \le \frac{C_l}{r^l}$$

*where $0 < t \le \theta A^{-1}$ and $x \in B(p, r/2, 0)$. Here the tensor norms are taken with respect to the evolving metric.*

Proof. We just give a proof for the case $m = 1$. The higher order case is similar and is left as an exercise. The proof is almost identical to the previous theorem. However, since the initial data has derivative bound, we can replace the barrier function $H$ in (5.3.8) by

$$J = \frac{a^2}{\phi^2} + A$$

in $B(p, r, 0) \times [0, \theta A^{-1}]$. By choosing $a$ suitably, we deduce

$$0 = \partial_t J \ge \Delta J - J^2 + A^2, \quad \text{in} \quad B(p, r, 0) \times [0, T].$$

with $T$ being the largest time when $F \le J$. From this and (5.3.7), we deduce

$$\partial_t (F - J) \le \Delta(F - J) - (F + J)(F - J), \quad \text{in} \quad B(p, r, 0) \times [0, T].$$

By the initial bound on $\nabla Rm$ and the fact that $J = \infty$ on the sides $\partial B(p, r, 0) \times [0, \theta A^{-1}]$, we also know that $J$ dominates $F$ on the parabolic boundary of $B(p, r, 0) \times [0, \theta A^{-1}]$. Here the constant $a = a(A, B, \theta)$ in $J$ is chosen sufficiently large. The strong maximum principle tells us $F < J$ in $B(p, r, 0) \times [0, T]$, which implies that $T = \theta A^{-1}$. The desired inequality when $m = 1$ follows. □

**Exercise 5.3.1** *Prove Proposition 5.3.1 when* $m > 1$.

From the results in [LY], [Ha5] presented in last chapter, we have seen that positive solutions to the heat equation satisfy differential Harnack inequalities. The scalar curvature under Ricci flow is a solution of a nonlinear heat equation. Does it satisfy a Harnack inequality? Hamilton [Ha6] shows that the answer is yes if the curvature operator is nonnegative.

**Theorem 5.3.3** *(Harnack inequality for the scalar curvature) Let* $g(t)$, $t \in (0, T)$, *be a complete solution of the Ricci flow on a manifold* $\mathbf{M}$. *Suppose the curvature operator of* $g(t)$, $t \in (0, T)$, *is nonnegative and bounded. Then for any one form* $W_a$ *and any two forms* $U_{ab}$, *there holds, under an evolving orthonormal frame,*

$$M_{ab} W_a W_b + 2P_{abc} U_{ab} W_c + R_{abcd} U_{ab} U_{dc} \geq 0.$$

*Here*

$$M_{ab} = \Delta R_{ab} - \frac{1}{2} \nabla_a \nabla_b R + 2R_{cabd} R_{cd} - R_{ac} R_{bc} + \frac{1}{2t} R_{ab},$$

$$P_{abc} = \nabla_a R_{bc} - \nabla_b R_{ac}.$$

**Remark 5.3.1** *As one can expect, the proof of this Harnack inequality requires rather intense and technical calculation, although the underlying idea is still the maximum principle. We refer to the original paper [Ha6] and the book [Cetc] for detailed proof and motivations. Hamilton first established a similar Harnack inequality for positive solutions of the heat equation with fixed metric in [Ha5].*

*The assumption that the curvature operator is nonnegative means* $R_{abcd} U_{ab} U_{dc} \geq 0$ *for any two forms* $U$. *See also Definition 5.1.3. Note the order of the indices* $d$ *and* $c$ *are reversed on* $U$. *This is due to the contraction convention for tensors we are using. It may seem quite restrictive at the first glance. However, in three dimension case, this is sufficient for singularity analysis. Recently S. Brendle [Br] is able to relax the curvature condition to certain extent.*

In many applications, the following trace form of the theorem is enough.

**Corollary 5.3.1** *(trace Harnack inequality under Ricci flow) Under the same assumption as the theorem, for any one form* $V_a$, *it holds*

$$\partial_t R + \frac{R}{t} + 2\nabla_a R V_a + 2R_{ab} V_a V_b \geq 0.$$

PROOF. Take

$$U_{ab} = \frac{1}{2}(V_a W_b - V_b W_a)$$

and take trace over $W_a$ in the previous theorem. $\square$

Let us mention an important classification result of ancient solutions by Hamilton, which is just one application of the trace Harnack inequality.

**Theorem 5.3.4** *Let $(\mathbf{M}, g(t))$, $t \in (-\infty, 0)$ be a complete solution to the Ricci flow with nonnegative curvature operator, positive Ricci curvature. Denote by $R$ the scalar curvature. Suppose that $\sup_{M\times(-\infty,0)} R$ is attained at some point in space time. Then $(\mathbf{M}, g(t))$ is a steady gradient soliton.*

A proof can be found in [Ha4] or Theorem 10.48 [CLN]. The concept of gradient soliton is given in Definition 5.4.2.

Next we turn to the concepts of convergence and compactness for Ricci flow. They are rooted in the classical theory on convergence of manifolds developed by Cheeger, Gromov and others. They are is very useful in the analysis of singularities of Ricci flow, when one needs to extract convergent subsequence from a sequence of scaled metrics.

**Definition 5.3.1** *($C^\infty_{loc}$ convergence of manifolds) Let $(\mathbf{M}_k, g_k, p_k)$ be a sequence of marked complete Riemann manifolds. Let $B(p_k, r_k)$ be a sequence of geodesic balls in $\mathbf{M_k}$ with $r_k \to r_\infty \le \infty$ as $k \to \infty$. We say that $B(p_k, r_k)$ converges in the $C^\infty_{loc}$ topology to a marked manifold $(\mathbf{M}_\infty, g_\infty, p_\infty) = \{p \,|\, d(p_\infty, p, g_\infty) < r_\infty\}$ if there exists a sequence of exhausting open sets $O_k \subset \mathbf{M}_\infty$ containing $p_\infty$ and a sequence of diffeomorphisms $f_k : O_k \to V_k \subset B(p_k, r_k) \subset \mathbf{M_k}$ satisfying the following conditions: $f_k(p_\infty) = p_k$ and the pull-back metrics $\tilde{g}_k = (f_k)^* g_k$ converge in $C^\infty$ topology to $g_\infty$ on every compact subset of $\mathbf{M}_\infty$, when $k \to \infty$.*

*We say the marked manifolds $(\mathbf{M}_k, g_k, p_k)$ converge to $(\mathbf{M}_\infty, g_\infty, p_\infty)$ in $C^\infty_{loc}$ topology if in addition $f_k(O_k)$ contains any ball $B(p_k, r) \subset \mathbf{M_k}$ with $r > 0$ when $k$ is sufficiently large.*

**Remark 5.3.2** *This notion of convergence is also called Cheeger-Gromov, or geometric convergence.*

The above definition of convergence can be modified to fit metrics evolving by Ricci flow in the following manner.

**Definition 5.3.2** *($C^\infty_{loc}$ convergence of evolving manifolds) Let $(\mathbf{M}_k, g_k(t), p_k)$ be a sequence of marked complete Riemann manifolds where the metrics $g_k(t)$ evolve smoothly by Ricci flow in the time interval $t \in (a, b]$ with $a < 0 < b$. Let*

$$B(p_k, r_k, 0) = \{p \,|\, d(p, p_k, g_k(0)) < r_k\}$$

*be a sequence of geodesic balls in $\mathbf{M_k}$ with $r_k \to r_\infty \le \infty$ as $k \to \infty$. We say that the marked sequence $\{B(p_k, r_k, 0), g_k(t), p_k\}$, $t \in (a, b]$, converges in the $C^\infty_{loc}$ topology to an evolving marked manifold $(\mathbf{M}_\infty, g_\infty(t), p_\infty)$, $t \in (a, b]$ if there exists a sequence of exhausting open sets $O_k \subset \mathbf{M}_\infty$ containing $p_\infty$ and a sequence of diffeomorphisms*

$$f_k : O_k \to V_k \subset B(p_k, r_k, 0) \subset \mathbf{M_k}$$

*satisfying the following conditions:*

*$f_k(p_\infty) = p_k$ and the pull-back metrics $\tilde{g}_k(t) = (f_k)^* g_k(t)$ converge in $C^\infty$ topology to $g_\infty(t)$ on every compact subset of $\mathbf{M}_\infty \times (a, b]$, when $k \to \infty$.*

*We say the marked evolving manifolds $(\mathbf{M}_k, g_k(t), p_k)$ converge to $(\mathbf{M}_\infty, g_\infty(t), p_\infty)$ in $C^\infty_{loc}$ topology if in addition $f_k(O_k)$ contains any ball $B(p_k, r, 0) \subset \mathbf{M_k}$ with $r > 0$ when $k$ is sufficiently large.*

**Remark 5.3.3** *In the triple $\{B(p_k, r_k, 0), g_k(t), p_k\}$, the ball $B(p_k, r_k, 0)$ is defined in terms of the metric $g_k(0)$. But it is equipped with the evolving metric $g_k(t)$, under which all geometric objects are computed.*

In [Ha8] Hamilton proves the next compactness result for Ricci flows. It roughly says that a sequence of Ricci flow is compact in $C^\infty_{loc}$ topology if the curvature tensors are uniformly bounded on compact sets and the injectivity radii are uniformly bounded from below at certain marked points at one time level. We refer the reader to [Ha8], or Section 4.1 of [CZ] for a detailed proof.

**Theorem 5.3.5** *(Hamilton compactness theorem for Ricci flow)*

*Let $(\mathbf{M}_k, g_k(t), p_k)$, $t \in (a, b]$ with $-\infty \le a < 0 < b \le +\infty$ be a sequence of marked complete Riemann manifolds where the metrics $g_k(t)$ evolve by Ricci flow smoothly. Let $B(p_k, r_k, 0) = \{p \,|\, d(p, p_k, g_k(0)) < r_k\}$ be a sequence of geodesic balls in $\mathbf{M_k}$ with $r_k \to r_\infty \le \infty$ as $k \to \infty$. Suppose the following conditions are met:*

*(i) for every $r \in (0, r_\infty)$, there exist positive constants $C(r)$, independent of $k$ such that the curvature tensor $Rm_{g_k(t)}$ satisfies*

$$|Rm_{g_k(t)}| \le C(r), \quad on \quad B(p_k, r, 0) \times (a, b]$$

*when $k$ is sufficiently large;*

*(ii) there exists a constant $\delta > 0$ such that the injectivity radii of $\mathbf{M_k}$ at $p_k$ under the metric $g_k(0)$ satisfy the lower bound*

$$inj(\mathbf{M}_k, p_k, g_k(0)) \geq \delta > 0, \qquad k = 1, 2, \ldots.$$

*Then there exists a subsequence of evolving marked sequence $((B(p_k, r_k, 0), g_k(t), p_k)$ over $t \in (a, b]$, which converge in $C^\infty_{loc}$ topology to a marked evolving manifold $(\mathbf{M}_\infty, g_\infty(t), p_\infty)$ where the metric $g_\infty(t)$ evolves by Ricci flow smoothly in $(a, b]$.*

*Moreover, at $t = 0$, $\mathbf{M}_\infty$ is the open geodesic ball centered at $p_\infty$ with radius $r_\infty$. Further $\mathbf{M}_\infty$ is complete if $r_\infty = \infty$.*

In the study of singularity of Ricci flow, one often needs to scale the flow near singularity and take the limit. Thus the compactness theorem is quite useful if one has the requisite injectivity radius lower bound. Historically, finding such a lower bound under Ricci flow is very difficult. The following property, introduced by Perelman [P1], implies the injectivity radius lower bound (cf. Theorem 3.6.2). In the next chapter, we will prove Perelman's theorem that any Ricci flow satisfies this property in finite time.

**Definition 5.3.3** *($\kappa$ noncollapsing or $\kappa$ noncollapsed property)*

*Let $(\mathbf{M}, g(t))$ be a Ricci flow defined on the time interval $[a, b]$ and $\kappa$ be a positive number. Let $x_0 \in \mathbf{M}$, $t_0 \in [a, b]$, $r > 0$ and suppose $t_0 - r^2 \geq a$. Then $\mathbf{M}$ is $\kappa$ noncollapsing or $\kappa$ noncollapsed at $(x_0, t_0)$ at scale $r$ if $|Rm| \leq r^{-2}$ on*

$$P(x_0, t_0, r, -r^2) = \{(x, t) \,|\, x \in \mathbf{M}, t \in (t_0 - r^2, t_0),\, d(x, x_0, t) < r\}$$

*and $|B(x_0, r, t_0)|_{g(t_0)} \geq \kappa r^3$.*

## 5.4 Special solutions: Solitons, ancient solutions, singularity models

Generally speaking, explicit or special solutions to partial differential equations are very useful, in part because there are not many of them available. The fundamental solution to the heat equation in $\mathbf{R^n}$ is one such example. In the case of the Ricci flow, two classes of special solutions, namely Ricci solitons and ancient solutions, provide important clues to the structure of solutions to Ricci flow equation near singularity and the topology of the underlying manifold. An oversimplified description of a result by Perelman [P1] goes like this. When the solution of Ricci flow develops singularity at a time $T$, the norm of the curvature tensor at time $t$ will tend to infinity as $t \to T^-$. By scaling the metrics suitably, one obtains, as limiting metrics, a solution to a new Ricci flow of a new time variable, say $s$. This solution exists when $s$ is between $-\infty$ and a fixed time say $s = 0$. Therefore it is dubbed an ancient solution. Ancient solutions are already special. In three dimension case, when $s \to -\infty$, one can scale the metrics again to extract a subsequence which converges to even more special solutions: Ricci solitons. These Ricci solitons are actually generated by the gradient of a scalar function, which makes them ever more special. In fact there are only a few of them which can exist only on a few particular manifolds such as $S^3$, $S^2 \times \mathbf{R}$ etc. Perelman then proves that a three dimensional Ricci flow near singularity behaves just like these gradient Ricci solitons. Clearly this result has important topological consequences.

Let us present the formal definition of these special solutions.

**Definition 5.4.1** *(ancient solutions, $\kappa$ solutions) A smooth, compact or noncompact Ricci flow is called an ancient solution if it is complete and the time of existence is $(-\infty, T)$ for some finite number $T$.*

*A $\kappa$ solution or ancient $\kappa$ solution is an ancient solution such that:*

*(i) It is $\kappa$ noncollapsed at all scales.*

*(ii) It has nonnegative curvature operator.*

*(iii) It has bounded curvature at each time slice.*

*(iv) It is nonflat, which means the curvature tensor is not identically 0.*

A Ricci soliton $g(t) = g_{ij}(x,t)$ is, after all, a solution of the Ricci flow equation. What is special is that $g(t)$ is obtained as a pullback of the initial metric $g(0)$ by a diffeomorphism $\phi_t$, modulo a scaling function of time only. The Ricci flow equation then forces the initial metric to satisfy certain time independent equation. The diffeomorphism $\phi_t$ can not be arbitrary either. They are generated by a time independent vector field. On the other hand, any initial metric satisfying this time independent equation produces a Ricci soliton. For this reason one does not necessarily distinguish the initial value $g(0)$ with the Ricci soliton $g(t)$ it generates.

**Definition 5.4.2** *(Ricci solitons) Let* $\mathbf{M}$ *be a Riemann manifold equipped with a metric $g_0$. Let $g = g(t)$ be the Ricci flow generated* *by $g_0$ as the initial value. Suppose, for a number $\lambda \in \mathbf{R}$ and a vector field $V$, $g_0$ satisfies,*

$$2Ric + L_V g_0 + 2\lambda g_0 = 0. \tag{5.4.1}$$

*Then $g_0$ or $g = g(t)$ is called a steady, shrinking and expanding Ricci soliton respectively if $\lambda = 0$, $\lambda < 0$ and $\lambda > 0$ respectively. Here Ric is associated with $g_0$.*

*If $V$ is a gradient field, i.e. there exists a scalar function $f$ on $\mathbf{M}$ such that $V = grad\, f$, then $g_0$ or $g = g(t)$ is called a steady, shrinking and expanding gradient Ricci soliton respectively.*

**Remark 5.4.1** *By Proposition 3.1.3, for any vector fields $Y$, $Z$ on $\mathbf{M}$,*

$$(L_V g_0)(Y, Z) = g_0(\nabla_Y V, Z) + g_0(Y, \nabla_Z V).$$

*So, under a local coordinates $\{x^1, \ldots, x^n\}$, the Ricci soliton equation can be written as*

$$2R_{ij} + g_{ik}\nabla_j V^k + g_{jk}\nabla_i V^k + 2\lambda g_{ij} = 0.$$

*Here $V^k$ is given by $V = V^k \frac{\partial}{\partial x^k}$. Also $\nabla_j V^k$ is given by $\nabla_{\frac{\partial}{\partial x^j}} V = \nabla_j V^k \frac{\partial}{\partial x^k}$.*

**Remark 5.4.2** *By the formula for Hessian of $f$, for gradient Ricci solitons, $g_0$ satisfies*

$$R_{ij} + \nabla_i \nabla_j f + \lambda g_{ij} = 0.$$

**Proposition 5.4.1** *Let $g_0$ (or $g(t)$) be the Ricci soliton given in the definition. Let $\phi_t$ be the one-parameter family of diffeomorphisms generated by the vector field $(1 + 2\lambda t)^{-1}V$, i.e. $\frac{d}{dt}\phi_t(x) = (1 + 2\lambda t)^{-1}V(\phi_t(x))$, $\phi_0 = I$. Then*

$$g(t) = (1 + 2\lambda t)\phi_t^* g_0.$$

PROOF. To prove this identity, we just need to verify that its right-hand side satisfies the Ricci flow equation. By Proposition 3.1.4 (ii) and the semigroup property of $\phi_t$, for time independent vector fields $Y, Z$ on $\mathbf{M}$,

$$\begin{aligned}
&(1 + 2\lambda t)\,(\frac{d}{dt}\phi_t^* g_0)(Y, Z) = (1 + 2\lambda t)\,\phi_t^*(L_{(1+2\lambda t)^{-1}V}\, g_0)(Y, Z)\\
&= L_V g_0((\phi_t)_* Y, (\phi_t)_* Z)\\
&= -2Ric_{g_0}((\phi_t)_* Y, (\phi_t)_* Z) - 2\lambda g_0((\phi_t)_* Y, (\phi_t)_* Z), \qquad \text{by} \quad (5.4.1)\\
&= -2(\phi_t^* Ric_{g_0})(Y, Z) - 2\lambda(\phi_t^* g_0)(Y, Z)
\end{aligned}$$

This shows

$$\begin{aligned}\frac{d}{dt}[(1+2\lambda t)\phi_t^* g_0](Y,Z) &= 2\lambda\phi_t^* g_0(Y,Z) + (1+2\lambda t)\frac{d}{dt}\phi_t^* g_0(Y,Z)\\ &= 2\lambda\phi_t^* g_0(Y,Z) - (2Ric_{\phi_t^* g_0} + 2\lambda\phi_t^* g_0)(Y,Z)\\ &= -2Ric_{\phi_t^* g_0}(Y,Z) = -2Ric_{(1+2\lambda t)\phi_t^* g_0}(Y,Z).\end{aligned}$$

In the last step, we have used the fact that the Ricci tensor is invariant relative to the metric under scaling by constants. This means that under a fixed background local coordinates, which is determined by some vector fields independent of time and metrics, the local expression for the Ricci tensors of $cg$ and $g$ are the same. □

Interestingly, there are not many 3-dimensional gradient shrinking solitons which are noncollapsed. They are all classified by Perelman [P1] in

**Proposition 5.4.2** *(Classification of 3-dimensional gradient shrinking solitons)*

*Let $(M, g(t))$ be a nonflat, three dimensional gradient shrinking soliton. Assume also $(M, g(t))$ has bounded and nonnegative sectional curvature and is $\kappa$ noncollapsed at all scales for some $\kappa > 0$.*

*Then $(M, g(t))$ is either*

*(i) the standard three sphere $S^3$, or one of its metric quotients, or*

*(ii) the standard $S^2 \times \mathbf{R}$, or one of its $Z_2$ quotients.*

Note the $Z_2$ quotients are $RP^2 \times \mathbf{R}$ and the twisted product of $S^2 \tilde{\times} \mathbf{R}$ where the group $Z_2$ flips both $S^2$ and $\mathbf{R}$.

Proof. Case 1. The first case is when the sectional curvature is 0 somewhere. Consider the pull back of the soliton to its universal cover, $(\tilde{M}, \tilde{g})$. This is a simply connected (by definition of universal cover), nonflat $\kappa$ solution with 0 sectional curvature somewhere. According to Theorem 5.2.1, the pull back $(\tilde{M}, \tilde{g})$ splits as the metric product of a 2 dimensional $\kappa$ solution and $\mathbf{R}$. By [Ha3], the only 2 dimensional $\kappa$ solution are the standard $S^2$ or $RP^2$. Since $\tilde{M}$ is simply connected but $RP^2$ is not, we know that $\tilde{M} = S^2 \times \mathbf{R}$, and the radius of the 2 sphere can be chosen as $\sqrt{-2t}$ at time $t < 0$. Hence the original gradient soliton is $M = S^2 \times \mathbf{R}/\Gamma$, a metric quotient of the round cylinder. If $M$ were compact, then at a very ancient time $t$, the scalar curvature would be $c/|t|$. Since the size of $M$ in the $\mathbf{R}$ direction does not change with time, the volume of the ball with radius $\sqrt{|t|}$ would be comparable to $c|t|$, the volume of the 2 sphere at the very ancient time $t$. But this contradicts with the fact that $\kappa$ solutions are $\kappa$ noncollapsed at all scales at any time. Therefore $M$ is noncompact. Since the projection $\Gamma_2$ of $\Gamma$ on the factor $\mathbf{R}$ is an isometry group of $\mathbf{R}$ and the image of $\mathbf{R}$ under $\Gamma_2$ is noncompact, we know $\Gamma_2$ is either $\{1\}$ or $Z_2$. Thus there is a $\Gamma$ invariant cross sphere, say $S^2 \times \{0\}$, on which $\Gamma$ acts isometrically and without fixed point. This implies that $\Gamma$ itself is either $\{1\}$ or $Z_2$. Consequently $M$ is $S^2 \times \mathbf{R}$, or one of its $Z_2$ quotients.

Case 2. The second case is when the manifold $M$ is compact and has positive sectional curvature everywhere. Let $t_0$ be a very ancient time. With $(M, g(t_0))$ as the initial value, Hamilton's rounding Theorem 5.2.7 shows that $(M, g(t))$, $t > t_0$, is getting rounder as $t$ increases. This means that the ratio between highest and lowest sectional curvature tends to 1. Thus $(M, g(0))$ is rounder than $(M, g(t_0))$. But these two only differ by homothetical scaling, modulo diffeomorphism. Therefore, $M$ must be the standard $S^3$ or its metric quotients.

Case 3. The third case is when the manifold $M$ is noncompact and has positive sectional curvature everywhere. But such gradient shrinking soliton actually does not exist. The proof of this is the hardest part of the proposition.

We divide the rest of the proof into several steps.

*Step 1.* Properties of the potential function.

Without loss of generality, we can assume that the soliton becomes singular at time $t = 0$. By definition of gradient shrinking soliton, there exists a potential function $f = f(x, t)$ such that, in a local orthonormal coordinates,

$$\nabla_i \nabla_j f + R_{ij} + \frac{1}{2t} g_{ij} = 0, \qquad -\infty < t < 0. \tag{5.4.2}$$

Taking the divergence (cf. Definition 3.3.2), we deduce

$$\Delta \nabla_j f + \nabla_i R_{ij} = 0.$$

Recall from Proposition 3.2.3 and Bochner's formula:

$$\nabla_i R_{ij} = \frac{1}{2} \nabla_j R,$$

$$\Delta \nabla_j f = \nabla_j \Delta f + R_{jk} \nabla_k f = \nabla_j (-R - \frac{3}{2t}) + R_{jk} \nabla_k f.$$

Therefore

$$\nabla_i R = 2 R_{ij} \nabla_j f. \tag{5.4.3}$$

Fixing time at $t=-1$ and choosing a base point $x_0$, let $c=c(s)$ be a minimum geodesic connecting $x_0$ with a point $x$, parameterized by arclength. Write $d(x_0,x,-1)=l$ and $X(s)=c'(s)$. To save notations, we still use $f$ to denote the function $f(\cdot,-1)$. Then, from (5.4.2),

$$\frac{d^2f(c(s))}{ds^2}=\frac{d}{ds}<\nabla f,c'(s)>=<\nabla_X\nabla f,X>=-Ric(X,X)+\frac{1}{2}.$$

Therefore

$$\frac{df(c(s))}{ds}|_{s=l}=\frac{df(c(s))}{ds}|_{s=0}+\frac{1}{2}l-\int_0^l Ric(X,X)ds.$$

Since the curvature is bounded globally, by Proposition 5.1.5, there is the bound

$$\int_0^l Ric(X,X)ds\leq C$$

which is independent of $l$. Hence

$$\frac{df(c(s))}{ds}|_{s=l}\geq\frac{1}{2}l-C. \tag{5.4.4}$$

Next let $Y=Y(s)$ be a unit parallel vector field along $c=c(s)$, which is perpendicular to $X=X(s)$. Then, by equation (5.4.2) again,

$$\frac{d}{ds}Y(f(c(s)))=\frac{d}{ds}<\nabla f,Y>=<\nabla_X\nabla f,Y>=-Ric(X,Y).$$

After integration, this becomes

$$Y(f)(x)=Y(f)(x_0)-\int_0^l Ric(X,Y)ds.$$

By Cauchy-Schwarz inequality

$$\begin{aligned}&\left[\int_0^l|Ric(X,Y)|ds\right]^2\leq l\int_0^l|Ric(X,Y)|^2ds\\&\leq l\int_0^l|Ric(Y,Y)|\,|Ric(X,X)|ds\leq Cl\int_0^l|Ric(X,X)|ds\leq Cl.\end{aligned}$$

Hence

$$|Y(f)(x)|\leq C(\sqrt{l}+1). \tag{5.4.5}$$

By this and (5.4.4), we know that

$$\begin{aligned}&\frac{l}{2}-C\leq<X,\nabla f>(x)\leq\frac{l}{2}+C,\\&|<Y,\nabla f>(x)|\leq C\sqrt{l}+C,\qquad d(x_0,x,-1)=l.\end{aligned} \tag{5.4.6}$$

When $l$ is large, these two inequalities imply that $f$ has no critical point and the gradient of $f$ becomes more and more parallel to the tangent of the minimum geodesic. So the gradient shrinking soliton looks like a cylinder near infinity.

*Step 2.* We show that the scalar curvature $R$ tends to 1 at infinity.

By (5.4.6), for large $l$, it holds

$$|f(x) - \frac{l^2}{4}| \le C(l+1). \tag{5.4.7}$$

Hence the portion of $M$ near infinity, relative to $x_0$, is covered by $\cup_{a>1}^{\infty}\{x \,|\, f(x) \ge a\}$. Let $q = q(s)$ be an integral curve of the gradient flow $q'(s) = \nabla f(q(s))$. Then, by (5.4.3),

$$\frac{dR(q(s))}{ds} =< \nabla R, q'(s) >= 2Ric(\nabla f, \nabla f) > 0. \tag{5.4.8}$$

These show that

$$\bar{R} \equiv \limsup_{d(x_0,x,-1)\to\infty} R(x,-1) > 0.$$

Pick a sequence of points $(x_k, -1)$ such that $R(x_k, -1) \to \bar{R}$. By the $\kappa$ noncollapsing assumption, we can apply Hamilton's compactness Theorem 5.3.5. So, there is a subsequence, still denoted by $(M, (x_k, -1))$, of marked manifolds, which converges to an ancient solution $(\bar{M}, \bar{g})$ in $C^{\infty}_{loc}$ topology. By Proposition 7.1.1, the limit solution splits off a line, i.e. $\bar{M} = \bar{M}_2 \times \mathbf{R}$. Observe that $\bar{M}_2$ is a 2 dimensional ancient solution which is also $\kappa$ noncollapsed at all scales. By [Ha3] (Theorem 7.1.3), $\bar{M}_2$ is either the standard $S^2$ or $RP^2$. Note that $M$, with strictly positive sectional curvature, is orientable by [CG2]. Hence $\bar{M}_2 = S^2$ and $\bar{M}$ is the standard shrinking cylinder $S^2 \times \mathbf{R}$, existing in the time interval $(-\infty, 0]$. At time $t = -1$ the scalar curvature of the limiting flow is $\bar{R}$. Thus $\bar{R} \le 1$, for the limit flow would blow up before time 0 otherwise.

Since the soliton $(M, g)$ is assumed singular at time 0, Proposition 5.4.1 shows that

$$g(t) = -t\phi_t^* g(-1)$$

where $\phi_t$ is the one parameter family of diffeomorphisms generated by $\nabla f$. Hence

$$\inf R(x,t) = |t|^{-1} \inf R(x,-1) > C|t|^{-1} > 0.$$

This implies that the limit soliton $(\bar{M}, \bar{g})$ must blow up at time 0. By definition of $\bar{R}$, the limit soliton is the standard shrinking cylinder with scalar curvature $\bar{R}$ at $t = -1$. To blow up exactly at time $t = 0$, there is no other choice but $\bar{R} = 1$. We can carry out the same argument for any sequence $\{x_k\}$, tending to infinity, with the property that $R(x_k, -1)$ has a limit. Therefore

$$\lim_{d(x_0,x,-1)\to\infty} R(x,-1) = 1, \quad \text{and } R(x,-1) < 1, \qquad \text{when} \quad d(x_0,x,-1) \quad \text{large.} \tag{5.4.9}$$

The last statement is due to (5.4.8).

Step 3. We prove

$$Area\{f = a\} < 8\pi \tag{5.4.10}$$

when $a$ is large. Here $\{f = a\}$ is the level surface of $f$ at value $a$. If it has more than one connected component, then we just pick one of them.

Let $\{e_1, e_2\}$ be an orthonormal basis of $\{f = a\}$ and $e_3$ be the unit outward normal. Then $\{e_1, e_2, e_3\}$ is an orthonormal basis of $M$. Let $\nabla^2 f \equiv (f_{ij})$ be the Hessian of $f$ under this basis. For the rest of the proof, expressions involving derivatives and components of curvature tensor are exclusively under this basis.

The second fundamental form of the level surface is

$$h_{ij} \equiv < \nabla_i e_3, e_j > = < \nabla_i \frac{\nabla f}{|\nabla f|}, e_j > = \frac{f_{ij}}{|\nabla f|}, \quad i,j = 1,2. \tag{5.4.11}$$

Here we have used $< \nabla f, e_j > = 0$, $j = 1, 2$. Observe, by (5.4.2) and (5.4.9), and $t = -1$,

$$f_{ii} = \frac{1}{2} - Ric(e_i, e_i) \geq \frac{1}{2} - \frac{R}{2} > 0, \quad i = 1, 2$$

when $a$ is large. Here we also have used the property that $Rg \geq 2Ric$, which is equivalent to positivity of sectional curvature in dimension 3. This is confirmed by the following calculation in a local orthonormal coordinates which diagonalizes the Ricci curvature.

Since $Ric_{ij} = R_{ikkj}$, we have

$$Ric_{11} = R_{1221} + R_{1331}, \quad Ric_{22} = R_{2112} + R_{2332}, \quad Ric_{33} = R_{3113} + R_{3223}.$$

Therefore

$$R = Ric_{11} + Ric_{22} + Ric_{33} = 2R_{1221} + 2R_{1331} + 2R_{2332}.$$

This shows $Rg \geq 2Ric$ and

$$R - 2Ric_{33} = 2R_{1221}. \tag{5.4.12}$$

By the first variation formula for area (see Chapter 1, section 8 of [CLN] e.g.),

$$\frac{d}{da} Area\{f = a\} = \int_{\{f=a\}} \frac{f_{11} + f_{22}}{|\nabla f|} \geq \int_{\{f=a\}} \frac{1 - R}{|\nabla f|} > 0$$

when $a$ is large. So $Area\{f = a\}$ is an increasing function for large $a$. We claim

$$\lim_{a \to \infty} Area\{f = a\} = 8\pi \tag{5.4.13}$$

which then implies (5.4.10) and completes the step.

By the last step, we can pick a sequence $a_i \to \infty$ and points $x_i$ tending to infinity such that $x_i \in \{f = a_i\}$ and that $(M, x_i, g(t))$ converges to the standard cylinder with scalar curvature 1 at time $-1$. Consider the functions

$$F_i(x) \equiv 2\sqrt{f(x)} - 2\sqrt{f(x_i)}.$$

By (5.4.6), (5.4.7) and (5.4.2),

$$\begin{aligned} |\nabla F_i| = |\nabla f|/\sqrt{f} \to 1, \quad &|\nabla^2 F_i| \\ &= \left|\frac{\nabla^2 f}{f^{1/2}} - \frac{1}{2}\frac{|\nabla f|^2}{f^{3/2}}\right| \leq \frac{C}{f^{1/2}} \to 0, \quad a \to \infty. \end{aligned}$$

By (5.4.2) again, $|\nabla^3 f| = |\nabla Ric| \leq C$, which implies by straightforward calculation that $|\nabla^3 F_i| \to 0$ when $a \to \infty$. Hence the function $F_i$ converge to a $C^2$ function $F_\infty$ satisfying

$$\nabla^2 F_\infty = 0, \quad |\nabla F_\infty| = 1.$$

By standard elliptic theory, $F$ is a $C^\infty$ function defined on the limit soliton: the standard cylinder. Now we know $F_\infty$ is a radial function and $F_\infty^{-1}(a)$ are totally geodesic 2 spheres with constant curvature, which are normal to the geodesic lines. Since (each component of) $\{f = a\}$ converges to a 2 sphere of scalar curvature 1, and the scalar curvature of a sphere with radius $r$ is $2/r^2$, the radius of the 2 sphere is $\sqrt{2}$ and its area is $8\pi$. Hence the claim (5.4.13) is true, completing the proof of Step 3.

*Step 4.* Reaching a contradiction with the Gauss-Bonnet formula.

Denote by $N$ the level surface $\{f = a\}$, $X$ its unit normal vector and $R_N$ the intrinsic (Gauss) curvature, which is 1/2 of the scalar curvature. Let $h_{ij}$ be the second fundamental form. According to the Gauss-Codazzi equation, (5.4.2) and (5.4.12) in step 3, we have

$$\begin{aligned}
R_N &= R_{1221} + det(h_{ij})\\
&= \frac{1}{2}(R - 2Ric(X,X)) + \frac{f_{11}f_{22} - f_{12}^2}{|\nabla f|^2}\\
&\le \frac{1}{2}(R - 2Ric(X,X)) + \frac{1}{4|\nabla f|^2}[f_{11} + f_{22}]^2\\
&= \frac{1}{2}(R - 2Ric(X,X)) + \frac{1}{4|\nabla f|^2}[1 - (R - Ric(X,X))]^2\\
&= \frac{1}{2}\Bigg[1 - Ric(X,X) - (1 - R + Ric(X,X))\\
&\qquad + \frac{(1 - R + Ric(X,X))^2}{2|\nabla f|^2}\Bigg].
\end{aligned}$$

When $a$ is sufficiently large, we already know that $1 - R + Ric(X,X)$ is bounded and positive. Also $|\nabla f|$ is large by (5.4.6). Consequently

$$R_N < \frac{1}{2}.$$

Now by Gauss-Bonnet formula

$$4\pi = \int_N R_N dA < \frac{1}{2} Area(N)$$

so that

$$Area(N) > 8\pi.$$

This is a contradiction with (5.4.10), which proves the proposition. □

Detailed proofs can also be found in [Cetc], [CZ], [KL] and [MT]. Generalizations and improvement of this result have appeared in [Ni3], [NW], [PW], [Nab].

Like many nonlinear evolution equations, the Ricci flow may develop singularity in finite time. The analysis of singularity is important for the understanding of the equation and the underlying manifold.

**Definition 5.4.3** *(maximal solutions, almost maximum points) Let* $g(t) = g_{ij}(x,t)$ *be a Ricci flow on* $\mathbf{M} \times [0,T)$, $T \le \infty$, *where* $\mathbf{M}$ *is either compact or* $(\mathbf{M}, g(t))$ *is complete and has bounded curvature. If* *the supremum norm of Rm under $g(t)$ becomes unbounded as $t \to T$, then $g = g(t)$ is called a maximal solution.*

*A point $(x,t) \in \mathbf{M} \times [0,T)$ is called an almost maximum point if there exist positive constants $a$ and $\alpha \in (0,1]$ such that*

$$|Rm(x,t)| \geq aK_{max}(s), \qquad s \in [t - \frac{\alpha}{K_{max}(t)}, t]$$

*where*

$$K_{max}(s) \equiv \sup_{x \in \mathbf{M}} |Rm(x,s)|_{g(s)}.$$

Hamilton [Ha7] introduced the following notions for maximal solutions, which classify all maximal solutions according to the rate the curvature tensor tends to infinity.

**Definition 5.4.4** *(types of maximal solutions) A maximal solution of the Ricci flow in $[0,T)$ is called*

*Type I, if $T < \infty$ and $\sup_{[0,T)}(T-t)K_{max}(t) < \infty$;*

*Type II (a), if $T < \infty$ but $\sup_{[0,T)}(T-t)K_{max}(t) = \infty$;*

*Type II (b), if $T = \infty$ and $\sup_{[0,T)} tK_{max}(t) = \infty$;*

*Type III (a), if $T = \infty$, $\sup_{[0,T)} tK_{max}(t) < \infty$ and $\limsup_{t\to\infty} tK_{max}(t) > 0$;*

*Type III (b), if $T = \infty$, $\sup_{[0,T)} tK_{max}(t) < \infty$ and $\limsup_{t\to\infty} tK_{max}(t) = 0$.*

The next theorem states that maximal solutions with certain injectivity radius bound can be scaled to one of the singularity models given by

**Definition 5.4.5** *(singularity models) Let $g(t) = g_{ij}(x,t)$ be a Ricci flow on $\mathbf{M}$ which is either compact or $(\mathbf{M}, g(t))$ is complete with bounded curvature. Then $g(t)$ is called a singularity model if it is not flat and is one of the three types.*

*Type I: The solution exists for all $t \in (-\infty, T)$ for some constant $T \in (0,\infty)$ and*

$$|Rm(x,t)| \leq T/(T-t), \qquad (x,t) \in \mathbf{M} \times (-\infty, T)$$

*with equality holds at some $x \in \mathbf{M}$ and $t = 0$;*

*Type II: The solution exists for all $t \in (-\infty, +\infty)$ and*

$$|Rm(x,t)| \leq 1, \qquad (x,t) \in \mathbf{M} \times (-\infty, +\infty)$$

*with equality holds at some $x \in \mathbf{M}$ and $t = 0$;*

*Type III: The solution exists for all $t \in (-A, +\infty)$ for some positive constant $A$ and*

$$|Rm(x,t)| \le A/(A+t), \qquad (x,t) \in \mathbf{M} \times (-A, +\infty)$$

*with equality holds at some $x \in \mathbf{M}$ and $t = 0$;*

**Theorem 5.4.1** *(scaling of maximal solutions to singularity models) Let $\mathbf{M}$ be a compact manifold. Suppose $(\mathbf{M}, g(t))$, $t \in [0, T)$ is a maximal solution of Type I, II (a), (b) or III (a) satisfying the injectivity radius condition:*

*for any sequence of almost maximum points $\{(x_k, t_k)\}$, $t_k \to T$, $k \to \infty$, there exists a constant $c > 0$ such that*

$$inj(\mathbf{M}, x_k, g(t_k)) \ge \frac{c}{\sqrt{K_{max}(t_k)}}, \quad k = 1, 2, \ldots .$$

*Then there exists a sequence of dilations of the solution around $(x_k, t_k)$ which converges in the $C^\infty_{loc}$ topology to a singularity model of the corresponding type.*

Proof. We will just present proofs of the cases for Type I and Type II (a) maximal solutions. The rest is similar. The proofs are modeled after Theorem 4.3.4 of [CZ], which also contains the details of the remaining cases.

Type I case. Define

$$\omega \equiv \limsup_{t \to T} (T-t) K_{max}(t).$$

Observe that $\omega$ is a finite positive number. The finiteness of $\omega$ comes from the assumption of Type I maximum solution. That $\omega$ is positive is a result of the maximum principle working on the evolution equation of curvature. Indeed from Proposition 5.1.1, item (5), it is easy to see that

$$\partial_t |Rm| \le \Delta |Rm| + c|Rm|^2.$$

Since $\mathbf{M}$ is a compact manifold, $K_{max}(t)$ is reached by $|Rm|$ somewhere. Therefore

$$\partial_t K_{max} \le c K^2_{max}.$$

This implies, after integration $K_{max}(t) \ge c/(T-t) > 0$, which shows $\omega > 0$. With a little more assumptions and efforts, this argument actually works for certain noncompact manifolds too.

Now we take a sequence $\{(x_k, t_k)\}$ with $t_k \to T$ such that

$$\omega = \limsup_{k\to\infty}(T - t_k)K_{max}(t_k).$$

With the scaling factor

$$\epsilon_k = |Rm(x_k, t_k)|^{-1/2}$$

we introduce the scaled metrics

$$g^{(k)}(\cdot, \tilde{t}) = \epsilon_k^{-2} g(\cdot, t_k + \epsilon_k^2 \tilde{t}), \quad \tilde{t} \in [-t_k/\epsilon_k^2, (T - t_k)/\epsilon_k^2).$$

Evidently $g^{(k)}$ is also a Ricci flow, namely

$$\partial_{\tilde{t}} g^{(k)}(\cdot, \tilde{t}) = -2Ric_{g^{(k)}(\cdot,\tilde{t})}.$$

When $k \to \infty$, the life span of $g^{(k)}$ expands to $(-\infty, \omega)$ since

$$t_k/\epsilon_k^2 = t_k|Rm(x_k, t_k)| \to \infty, \quad (T-t_k)/\epsilon_k^2 = (T-t_k)|Rm(x_k, t_k)| \to \omega.$$

Since $\omega = \limsup_{t\to T}(T - t)K_{max}(t)$, for any $\delta > 0$, there exists $S < T$ such that

$$|Rm(x, t)| \le (\omega + \delta)/(T - t)$$

holds for all $t \in [S, T]$. When $k$ is sufficiently large,

$$[t_k - K_{max}^{-1}(t_k), t_k] \subset [S, T).$$

Hence, for $t \in [t_k - K_{max}^{-1}(t_k), t_k]$, it holds, for a sufficiently small $\delta$

$$\begin{aligned} K_{max}(t) = \sup_{x\in\mathbf{M}} |Rm(x, t)| &\le (\omega + \delta)/(T - t) \le (\omega + \delta)/(T - t_k) \\ &\le C|Rm(x_k, t_k)|. \end{aligned}$$

For $t \le S$, it holds $(T - t)^{-1} \le (T - S)^{-1} \le (T - t_k)^{-1}$ and hence we also have

$$K_{max}(t) \le C/(T - S) \le C|Rm(x_k, t_k)|.$$

These two bounds on $K_{max}$ mean $(x_k, t_k)$ is an almost maximum point for the maximal solution $g(t)$, $t \le t_k$. The assumption on the injectivity radius shows

$$inj(\mathbf{M}, x_k, g(t_k)) \ge \frac{c}{\sqrt{K_{max}(t_k)}} \ge C\epsilon_k.$$

Therefore, for the scaled metric $g^{(k)}$, it holds

$$inj(\mathbf{M}, x_k, g^{(k)}(0)) \geq C.$$

Denote by $Rm^{(k)}$ the curvature tensor of $g^{(k)}$. Then, for $\tilde{t} \in [(S - t_k)/\epsilon_k^2, (T - t_k)/\epsilon_k^2)$,

$$\begin{aligned}|Rm^{(k)}(x, \tilde{t})| &= \epsilon_k^2 |Rm(x, t)| \\ &\leq (\omega + \delta)[(T - t)\,|Rm(x_k, t_k)|]^{-1} \\ &= (\omega + \delta)[(T - t_k)\,|Rm(x_k, t_k)| + (t_k - t)\,|Rm(x_k, t_k)|]^{-1} \\ &\to (\omega + \delta)/(\omega - \tilde{t}), \qquad k \to \infty.\end{aligned}$$

This curvature bound and the above injectivity lower bound allow us to use Hamilton's compactness Theorem 5.3.5 to conclude: There exists a subsequence, still denoted by $\{g^{(k)}(\tilde{t})\}$, which converges in $C^\infty_{loc}$ topology to a limit metric $g^{(\infty)}(\tilde{t})$ on a limiting manifold $\tilde{\mathbf{M}}$. Moreover $g^{(\infty)}(\tilde{t})$ is a complete solution, existing for $\tilde{t} \in (-\infty, \omega)$, satisfying

$$|Rm^{(\infty)}(\tilde{t})| \leq \omega/(\omega - \tilde{t})$$

everywhere on $\tilde{\mathbf{M}} \times (-\infty, \omega)$ with equality holding somewhere at $\tilde{t} = 0$.

Type II (a) case.

This time $g = g(t)$ exists for $t \in [0, T)$ but $\limsup_{t \to T}(T - t)K_{max}(t) = \infty$. Pick $x_k \in \mathbf{M}$ and times $t_k$, $T_k$ such that $t_k < T_k \leq T$ and $t_k \to T$ when $k \to \infty$. We also require that, for a sequence of numbers $a_k \to 1_-$,

$$(T_k - t_k)|Rm(x_k, t_k)| \geq a_k \sup_{x \in \mathbf{M}, t \leq T_k} (T_k - t)|Rm(x, t)| \to \infty$$

when $k \to \infty$.

As in the previous case, we take

$$\epsilon_k = |Rm(x_k, t_k)|^{-1/2}$$

and define the scaled metrics

$$g^{(k)}(\cdot, \tilde{t}) = \epsilon_k^{-2} g(\cdot, t_k + \epsilon_k^2 \tilde{t}), \quad \tilde{t} \in [-t_k/\epsilon_k^2, (T_k - t_k)/\epsilon_k^2).$$

Note that, when $k \to \infty$,

$$t_k/\epsilon_k^2 = t_k|Rm(x_k, t_k)| \to \infty, \quad (T - t_k)/\epsilon_k^2 = (T - t_k)|Rm(x_k, t_k)| \to \infty.$$

This means the life span of $g^{(k)}$ tends to $(-\infty, \infty)$.

Also, for $\tilde{t} \in [-t_k \epsilon_k^{-2}, (T_k - t_k)\epsilon_k^{-2})$, using $t = t_k + \epsilon_k^2 \tilde{t}$, we deduce,

$$\begin{aligned}|Rm^{(k)}(x, \tilde{t})| &= \epsilon_k^2 |Rm(x,t)| \\ &\le a_k^{-1}(T_k - t_k)(T_k - t)^{-1} \\ &= a_k^{-1}(T_k - t_k)\, \frac{|Rm(x_k, t_k)|}{(T_k - t_k)|Rm(x_k, t_k)| - \tilde{t}} \\ &\to 1, \qquad k \to \infty.\end{aligned}$$

As in the previous case, $(x_k, t_k)$ is an almost maximum point. By Hamilton's compactness Theorem 5.3.5 again, there exists a subsequence, still denoted by $\{g^{(k)}(\tilde{t})\}$, which converges in $C^\infty_{loc}$ topology to a limit metric $g^{(\infty)}(\tilde{t})$ on a limiting manifold $\tilde{\mathbf{M}}$. Moreover $g^{(\infty)}(\tilde{t})$ is a complete solution, existing for $\tilde{t} \in (-\infty, \infty)$, satisfying

$$|Rm^{(\infty)}| \le 1$$

everywhere on $\tilde{\mathbf{M}} \times (-\infty, \infty)$ with equality holding somewhere at $\tilde{t} = 0$. □

Singularity models should be much simpler than a general Ricci flow since the former reflects the microstructure of the later. This belief is partially confirmed by the following result of Hamilton's:

**Theorem 5.4.2** *(Type II singularity model with nonnegative curvature) Any Type II singularity model with nonnegative curvature operator and positive Ricci curvature to the Ricci flow must be a steady Ricci soliton.* □

PROOF. See [Ha4] and Proposition 9.29 [CLN]. In a 3-dimension case, the Hamilton-Ivey pinching theorem Theorem 5.2.4 tells us that the curvature operator of a Type I and II singularity model is nonnegative. So, a natural question is: what can one say about the structure of these singularity models? Of course, Theorem 5.4.1 also hinges on the injectivity radius lower bound, which is only an assumption so far. These two issues had been big obstacles in the study of Ricci flow. There is also a problem about the structure of limiting solutions when the curvature at the blow up points are not comparable with maximal ones. These were all overcome by Perelman [P1] in 2002. We will start to explain Perelman's work in the next chapter.

We should mention that these issues for the special case of three manifolds with positive Ricci curvature were largely understood by Hamilton [Ha1]. See Theorem 5.2.7 here.

# Chapter 6

# Perelman's entropies and Sobolev inequality for Ricci flow, the smooth case

## 6.1 Perelman's entropies and their monotonicity

In a truly remarkable paper [P1], Perelman discovered several quantities which are monotone under Ricci flow. These are the analytical breakthroughs that led him to the proof of Poincaré and geometrization conjectures. One of the monotone quantities is the $F$ entropy. Its monotonicity implies the monotonicity of the first eigenvalue of the operator $-4\Delta + R$. This can also be regarded as a family of Poincaré inequalities which are uniform in time. The other quantity is called the $W$ entropy, which is a family of log Sobolev inequalities in disguise. Its monotonicity implies the crucial no local collapsing result under Ricci flow.

Perelman's entropies are constructed via solutions of the conjugate heat equation

$$H^* u \equiv \Delta u - Ru + \partial_t u = 0 \tag{6.1.1}$$

which is attached to the Ricci flow artificially. Here $\Delta$ is the Laplace-Beltrami operator and $R$ is the scalar curvature, both under the metric $g = g(t)$. Perelman [P1] mentioned that he was inspired by similar formulas in string theory. We would like to provide another motivation which is rooted in the classical Boltzmann entropy. Such a view has been mentioned in [Cetc] and [To] e.g. This point of view seems more natural and results in somewhat simpler proofs. Later in the section we will present Perelman's original proof for a slight generalization of the $W$ entropy.

Let $u$ be a positive solution to the conjugate heat equation. The classical Boltzmann entropy is

$$\mathbf{B} = \int_{\mathbf{M}} u \ln u d\mu(g(t)). \tag{6.1.2}$$

Our first observation is

**Proposition 6.1.1** *Let $u$ be a positive solution to (6.1.1). Then*

$$H^*(u \ln u) = \frac{|\nabla u|^2}{u} + Ru.$$

PROOF.

$$\begin{aligned} H^*(u \ln u) &= \Delta(u \ln u) - Ru \ln u + \partial_t (u \ln u) \\ &= \Delta u \;\ln u + 2\nabla u \nabla \ln u + u \Delta \ln u - Ru \ln u + \partial_t u \;\ln u + \partial_t u \\ &= 2\frac{|\nabla u|^2}{u} + u div(\frac{\nabla u}{u}) + \partial_t u \\ &= 2\frac{|\nabla u|^2}{u} - \frac{|\nabla u|^2}{u} + \Delta u + \partial_t u \\ &= \frac{|\nabla u|^2}{u} + Ru. \quad \Box \end{aligned}$$

**Proposition 6.1.2** *Let $u$ be a positive solution to (6.1.1). Then*

$$\begin{aligned} H^*(\frac{|\nabla u|^2}{u} + Ru) = \frac{2}{u}\left(u_{ij} - \frac{u_i u_j}{u}\right)^2 + 2\nabla R \nabla u + \frac{4}{u} Ric(\nabla u, \nabla u) \\ + 2|Ric|^2 u + 2\nabla R \nabla u + 2u\Delta R. \end{aligned}$$

*Here* $\left(u_{ij} - \frac{u_i u_j}{u}\right)^2 \equiv |Hess\ u - \frac{du \otimes du}{u}|^2$*;* $\nabla R \nabla u = g(\nabla R, \nabla u)$.

PROOF. The proof is very similar to that in [Ha5] where Hamilton considered the evolution of $\frac{|\nabla u|^2}{u}$ where $u$ is a positive solution of the linear heat equation.

Write $v = \sqrt{u}$. Then

$$\Delta(\frac{|\nabla u|^2}{u}) = 4\Delta(|\nabla v|^2) = 4(v_i^2)_{kk} = 8(v_{ik} v_i)_k = 8v_{ikk} v_i + 8v_{ik}^2.$$

Here and later we are using local orthonormal coordinates when necessary.

$$\partial_t(\frac{|\nabla u|^2}{u}) = 4\partial_t(g^{ij}v_iv_j) = 8v_iv_{ti} + 8Ric(\nabla v, \nabla v).$$

Since $u$ is a solution of the conjugate heat equation, it is easy to check that

$$v_t = -\Delta v - \frac{|\nabla v|^2}{v} + Rv/2.$$

Hence

$$\partial_t(\frac{|\nabla u|^2}{u}) = 8v_i(-v_{kk} - v_k^2v^{-1})_i + 8v_i(Rv/2)_i + 8Ric(\nabla v, \nabla v).$$

Combining the above equalities and applying Bochner-Weitzenbock formula, we deduce

$$\begin{aligned} &H^*(\frac{|\nabla u|^2}{u}) \\ &= 8v_i(v_{ikk} - v_{kki}) + 8\left(v_{ik}^2 - \frac{2v_iv_kv_{ki}}{v} + \frac{v_k^2v_i^2}{v^2}\right) + 4v_iR_iv + 8Ric(\nabla v, \nabla v) \\ &= 16Ric(\nabla v, \nabla v) + 8(v_{ik} - \frac{v_kv_i}{v})^2 + 4v_iR_iv. \end{aligned}$$

Also observe that

$$H^*(Ru) = 2|Ric|^2u + 2\nabla R\nabla u + 2u\Delta R.$$

Therefore we have, after bring back $u = v^2$,

$$\begin{aligned} H^*(\frac{|\nabla u|^2}{u} + Ru) = &\frac{2}{u}\left(u_{ij} - \frac{u_iu_j}{u}\right)^2 + 2\nabla R\nabla u + \frac{4}{u}Ric(\nabla u, \nabla u) \\ &+ 2|Ric|^2u + 2\nabla R\nabla u + 2u\Delta R. \end{aligned}$$

□

**Definition 6.1.1** *(F entropy and W entropy) Perelman's F entropy is the integration of $H^*(u\ln u)$, i.e.*

$$\mathbf{F} = \int_{\mathbf{M}} (\frac{|\nabla u|^2}{u} + Ru)d\mu(g(t)). \tag{6.1.3}$$

*His W entropy is a combination of the F entropy and the Bolzman entropy together with certain scaling factor. Let $\tau$ be such that $\frac{d\tau}{dt} = -1$, define*

$$\mathbf{W} = \tau\mathbf{F} - \mathbf{B} - \frac{n}{2}(\ln 4\pi\tau) - n. \tag{6.1.4}$$

*i.e.*

$$\mathbf{W} = \int_{\mathbf{M}} \left[\tau(\frac{|\nabla u|^2}{u} + Ru) - u\ln u - \frac{n}{2}(\ln 4\pi\tau)\ u - nu\right] d\mu(g(t)).$$

With the above preparation we can give a short proof of:

**Theorem 6.1.1** *( [P1]) Let $u$ be a positive solution of the conjugate heat equation then Perelman's $F$ and $W$ entropy are nondecreasing in time $t$. Moreover*

$$\frac{d}{dt}\mathbf{F} = 2\int_{\mathbf{M}} |Ric - Hess(\ln u)|^2 \ u \ d\mu(g(t)),$$

$$\frac{d}{dt}\mathbf{W} = 2\tau \int_{\mathbf{M}} \left| Ric - Hess(\ln u) - \frac{1}{2\tau}g \right|^2 \ u \ d\mu(g(t)).$$

Proof. Note that

$$\frac{d}{dt}\mathbf{F} = \int_{\mathbf{M}} (\partial_t - R)(\frac{|\nabla u|^2}{u} + Ru) d\mu(g(t)) = \int_{\mathbf{M}} H^*(\frac{|\nabla u|^2}{u} + Ru) d\mu(g(t)).$$

By Proposition 6.1.2, we have

$$\frac{d}{dt}\mathbf{F} = \int_{\mathbf{M}} \left[ \frac{2}{u}\left(u_{ij} - \frac{u_i u_j}{u}\right)^2 + 2\nabla R \nabla u + \frac{4}{u} Ric(\nabla u, \nabla u) \right.$$
$$\left. + 2|Ric|^2 u \right] d\mu(g(t))$$

where we have used the identity

$$\int_{\mathbf{M}} (2\nabla R \nabla u + 2u\Delta R) d\mu(g(t)) = 0.$$

With the help of the contracted second Bianchi identity in Proposition 3.2.3: $\nabla_i R = 2\nabla_j R_{ij}$ (in orthonormal system) and integration by parts, we realize that

$$\int_{\mathbf{M}} \nabla R \nabla u d\mu(g(t)) = 2\int_{\mathbf{M}} \nabla_j R_{ij} \nabla_i u d\mu(g(t))$$
$$= -2\int_{\mathbf{M}} < Ric, \ Hess u > d\mu(g(t)).$$

Therefore

$$\begin{aligned} \frac{d}{dt}\mathbf{F} &= \int_{\mathbf{M}} \left[ \frac{2}{u}\left(u_{ij} - \frac{u_i u_j}{u}\right)^2 + 4R_{ij}(\frac{u_i u_j}{u} - u_{ij}) + 2R_{ij}^2 u \right] d\mu(g(t)) \\ &= 2\int_{\mathbf{M}} \left[ \frac{1}{\sqrt{u}}\left(u_{ij} - \frac{u_i u_j}{u}\right) - R_{ij}\sqrt{u} \right]^2 d\mu(g(t)) \\ &= 2\int_{\mathbf{M}} |Ric - Hess(\ln u)|^2 \ u \ d\mu(g(t)) \end{aligned} \tag{6.1.5}$$

This is the desired formula for the F entropy.

For the W entropy, note that $\frac{d}{dt}\mathbf{B} = \mathbf{F}$ by Proposition 6.1.1. Hence

$$\frac{d}{dt}\mathbf{W} = -\mathbf{F} + \tau \frac{d}{dt}\mathbf{F} - \mathbf{F} + \frac{n}{2\tau}.$$

From (6.1.5)

$$\frac{d}{dt}\mathbf{W} = 2\tau \int_{\mathbf{M}} |Ric - Hess(\ln u)|^2 \ u \ d\mu(g(t)) \\ - 2\int_{\mathbf{M}} \left( \frac{|\nabla u|^2}{u} + Ru \right) d\mu(g(t)) + \int_{\mathbf{M}} \frac{n}{2\tau} u d\mu(g(t)).$$

In a local frame, it holds

$$\int_{\mathbf{M}} g^{ij} \left[ R_{ij} - (\ln u)_{ij} \right] u \ d\mu \\ = \int_{\mathbf{M}} \left[ R - \frac{\Delta u}{u} + \frac{|\nabla u|^2}{u^2} \right] \ u \ d\mu \\ = \int_{\mathbf{M}} (\frac{|\nabla u|^2}{u} + R \ u) d\mu.$$

Here $d\mu = d\mu(g(t))$. Substituting this into the formula for $\frac{d}{dt}\mathbf{W}$, we obtain

$$\frac{d}{dt}\mathbf{W} = 2\tau \int_{\mathbf{M}} \left| Ric - Hess(\ln u) - \frac{1}{2\tau} g \right|^2 \ u \ d\mu.$$

□

An immediate consequence of the monotonicity of $W$ entropy is Perelman's finite time $\kappa$ noncollapsing theorem, a result of fundamental importance.

**Theorem 6.1.2** *(local noncollapsing theorem) Let $(M, g(t))$, $t \in [0, T)$ be a smooth Ricci flow on a closed manifold $M$. If $T < \infty$, then for any $r > 0$, there exists $\kappa = \kappa(g(0), T, r) > 0$ such that the flow $(M, g(t))$ is $\kappa$ noncollapsed, in the sense of Definition 5.3.3, below the scale $r$ for all $t \in [0, T)$.*

We will not repeat Perelman's original proof of the theorem here. Instead, let us mention that the theorem is an immediate consequence of the Sobolev imbedding Theorem 6.2.1 below and Theorem 4.1.2.

As a comparison, we outline Perelman's ([P1]) original proof of the monotonicity of the F entropy, which has a variational flavor. A careful proof can be found in Section 1.5 of [CZ] e.g.

**Lemma 6.1.1** *Let $g = g(t)$ be a solution to the Ricci flow and $f$ be a solution to the equation*

$$\Delta f + R + \partial_t f = |\nabla f|^2.$$

*Then the $F$ entropy defined by $F(g, f) := \int \big(R + |\nabla f|^2\big) e^{-f}\, d\mu$ satisfies*

$$\frac{\partial F(g,f)}{\partial t} = 2\int |Ric + Hess(f)|^2 e^{-f}\, d\mu \geq 0. \tag{6.1.6}$$

*Here $d\mu = d\mu(g(t))$.*

PROOF. (sketch).

Let $\delta g_{ij} = v_{ij}$ and $\delta f = h$ be variations of $g_{ij}$ and $f$ respectively. First we show that the first variation of the $F$ entropy is given by

$$\delta F(v_{ij}, h) = \int_M \Big[-v_{ij}(R_{ij} + (Hess f)_{ij}) + (\frac{v}{2} - h)(2\Delta f - |\nabla f|^2 + R)\Big]\, e^{-f} d\mu. \tag{6.1.7}$$

Here $v \equiv g^{ij} v_{ij}$.

By direct computation, we have

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

$$\delta |\nabla f|^2 = -v^{ij} \nabla_i f \nabla_j f + 2 < \nabla f, \nabla h >,$$

and

$$\delta(e^{-f} d\mu) = (\frac{v}{2} - h) e^{-f} d\mu.$$

From these the variation formula (6.1.7) follows.

Now, we take $v_{ij} = -2R_{ij}$, $v = -2R$ and $h = \partial_t f$ in (6.1.7). Using integration by parts and contracted second Bianchi identity, we can then prove the lemma. □

The next pointwise inequality due to Perelman [P1] clearly implies the monotonicity of $W$ entropy. It is also useful when one attempts to localize. The quantity $P$ below seems mysterious at the first glance. However, it arises from the Euler-Lagrange equation associated with the $W$ entropy.

**Proposition 6.1.3** *Let*

$$u \equiv \frac{e^{-f}}{(4\pi\tau)^{\frac{n}{2}}}$$

*be a positive solution to the conjugate heat equation (6.1.1), where* $d\tau/dt = -1$. *Define*

$$P \equiv [\tau(2\triangle f - |\nabla f|^2 + R) + f - n]u$$
$$= \tau(-2\Delta u + \frac{|\nabla u|^2}{u} + Ru) - u\ln u - \frac{n}{2}(\ln 4\pi\tau)u - nu.$$

*Then*

$$H^*P = 2\tau|Ric + Hess(f) - \frac{g}{2\tau}|^2 u\,. \tag{6.1.8}$$

*Here* $H^*$ *is again the conjugate heat operator* $\Delta - R + \partial_t$.

See [To].

PROOF. Notice that

$$\begin{cases} f = -\ln u - \frac{n}{2}\ln(4\pi\tau) \\ \nabla f = -\frac{\nabla u}{u} \\ \triangle f = -\frac{\triangle u}{u} + |\nabla f|^2 \\ \frac{\partial f}{\partial t} = -\frac{u_t}{u} + \frac{n}{2\tau}. \end{cases} \tag{6.1.9}$$

We obtain the evolution equation for $f$,

$$\frac{\partial f}{\partial t} = -\triangle f + |\nabla f|^2 - R + \frac{n}{2\tau}\,. \tag{6.1.10}$$

Write $P = \frac{P}{u}u$, we find that

$$H^*P = H^*(\frac{P}{u}u) = \frac{P}{u}H^*u + u\left(\frac{\partial}{\partial t} + \triangle\right)(\frac{P}{u}) + 2\langle\nabla\frac{P}{u}, \nabla u\rangle\,.$$

Since $H^*u = 0$, and $\nabla f = -\frac{\nabla u}{u}$, we have

$$\frac{H^*P}{u} = \left(\frac{\partial}{\partial t} + \triangle\right)(\frac{P}{u}) - 2\langle\nabla\frac{P}{u}, \nabla f\rangle. \tag{6.1.11}$$

For the first term on the right-hand side,

$$-\left(\frac{\partial}{\partial t} + \triangle\right)(\frac{P}{u}) = -\left(\frac{\partial}{\partial t} + \triangle\right)[\tau(2\triangle f - |\nabla f|^2 + R) + f - n]$$
$$= (2\triangle f - |\nabla f|^2 + R)$$
$$- \tau\left(\frac{\partial}{\partial t} + \triangle\right)(2\triangle f - |\nabla f|^2 + R) - \left(\frac{\partial}{\partial t} + \triangle\right)f.$$

Using the evolution equation for $f$ in (6.1.10) on the final term, we reduce this to

$$-\left(\frac{\partial}{\partial t}+\triangle\right)(\frac{P}{u}) = 2\triangle f - 2|\nabla f|^2 + 2R - \frac{n}{2\tau} - \tau\left(\frac{\partial}{\partial t}+\triangle\right)(2\triangle f - |\nabla f|^2 + R)\,.$$

Recall from Proposition 5.1.1 that

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\triangle\right)&(2\triangle f - |\nabla f|^2 + R)\\ &= 4\langle\, Ric, Hess(f)\rangle + \triangle|\nabla f|^2 - 2Ric(\nabla f, \nabla f)\\ &\quad - 2\langle\, \nabla f, \nabla(-\triangle f + |\nabla f|^2 - R)\rangle + 2|Ric|^2\,.\end{aligned} \tag{6.1.12}$$

Also

$$2\langle\, \nabla\frac{P}{u}, \nabla f\rangle = 2\tau\langle\, \nabla(2\triangle f - |\nabla f|^2 + R), \nabla f\rangle + 2|\nabla f|^2\,.$$

Combining the above expressions altogether, we find that

$$\begin{aligned}-\frac{H^*P}{u} &= 2\triangle f + 2R - \frac{n}{2\tau} - \tau(4\langle\, Ric, Hess(f)\rangle + 2|Ric|^2)\\ &\quad + \tau[-\triangle|\nabla f|^2 + 2Ric(\nabla f, \nabla f) + 2\langle\, \nabla f, \nabla(\triangle f)\rangle]\,.\end{aligned}$$

The three terms in the square brackets simplified to $-2|Hess(f)|^2$, so

$$\begin{aligned}-\frac{H^*P}{u} &= 2\triangle f + 2R - \frac{n}{2\tau}\\ &\quad - \tau[4\langle Ric, Hess(f)\rangle + 2|Ric|^2 + 2|Hess(f)|^2]\\ &= 2\triangle f + 2R - \frac{n}{2\tau} - 2\tau\left(|Ric + Hess(f)|\right)^2\\ &= 2\langle Ric + Hess(f), g_{ij}\rangle - \frac{g_{ij}^2}{2\tau} - 2\tau\left(|Ric + Hess(f)|\right)^2\\ &= -2\tau|Ric + Hess(f) - \frac{g}{2\tau}|^2\,.\end{aligned}$$

□

**Remark 6.1.1** *One can also prove the proposition by using Propositions 6.1.1, 6.1.2 and by computing $H^*(-2\tau\Delta u)$ directly.*

**Exercise 6.1.1** *Give an alternative proof of Proposition 6.1.3 using the above remark.*

**Corollary 6.1.1** *Let $u = u(x,t) = G(x,t;y,T)$, $t < T$, be the fundamental solution of the conjugate heat equation and $f$ be given by $u = e^{-f}/(4\pi(T-t))^{n/2})$. Let $P = P(u)$ be as in the previous proposition. Then $P \le 0$. Moreover, for any smooth curve $c = c(t)$ on $M$, it holds*

$$-\frac{d}{dt}f(c(t),t) \le \frac{1}{2}(R(c(t),t) + |c'(t)|^2) - \frac{1}{2(T-t)}f(c(t),t).$$

PROOF. When $t \to T^-$, the fundamental solution $G$ is asymptotically the fundamental solution of the heat equation in $R^n$. For the latter the corresponding quantity $P$ is zero. By the previous proposition, $P(u)$ is a subsolution of the conjugate heat equation. Hence the maximum principle implies that $P \le 0$. A detailed proof can be found in [Ni2]. Note that the reduced distance was used in the proof. However one can also use the geodesic distance function. See [LX] e.g.

In terms of the function $f$, the inequality $P(u) \le 0$ can be written as

$$(T-t)(2\Delta f - |\nabla f|^2 + R) + f - n \le 0.$$

Since $u$ solves the conjugate heat equation, we know by (6.1.10) that $f$ solves

$$\partial_t f = -\Delta f + |\nabla f|^2 - R + \frac{n}{2(T-t)}.$$

These two expressions show

$$\partial_t f + \frac{1}{2}R - \frac{1}{2}|\nabla f|^2 - \frac{f}{2(T-t)} \ge 0.$$

On the other hand

$$-\frac{d}{dt}f(c(t),t) = -\partial_t f - < \nabla f, c'(t) > \le -\partial_t f + \frac{1}{2}|\nabla f|^2 + \frac{1}{2}|c'(t)|^2.$$

The desired inequality follows from adding the last two inequalities. □

As a comparison, here in this section, we will present a slight generalization of the $W$ entropy and outline a proof by Perelman's original method. The result first appeared in [Lj].

Define a family of generalized $W$ entropy for the Ricci flow by:

$$W(g,f,\tau) := \int_M \left(\frac{a^2}{2\pi}\tau(R + |\nabla f|^2) + f - n\right) u\, d\mu \qquad (6.1.13)$$

where $R$ is the scalar curvature, $\tau = T - t > 0$;

$$u = \frac{e^{-f}}{(4\pi\tau)^{\frac{n}{2}}}$$

such that $\int u\, d\mu = 1$, $d\mu = d\mu(g(t))$.

**Theorem 6.1.3** *Let $g(t)$ be a solution to the Ricci flow, that is, $\frac{\partial g}{\partial t} = -2Ric$ on a closed manifold $M$ for $t \in [0, T)$, and $u : M \times [0, T) \mapsto (0, \infty)$ with $u = \frac{e^{-f}}{(4\pi\tau)^{\frac{n}{2}}}$ be a positive solution to the conjugate heat equation (6.1.1). For $0 \le a^2 \le 2\pi$, the functional defined in (6.1.13) is increasing according to*

$$\frac{\partial}{\partial t} W(g, f, \tau) \ge \frac{a^2\tau}{\pi} \int_M |Ric + Hess(f) - \frac{g}{2\tau}|^2 u\, d\mu \ge 0. \qquad (6.1.14)$$

Proof. (of Theorem 6.1.3) Notice

$$\begin{aligned} W(g, f, \tau) &= \int_M \left( \frac{a^2}{2\pi}\tau(R + |\nabla f|^2) + f - n \right) u\, d\mu \\ &= \frac{a^2}{2\pi} \int_M [\tau(R + |\nabla f|^2) + f - n] u\, d\mu \\ &\quad + (1 - \frac{a^2}{2\pi}) \left( \int_M f\, u\, d\mu \right) - (1 - \frac{a^2}{2\pi}) n. \end{aligned}$$

Here $f$ is given in the statement of the theorem. We split the derivative of $W$ over time $t$ into two parts,

$$\begin{aligned} \frac{\partial}{\partial t} W(g, f, \tau) &= \frac{a^2}{2\pi} \frac{\partial}{\partial t} \left( \int_M [\tau(R + |\nabla f|^2) + f - n] u\, d\mu \right) \\ &\quad + (1 - \frac{a^2}{2\pi}) \frac{\partial}{\partial t} \left( \int_M f\, u\, d\mu \right) \\ &= \frac{a^2}{2\pi} \frac{\partial}{\partial t} \int_M P\, d\mu + (1 - \frac{a^2}{2\pi}) \frac{\partial}{\partial t} \left( \int_M f\, u\, d\mu \right), \end{aligned}$$

where the quantity $P$ is given in Proposition 6.1.3. We compute for each term,

$$\begin{aligned} \frac{a^2}{2\pi} \frac{\partial}{\partial t} \int_M P\, d\mu &= \frac{a^2}{2\pi} \int_M (P_t\, d\mu + P \frac{\partial d\mu}{\partial t}) = \frac{a^2}{2\pi} \int_M (P_t\, d\mu + P(-R)\, d\mu) \\ &= \frac{a^2}{2\pi} \int_M H^* P\, d\mu - \frac{a^2}{2\pi} \int_M \triangle P\, d\mu = \frac{a^2}{2\pi} \int_M H^* P\, d\mu, \end{aligned}$$

where the last equality comes from $\int_M \triangle P \, d\mu = 0$ for closed manifold $M$. By Proposition 6.1.3, we have

$$\frac{a^2}{2\pi}\frac{\partial}{\partial t}\int_M P\, d\mu = \frac{a^2\tau}{\pi}\int_M |Ric + Hess(f) - \frac{g}{2\tau}|^2 u\, d\mu \geq 0\,.$$

It suffices to prove the non-negativity of $\frac{\partial}{\partial t}\left(\int_M f\, u\, d\mu\right)$. By direct computation,

$$\begin{aligned}\frac{\partial}{\partial t}\left(\int_M f\, u\, d\mu\right) &= \int_M f_t\, u\, d\mu + f\, u_t\, d\mu + f\, u\, \frac{\partial(d\mu)}{\partial t} \\ &= \int_M (-\triangle f + |\nabla f|^2 - R + \frac{n}{2\tau}) u\, d\mu \\ &\quad + \int_M \big(f(-\triangle u + Ru) - Rfu\big)\, d\mu\,.\end{aligned}$$

Using integration by parts, we have

$$\begin{aligned}\frac{\partial}{\partial t}\left(\int_M f\, u\, d\mu\right) &= \int_M (-2\triangle f + |\nabla f|^2) u\, d\mu + \int_M (\frac{n}{2\tau} - R) u\, d\mu \\ &= \int_M (2\frac{\triangle u}{u} - |\nabla f|^2) u\, d\mu + \int_M (\frac{n}{2\tau} - R) u\, d\mu \\ &= \int_M -|\nabla f|^2\, u\, d\mu + \int_M (\frac{n}{2\tau} - R) u\, d\mu \\ &= \frac{n}{2\tau} - \int_M (|\nabla f|^2 + R)\, u\, d\mu\,.\end{aligned} \tag{6.1.15}$$

Now we turn to estimate of $F(g,\tau) = \int_M (|\nabla f|^2 + R) u\, d\mu$. From Lemma 6.1.1, we have

$$\begin{aligned}\frac{\partial F}{\partial t} &= 2\int |Ric + Hess(f)|^2 u\, d\mu = 2\int \big(\sum_{i,j} |R_{ij} + f_{ij}|^2\big) u\, d\mu \\ &\geq 2\int \big(\sum_{i=j} |R_{ij} + f_{ij}|^2\big) u\, d\mu \geq 2\int \frac{1}{n}\big(\sum R_{ii} + \sum f_{ii}\big)^2 u\, d\mu \\ &= \frac{2}{n}\int (R + \triangle f)^2\, u\, d\mu\,.\end{aligned}$$

The last inequality comes from $\sqrt{\frac{a_1^2+\cdots+a_n^2}{n}} \geq \frac{a_1+\cdots+a_n}{n}$ for $a_i \geq 0$. Also by Cauchy-Schwarz inequality, we have

$$\int (R + \triangle f)\sqrt{u}\,\sqrt{u}\, d\mu \leq \left(\int (R+\triangle f)^2 u\, d\mu\right)^{\frac{1}{2}} \left(\int u\, d\mu\right)^{\frac{1}{2}}\,.$$

Since $\int u\,d\mu = 1$, the above inequality can be simplified as

$$\left(\int (R+\triangle f)u\,d\mu\right)^2 \le \int (R+\triangle f)^2 u\,d\mu\,.$$

Then the evolution of $F$ along the time $t$ would be estimated by

$$\frac{\partial F}{\partial t} \ge \frac{2}{n}\left(\int (R+\triangle f)u\,d\mu\right)^2 = \frac{2}{n}\left(\int (R+|\nabla f|^2)u\,d\mu\right)^2$$

due to the following equality in closed manifold $M$

$$\begin{aligned}\int_M (\triangle f - |\nabla f|^2)u\,d\mu &= \int_M (-\frac{\triangle u}{u} + |\nabla f|^2 - |\nabla f|^2)u\,d\mu \\ &= -\int_M \triangle u\,d\mu = 0\end{aligned}$$

$$\Rightarrow \int_M (\triangle f)\,u\,d\mu = \int_M |\nabla f|^2 u\,d\mu\,.$$

From the definition $F = \int (R+|\nabla f|^2)u\,d\mu$, we get

$$\frac{\partial F}{\partial t} \ge \frac{2}{n}F^2 \ge 0\,.$$

We claim

$$F(t) \le \frac{n}{2(T-t)}\,.$$

Here is the proof of the above claim,

$$\begin{aligned}&\frac{dF}{dt} \ge \frac{2}{n}F^2 \;\Rightarrow\; \frac{dF}{F^2} \ge \frac{2}{n}dt \;\Rightarrow\; \int_t^T \frac{dF}{F^2} \ge \frac{2}{n}(T-t) \\ \Rightarrow\;& -(\frac{1}{F(T)} - \frac{1}{F(t)}) \ge \frac{2}{n}(T-t) \;\Rightarrow\; \frac{1}{F(t)} \ge \frac{2}{n}(T-t) + \frac{1}{F(T)}\,.\end{aligned}$$

If $F(T) > 0$, then $\frac{1}{F(t)} \ge \frac{2}{n}(T-t)$, that is, $F(t) \le \frac{n}{2(T-t)}$;
If $F(T) \le 0$, since $\frac{dF}{dt} \ge 0$, then $F(t) \le 0 \le \frac{n}{2(T-t)}$ for all $t \in [0,T)$, therefore,

$$F(t) = \int (R+|\nabla f|^2)u\,d\mu \le \frac{n}{2(T-t)} = \frac{n}{2\tau}$$

plugging into (6.1.15), we obtain

$$\frac{\partial}{\partial t}\left(\int_M f\,u\,d\mu\right) = \frac{n}{2\tau} - \int_M (|\nabla f|^2 + R)\,u\,d\mu \ge 0\,.$$

Thus we complete the proof of Theorem 6.1.3.

□

At the end of this section we briefly touch on the concepts of reduced distance and volumes, introduced by Perelman [P1]. The reduced distance is a space-time distance function weighted by the scalar curvature. The reduced volume, derived from the reduced distance, is another monotone quantity under Ricci flow. These two quantities are crucial to Perelman's proof of the Poincaré conjecture. Though not used in this book, they are presented here for completeness.

Let $M$ be a compact Riemann manifold or a complete one with bounded curvature. Here it is convenient to write the Ricci flow in the form of $\partial_\tau g = 2Ric$, where $\tau$ is a backward time, i.e. $\tau = T - t$ for some fixed $T$. Given a smooth curve $c = c(\tau)$, $\tau \in [\tau_1, \tau_2]$ on $M$, parameterized by $\tau$, the $L$ length is defined by

$$L(c) = \int_{\tau_1}^{\tau_2} \sqrt{\tau}\left[R(c(\tau),\tau) + g(\tau)(c'(\tau), c'(\tau))\right] d\tau. \tag{6.1.16}$$

**Definition 6.1.2** *(L distance and Reduced distance ) Let $(p, \tau_1)$ and $(q, \tau_2)$ be two space time points in $M \times [a, b]$ where a smooth Ricci flow is defined.*

*The L distance, denoted by $L(p, \tau_1, q, \tau_2)$, is the infimum of the L lengths of the smooth curves $c = c(\tau)$ such that $c(\tau_1) = p$ and $c(\tau_2) = q$.*

*The reduced distance, denoted by $l(p, \tau_1, q, \tau_2)$, is*

$$l(p, \tau_1, q, \tau_2) = \frac{L(p, \tau_1, q, \tau_2)}{2\sqrt{|\tau_2 - \tau_1|}}.$$

**Definition 6.1.3** *(Reduced volume) Fixing a point $(p_0, \tau_0)$ in space time and $\tau > \tau_0$, the associated reduced volume is*

$$\tilde{V}(p_0, \tau_0, \tau) \equiv \int_M (4\pi(\tau - \tau_0))^{-n/2} \exp(-l(p_0, \tau_0, p, \tau)) d\mu(g(\tau)).$$

Perelman's reduced distance is related to the weighted distance introduced in [LY] for Schrödinger heat equation in the fixed metric case. Perelman discovered a number of amazing and sharp differential inequalities or even equalities for reduced distance and volume, which are not expected for such complex quantities. For instance, Perelman showed that the reduced volume is nondecreasing in $\tau$, with equality holding only on gradient shrinking solitons. We refer the reader to [P1] and [CZ], [KL] and [MT] for detailed information.

## 6.2 Log Sobolev inequality and Sobolev inequality under Ricci flow

In this section, we strengthen the monotonicity of Perelman's W entropy to a uniform Sobolev inequality along Ricci flow. This result first appeared in the arXiv version of [Z2] (June 2007). There was an error in the Sobolev coefficients for large time, which was corrected in the erratum. Later it appeared in [Y] with the same error in the first version (July 2007) and in [Hs]. We know that Sobolev inequality contains a host of analytical and geometric information. These include noncollapsing and isoperimetric inequality e.g. It is an important tool in studying elliptic and parabolic differential equations on manifolds. In the papers [CH], etc., [Se2] and [Ru1] some applications of such Sobolev inequality to Kähler Ricci flow are already found.

**Theorem 6.2.1** *(Sobolev inequality for smooth Ricci flow) Let* $\mathbf{M}$ *be a compact Riemann manifold with dimension* $n \geq 3$ *and the metrics* $g = g(t)$ *evolve by the Ricci flow* $\partial_t g = -2Ric$*. Let* $A$ *and* $B$ *be positive numbers such that the* $L^2$ *Sobolev inequality for* $(\mathbf{M}, g(0))$ *holds, i.e. for any* $v \in W^{1,2}(\mathbf{M})$*,*

$$\left( \int_{\mathbf{M}} v^{2n/(n-2)} \, d\mu(g(0)) \right)^{(n-2)/n} \leq A \int_{\mathbf{M}} |\nabla v|^2 \, d\mu(g(0)) + B \int_{\mathbf{M}} v^2 \, d\mu(g(0)).$$

*Also let* $\lambda_0$ *be the first eigenvalue of Perelman's* $F$ *entropy, i.e.*

$$\lambda_0 = \inf_{\|v\|_2 = 1} \int_{\mathbf{M}} (4|\nabla v|^2 + Rv^2) d\mu(g(0))$$

*Then the following conclusions are true.*

*(a) Suppose the Ricci flow is smooth for* $t \in (0, T_0)$ *where* $T_0 \leq \infty$ *is the life span of the Ricci flow. Then there exist positive functions* $A(t), B(t)$ *depending only on the initial metric* $g(0)$ *in terms of* $A$ *and* $B$*, and* $t$ *such that, for all* $v \in W^{1,2}(\mathbf{M}, g(t))$*,* $t \in [0, T_0)$*, it holds*

$$\left( \int v^{2n/(n-2)} d\mu(g(t)) \right)^{(n-2)/n} \leq A(t) \int (|\nabla v|^2 + \frac{1}{4} Rv^2) d\mu(g(t)) + B(t) \int v^2 d\mu(g(t)).$$

*Here* $R$ *is the scalar curvature with respect to* $g(t)$*.*

*Moreover, if $\lambda_0 > 0$, which holds if $R(x,0) > 0, x \in \mathbf{M}$, then $A(t)$ is independent of $t$ and $B(t) = 0$, i.e. there exists a constant $A_0$ independent of time $t$ such that, for all $v \in W^{1,2}(\mathbf{M}, g(t))$, $t \in [0, T_0)$, it holds*

$$\left( \int v^{2n/(n-2)} d\mu(g(t)) \right)^{(n-2)/n} \leq A_0 \int (|\nabla v|^2 + \frac{1}{4} R v^2) d\mu(g(t)).$$

*(b) Suppose the Ricci flow is smooth for $t \in (0,1)$ and is singular at $t = 1$. Let $\tilde{t} = -\ln(1-t)$ and $\tilde{g}(\tilde{t}) = \frac{1}{1-t} g(t)$ which satisfy a normalized Ricci flow*

$$\partial_{\tilde{t}} \tilde{g} = -2\tilde{Ric} + \tilde{g}.$$

*Then there exist positive constants $\tilde{A}, \tilde{B}$ depending only on the initial metric $g(0)$ such that, for all $v \in W^{1,2}(\mathbf{M}, \tilde{g}(\tilde{t}))$, $\tilde{t} > 0$, it holds*

$$\left( \int v^{2n/(n-2)} d\mu(\tilde{g}(\tilde{t})) \right)^{(n-2)/n} \leq \tilde{A} \int (|\tilde{\nabla} v|^2 + \frac{1}{4} \tilde{R} v^2) d\mu(\tilde{g}(\tilde{t})) + \tilde{B} \int v^2 d\mu(\tilde{g}(\tilde{t})).$$

*Here $\tilde{R}$ is the scalar curvature with respect to $\tilde{g}(\tilde{t})$.*

**Remark 6.2.1** *In part (b), $\tilde{B} = 0$ if the initial metric satisfies*

$$\lambda_0 = \inf_{v \in W^{1,2}(\mathbf{M}), \|v\|_2 = 1} \int (4|\tilde{\nabla} v|^2 + \tilde{R} v^2) d\mu(g(0)) > 0.$$

*Proof of the theorem*

Since Case (b) is an immediate consequence of Case (a) by scaling, we just prove Case (a). The proof is divided into a few steps.

*Step 1.* We show that the monotone property of $W$, the Perelman $W$ entropy, implies the Log Sobolev inequalities (6.2.8) below.

Let us assume that the Ricci flow exists in the time interval $[0, t_0]$.

It is convenient to work on the scaled time and metric $\tilde{t} = t/t_0$ and $\tilde{g} = g/t_0$. Clearly $\tilde{g}(\tilde{t})$ still satisfies the Ricci flow equation.

For any $\epsilon > 0$, we take

$$\tau = \tau(\tilde{t}) = \epsilon^2 + 1 - \tilde{t}$$

so that $\tau_1 = 1 + \epsilon^2$ and $\tau_2 = \epsilon^2$ (by taking $\tilde{t}_1 = 0$ and $\tilde{t}_2 = 1$).

Recall that Perelman's $W$ entropy is

$$W(\tilde{g}, f, \tau) = \int_{\mathbf{M}} \left(\tau(\tilde{R} + |\tilde{\nabla} f|^2) + f - n\right) u \, d\mu(\tilde{g}(\tilde{t}))$$

where $u = \frac{e^{-f}}{(4\pi\tau)^{n/2}}$. Let $u_2$ be a minimizer of the entropy $W(\tilde{g}, f, \tau_2)$ for all $u$ such that $\int u d\mu(\tilde{g}(\tilde{t}_2)) = 1$. We solve the backward heat equation with the final value chosen as $u_2$ at $\tilde{t} = \tilde{t}_2$. Let $u_1$ be the value of the solution of the backward heat equation at $\tilde{t} = \tilde{t}_1$. As usual, we define functions $f_i$ with $i = 1, 2$ by the relation $u_i = e^{-f_i}/(4\pi\tau_i)^{n/2}$, $i = 1, 2$. Then, by the monotonicity of the $W$ entropy

$$\begin{aligned} inf_{\int u_0 d\mu(\tilde{g}(\tilde{t}_1))=1} W(\tilde{g}(\tilde{t}_1), f_0, \tau_1) &\leq W(\tilde{g}(\tilde{t}_1), f_1, \tau_1) \leq W(\tilde{g}(\tilde{t}_2), f_2, \tau_2) \\ &= inf_{\int u d\mu(\tilde{g}(\tilde{t}_2))=1} W(\tilde{g}(\tilde{t}_2), f, \tau_2), \end{aligned} \tag{6.2.1}$$

where $f_0$ and $f$ are given by the formulas

$$u_0 = e^{-f_0}/(4\pi\tau_1)^{n/2}, \qquad u = e^{-f}/(4\pi\tau_2)^{n/2}.$$

Using these notations we can rewrite (6.2.1) as

$$\begin{aligned} &\inf_{\|u\|_1=1} \int_{\mathbf{M}} \left(\epsilon^2(\tilde{R} + |\tilde{\nabla} \ln u|^2) - \ln u - \ln(4\pi\epsilon^2)^{n/2}\right) u \, d\mu(\tilde{g}(\tilde{t}_2)) \\ &\quad \geq \inf_{\|u_0\|_1=1} \int_{\mathbf{M}} \Big((1+\epsilon^2)(\tilde{R} + |\tilde{\nabla} \ln u_0|^2) - \ln u_0 \\ &\qquad - \ln(4\pi(1+\epsilon^2))^{n/2}\Big) u_0 \, d\mu(\tilde{g}(0)). \end{aligned}$$

Observe that the $\ln(4\pi)^{n/2}$ terms on both sides of the above inequality can be canceled. Denote $\tilde{v} = \sqrt{u}$ and $\tilde{v}_0 = \sqrt{u_0}$. We obtain,

$$\begin{aligned} &\inf_{\|\tilde{v}\|_2=1} \int_{\mathbf{M}} \left(\epsilon^2(\tilde{R}\tilde{v}^2 + 4|\tilde{\nabla}\tilde{v}|^2) - \tilde{v}^2 \ln \tilde{v}^2\right) \, d\mu(\tilde{g}(\tilde{t}_2)) - n \ln \epsilon \\ &\quad \geq \inf_{\|\tilde{v}_0\|_2=1} \int_{\mathbf{M}} \left((1+\epsilon^2)(\tilde{R}\tilde{v}_0^2 + 4|\tilde{\nabla}\tilde{v}_0|^2) - \tilde{v}_0^2 \ln \tilde{v}_0^2\right) \, d\mu(\tilde{g}(0)) \\ &\qquad - \ln(1+\epsilon^2)^{n/2}. \end{aligned} \tag{6.2.2}$$

Now we want to go back to the original metric $g$ and time $t$. Using the conversion formulas $\tilde{R} = t_0 R$, $d\mu(\tilde{g}) = d\mu(g)/t_0^{n/2}$, $v = \tilde{v}/t_0^{n/4}$ and

$v_0 = \tilde{v}_0/t_0^{n/4}$, we obtain

$$\begin{aligned}&\inf_{\|v\|_2=1}\int_{\mathbf{M}}\left(t_0\epsilon^2(Rv^2+4|\nabla v|^2)-v^2\ln v^2\right)\,d\mu(g(t_0))-\ln(t_0\epsilon^2)^{n/2}\\&\qquad\geq\inf_{\|v_0\|_2=1}\int_{\mathbf{M}}\left(t_0(1+\epsilon^2)(Rv_0^2+4|\nabla v_0|^2)-v_0^2\ln v_0^2\right)\,d\mu(g(0))\\&\qquad\quad-\ln[t_0(1+\epsilon^2)]^{n/2}.\end{aligned}$$

Because $\epsilon$ is arbitrary, we can rename $\sqrt{t_0}\epsilon$ as $\epsilon$. Then the above inequality becomes

$$\begin{aligned}&\inf_{\|v\|_2=1}\int_{\mathbf{M}}\left(\epsilon^2(Rv^2+4|\nabla v|^2)-v^2\ln v^2\right)\,d\mu(g(t_0))-\frac{n}{2}\ln\epsilon^2\\&\qquad\geq\inf_{\|v_0\|_2=1}\int_{\mathbf{M}}\left((t_0+\epsilon^2)(Rv_0^2+4|\nabla v_0|^2)-v_0^2\ln v_0^2\right)\,d\mu(g(0))\\&\qquad\quad-\frac{n}{2}\ln(t_0+\epsilon^2).\end{aligned}\tag{6.2.3}$$

Since $(\mathbf{M}, g(0))$ is a compact Riemann manifold, the Sobolev inequality as described in Section 4.1 holds, i.e. for any $v_0 \in W^{1,2}(\mathbf{M})$, there exist positive constants $A$ and $B$ depending only on the metric $g(0)$ such that

$$\begin{aligned}\left(\int_{\mathbf{M}}v_0^{2n/(n-2)}\,d\mu(g(0))\right)^{(n-2)/n}&\leq A\int_{\mathbf{M}}|\nabla v_0|^2\,d\mu(g(0))\\&\quad+B\int_{\mathbf{M}}v_0^2\,d\mu(g(0)).\end{aligned}$$

By the work of [Heb1], we know that $A$ can be any number strictly larger than the Euclidean Sobolev constant; and $B$ depends on $A$, the injectivity radius and the lower bound of the Ricci curvature of $(\mathbf{M}, g(0))$ only.

Recall

$$\lambda_0=\inf_{\|v_0\|_2=1}\int_{\mathbf{M}}(4|\nabla v_0|^2+Rv_0^2)d\mu(g(0))$$

which is the infimum of Perelman's $F$ entropy for $(\mathbf{M}, g(0))$. Then we can convert the above Sobolev inequality to

$$\begin{aligned}\left(\int_{\mathbf{M}}v_0^{2n/(n-2)}\,d\mu(g(0))\right)^{(n-2)/n}&\leq A_0\int_{\mathbf{M}}(4|\nabla v_0|^2+Rv_0^2)\,d\mu(g(0))\\&\quad+B_0\int_{\mathbf{M}}v_0^2\,d\mu(g(0)).\end{aligned}\tag{6.2.4}$$

Here the constants $A_0$ and $B_0$ are given below.

$$\begin{aligned}&A_0 = A4^{-1} + B\lambda_0^{-1} + A4^{-1}\sup R^-(\cdot,0)\lambda_0^{-1}, \quad B_0 = 0 \quad \text{if} \quad \lambda_0 > 0;\\ &A_0 = A4^{-1}, \quad B_0 = A4^{-1}\sup R^-(\cdot,0) + B, \quad \text{if} \quad \lambda_0 \le 0.\end{aligned} \tag{6.2.5}$$

The following argument is similar to Proposition 4.2.1 (I) to (II). Applying Hölder and Jensen inequalities, we have, for all $v_0 \in W^{1,2}(\mathbf{M})$ verifying $\|v_0\|_2 = 1$,

$$\int_{\mathbf{M}} v_0^2 \ln v_0^2 \, d\mu(g(0)) \le \frac{1}{2} n \ln\left(A_0 \int_{\mathbf{M}} (4|\nabla v_0|^2 + Rv_0^2)\, d\mu(g(0)) + B_0\right).$$

By the elementary inequality $\ln z \le qz - \ln q - 1, \quad q, z > 0$, we know that

$$\begin{aligned}\int_{\mathbf{M}} v_0^2 \ln v_0^2 \, d\mu(g(0)) \le \frac{n}{2} q\left(A_0 \int_{\mathbf{M}} (4|\nabla v_0|^2 + Rv_0^2)\, d\mu(g(0)) + B_0\right)\\ - \frac{n}{2}\ln q - \frac{n}{2}.\end{aligned}$$

In the above, we choose $q$ so that $\frac{n}{2} q A_0 = t_0 + \epsilon^2$, i.e. $q = 2(t_0 + \epsilon^2)/(nA_0)$. Then

$$\begin{aligned}\int_{\mathbf{M}} v_0^2 \ln v_0^2 \, d\mu(g(0)) &\le (t_0+\epsilon^2) \int_{\mathbf{M}} (4|\nabla v_0|^2 + Rv_0^2)\, d\mu(g(0))\\ &+ \frac{(t_0+\epsilon^2)B_0}{A_0} - \frac{n}{2}\ln\frac{2(t_0+\epsilon^2)}{nA_0} - \frac{n}{2}.\end{aligned}$$

After rearrangement, this inequality becomes

$$\begin{aligned}&(t_0+\epsilon^2)\int_{\mathbf{M}} (4|\nabla v_0|^2 + Rv_0^2)\, d\mu(g(0))\\ &\quad - \int_{\mathbf{M}} v_0^2 \ln v_0^2 \, d\mu(g(0)) - \frac{n}{2}\ln(t_0+\epsilon^2)\\ &\quad \ge -(t_0+\epsilon^2)B_0A_0^{-1} - n2^{-1}\ln(nA_0 2^{-1}) + n2^{-1}.\end{aligned} \tag{6.2.6}$$

Substituting the log Sobolev inequality (6.2.6) to the right-hand side of (6.2.3), we deduce

$$\begin{aligned}\inf_{\|v\|_2=1} \int_{\mathbf{M}} \left(\epsilon^2(Rv^2 + 4|\nabla v|^2) - v^2\ln v^2\right)\, d\mu(g(t_0)) - n\ln\epsilon\\ \ge -(t_0+\epsilon^2)B_0A_0^{-1} - n2^{-1}\ln(nA_0 2^{-1}) + n2^{-1}.\end{aligned} \tag{6.2.7}$$

Therefore, we reach the uniform log Sobolev inequality:

$$\int_{\mathbf{M}} v^2 \ln v^2 \, d\mu(g(t_0)) \le \epsilon^2 \int_{\mathbf{M}} \big(4|\nabla v|^2 + Rv^2\big) d\mu(g(t_0)) - n \ln \epsilon$$
$$+ (t_0 + \epsilon^2) B_0 A_0^{-1} + n2^{-1} \ln(nA_0 2^{-1}) - n2^{-1}. \tag{6.2.8}$$

*Step 2.* Fix a time $t_0$ during the Ricci flow or $\tilde{t}_0$ during the normalized one. Suppose on $(\mathbf{M}, g(t_0))$ or $(\mathbf{M}, \tilde{g}(\tilde{t}_0))$, the Log Sobolev inequalities (6.2.8) holds. We show that they imply upper bound for the heat kernel (fundamental solution) of

$$\Delta u(x,t) - \frac{1}{4} R(x, t_0) u(x,t) - \partial_t u(x,t) = 0 \tag{6.2.9}$$

under the fixed metric $g(t_0)$ or $\tilde{g}(\tilde{t}_0)$. We stress that the time $t$ here is no longer the time in the Ricci flow which is frozen at time $t_0$.

The proof of the upper bound, which does not distinguish between the Ricci flow or the normalized Ricci flow case, follows the original ideas of Davies [Da]. There is only one extra issue to deal with here. Namely the negative part of the scalar curvature may make the semigroup generated by $\Delta - \frac{1}{4}R$ not contractive. However the modification in the proof is moderate since the most negative value of the scalar curvature does not decrease under either the Ricci flow or the normalized one in the theorem. This statement is a result of the maximum principle applied to the equation

$$\Delta R - \partial_t R + 2|Ric|^2 = 0.$$

For this reason, we will be brief in the presentation.

Let $u$ be a positive solution to (6.2.9). Given $T > 0$ and $t \in (0, T)$, we take

$$p(t) = T/(T-t)$$

so that $p(0) = 1$ and $p(T) = \infty$. By direct computation

$$\begin{aligned}
\partial_t \|u\|_{p(t)} &= \partial_t \Big( \int_{\mathbf{M}} u^{p(t)}(x,t) dx \Big)^{1/p(t)} \\
&= -\frac{p'(t)}{p^2(t)} \|u\|_{p(t)} \ln \int_{\mathbf{M}} u^{p(t)}(x,t) dx \\
&\quad + \frac{1}{p(t)} \Big( \int_{\mathbf{M}} u^{p(t)}(x,t) dx \Big)^{(1/p(t))-1} \\
&\quad \times \Big[ \int_{\mathbf{M}} u^{p(t)} (\ln u) p'(t) dx + p(t) \int_{\mathbf{M}} u^{p(t)-1} (\Delta u - \frac{1}{4} Ru) dx \Big].
\end{aligned}$$

Here $dx$ means the integral element with respect to $g(t_0)$. We adopt this notation to emphasize that $g(t_0)$ is not evolving with respect to $t$. Using integration by parts on the term containing $\Delta u$ and multiplying both sides by $p^2(t)\|u\|_{p(t)}^{p(t)}$, we reach

$$\begin{aligned} p^2(t)\|u\|_{p(t)}^{p(t)}\partial_t\|u\|_{p(t)} = &-p'(t)\|u\|_{p(t)}^{p(t)+1}\ln\int_{\mathbf{M}} u^{p(t)}(x,t)dx \\ &+ p(t)\|u\|_{p(t)}p'(t)\int_{\mathbf{M}} u^{p(t)}\ln u(x,t)dx \\ &- p^2(t)(p(t)-1)\|u\|_{p(t)}\int_{\mathbf{M}} u^{p(t)-2}|\nabla u|^2(x,t)dx \\ &- p^2(t)\|u\|_{p(t)}\int_{\mathbf{M}}\frac{1}{4}R(x,t_0)u^{p(t)}(x,t)dx. \end{aligned}$$

Dividing both sides by $\|u\|_{p(t)}$, we obtain

$$\begin{aligned} p^2(t)\|u\|_{p(t)}^{p(t)}\partial_t\ln\|u\|_{p(t)} = &-p'(t)\|u\|_{p(t)}^{p(t)}\ln\int_{\mathbf{M}} u^{p(t)}(x,t)dx \\ &+ p(t)p'(t)\int_{\mathbf{M}} u^{p(t)}\ln u(x,t)dx \\ &- 4(p(t)-1)\int_{\mathbf{M}}|\nabla(u^{p(t)/2})|^2(x,t)dx - p^2(t) \\ &\times\int_{\mathbf{M}}\frac{1}{4}R(x,t_0)(u^{p(t)/2})^2(x,t)dx. \end{aligned}$$

Merging the first two terms on the right-hand side of the above equality and making the substitution $v = u^{p(t)/2}/\|u^{p(t)/2}\|_2$, we arrive at, after dividing by $\|u\|_{p(t)}^{p(t)}$,

$$\begin{aligned} &p^2(t)\partial_t\ln\|u\|_{p(t)} \\ &= p'(t)\int_{\mathbf{M}} v^2\ln v^2(x,t)dx - 4(p(t)-1)\int_{\mathbf{M}}|\nabla v|^2(x,t)dx - p^2(t) \\ &\quad\times\int_{\mathbf{M}}\frac{1}{4}R(x,t_0)v^2(x,t)dx \\ &= p'(t)\int_{\mathbf{M}} v^2\ln v^2(x,t)dx - 4(p(t)-1)\int_{\mathbf{M}}(|\nabla v|^2(x,t) \\ &\qquad + \frac{1}{4}R(x,t_0)v^2)dx \\ &\qquad + (4(p(t)-1)-p^2(t))\int_{\mathbf{M}}\frac{1}{4}R(x,t_0)v^2(x,t)dx. \end{aligned}$$

It is easy to check $\|v\|_2 = 1$ and also

$$\frac{4(p(t)-1)}{p'(t)} = \frac{4t(T-t)}{T} \le T,$$

$$-T \le \frac{4(p(t)-1)-p^2(t)}{p'(t)} = \frac{4t(T-t)-T^2}{T} \le 0.$$

Hence

$$\begin{aligned} p^2(t)\partial_t \ln \|u\|_{p(t)} \le p'(t)\bigg( &\int_{\mathbf{M}} v^2 \ln v^2(x,t)dx \\ &- \frac{4(p(t)-1)}{p'(t)} \int_{\mathbf{M}} (|\nabla v|^2(x,t) + \frac{1}{4}R(x,t_0)v^2)dx + T \sup R^-(x,t_0)\bigg). \end{aligned}$$

Taking $\epsilon$ so that

$$\frac{\epsilon^2}{\pi} = \frac{4(p(t)-1)}{p'(t)} \le T$$

in the log Sobolev inequality (6.2.8), we deduce

$$p^2(t)\partial_t \ln \|u\|_{p(t)} \le p'(t)\Big(-n \ln \sqrt{\pi 4(p(t)-1)/p'(t)} + L + T \sup R^-(x,0)\Big)$$

where

$$\begin{aligned} L &\equiv (t_0+\epsilon^2)B_0 A_0^{-1} + n2^{-1}\ln(nA_0 2^{-1}) - n2^{-1} \\ &\le (t_0+\pi T)B_0 A_0^{-1} + n2^{-1}\ln(nA_0 2^{-1}) - n2^{-1}. \end{aligned} \tag{6.2.10}$$

Here we also used the fact that $\sup R^-(x,t_0) \le \sup R^-(x,0)$ as remarked earlier.

Note that

$$p'(t)/p^2(t) = 1/T, \qquad 4(p(t)-1)/p'(t) = 4t(T-t)/T.$$

Hence

$$\partial_t \ln \|u\|_{p(t)} \le \frac{1}{T}\Big( -\frac{n}{2}\ln \pi 4t(T-t)/T + L + T\sup R^-(x,0)\Big).$$

This yields, after integration from $t=0$ to $t=T$,

$$\ln \frac{\|u(\cdot,T)\|_\infty}{\|u(\cdot,0)\|_1} \le -\frac{n}{2}\ln(4\pi T) + L + T\sup R^-(x,0). \tag{6.2.11}$$

Denote by $p = p(x,t,y)$ the heat kernel of (6.2.9). Since

$$u(x,T) = \int_{\mathbf{M}} p(x,T,y)u(y,0)dy,$$

(6.2.11) shows

$$p(x,T,y) \le \frac{\exp(L + T \sup R^-(x,0))}{(4\pi T)^{n/2}}. \tag{6.2.12}$$

Recall that

$$L \le (t_0 + \pi T)B_0 A_0^{-1} + n2^{-1}\ln(nA_0 2^{-1}) - n2^{-1}. \tag{6.2.13}$$

If $\lambda_0 > 0$, then $B_0 = 0$ (cf. (6.2.5)). So the above bound becomes

$$\begin{aligned} p(x,T,y) &\le \frac{\exp(n2^{-1}\ln(nA_0 2^{-1}) - n2^{-1})}{(4\pi T)^{n/2}} e^{T \sup R^-(x,0)} \\ &\le \frac{c_1(A + B\lambda_0^{-1} + 1)^{c_2}}{T^{n/2}} e^{T \sup R^-(x,0)} \end{aligned} \tag{6.2.14}$$

where $c_1$ and $c_2$ are numerical constants. The last inequality is due to (6.2.5) where $A_0$ is given.

Using integration by parts and the monotonicity of Perelman's $F$ entropy (Theorem 6.1.1), it is easy to see that

$$\begin{aligned} \frac{d}{ds}\int_{\mathbf{M}} p^2(y,s,x)dy &= 2\int_{\mathbf{M}} p(\Delta p - Rp/4)dy \\ &= -\frac{1}{2}\int_{\mathbf{M}} (4|\nabla p|^2 + Rp^2)dy \\ &= -\frac{1}{2}\frac{\int_{\mathbf{M}}(4|\nabla p|^2 + Rp^2)dy}{\|p\|_2^2}\int_{\mathbf{M}} p^2(y,s,x)dy \\ &\le -\frac{1}{2}\int_{\mathbf{M}} p^2(y,s,x)dy \quad \inf\{F(v) \; | v \in W^{1,2}(\mathbf{M}, g(t_0)), \; \|v\|_2 = 1\} \\ &\le -\frac{\lambda_0}{2}\int_{\mathbf{M}} p^2(y,s,x)dy. \end{aligned}$$

Here $F(v)$ is the $F$ entropy. Therefore, for $s > 1$,

$$\begin{aligned} p(x,2s,x) = \int_{\mathbf{M}} p^2(y,s,x)dy &\le e^{-\lambda_0(s-1)/2}\int_{\mathbf{M}} p^2(y,1,x)dy \\ &= e^{-\lambda_0(s-1)/2} p(x,2,x). \end{aligned}$$

By the reproducing formula for the heat kernel $p$, it is easy to see that

$$p(x, 2s, y) \le \sqrt{p(x, 2s, x)}\sqrt{p(y, 2s, y)}.$$

The last two inequalities can be combined to give

$$p(x, 2s, y) \le e^{-\lambda_0(s-1)/2}\sqrt{p(x, 2, x)}\sqrt{p(y, 2, y)}.$$

By (6.2.14), we see that, for $s \ge 1$,

$$p(x, 2s, y) \le \frac{c_1(A + B\lambda_0^{-1} + 1)^{c_2}}{2^{n/2}} e^{2\sup R^-(x,0)}\ e^{-\lambda_0(s-1)/2}.$$

Thus we arrive at the following uniform bound that depends only on the initial metric $g(0)$ but not on the underlining metric $g(t_0)$.

If $\lambda_0 > 0$, then, for all $T > 0$,

$$p(x, T, y) \le \frac{c_1(A + B\lambda_0^{-1} + 1)^{c_2}}{T^{n/2}} e^{2\sup R^-(x,0)} e^{-\lambda_0 T/5}. \tag{6.2.15}$$

If $\lambda_0 \le 0$, from (6.2.12) and (6.2.13) again, we have

$$p(x, T, y) \le \frac{\exp(c_1[(BA^{-1} + \sup R^-(\cdot, 0))\ (t_0 + T) + A^{c_2} + 1)]}{T^{n/2}}. \tag{6.2.16}$$

*Step 3.* We show that the above heat kernel upper bound implies the Sobolev imbedding in Theorem 6.2.1.

This is more or less standard.

Case 1. Suppose $\lambda_0 < 0$.

Let $t_0$ be a fixed time during Ricci flow. Let $F = \sup R^-(x, 0)$ and $p_F$ be the heat kernel of the operator $\Delta - \frac{1}{4}R(x, t_0) - F - 1$. Since $R^-(x, t_0) \le F$, from the upper bound for $p$ in (6.2.16), we know that $p_F$ obeys the global upper bound

$$p_F(x, t, y) \le \frac{\Lambda}{t^{n/2}}, \qquad t > 0.$$

Here $\Lambda$ depends only on $L$ in (6.2.10) and $F$. Moreover $p_F$ is a contraction. By Hölder inequality, for any $f \in L^2(\mathbf{M})$, we have

$$|\int_{\mathbf{M}} p_F(x, t, y) f(y) dy| \le \left(\int_{\mathbf{M}} p_F^2(x, t, y) dy\right)^{1/2} \|f\|_2 \le \Lambda^{1/2} t^{-n/4} \|f\|_2.$$

The Sobolev inequality in Theorem 6.2.1 now follows from Theorem 2.4.2 in [Da] (see Theorem 4.2.1 here), i.e. there exist positive constants

$A(t_0), B(t_0)$ depending only on the initial metric through $\Lambda$ and $t_0$ such that, for all $v \in W^{1,2}(\mathbf{M}, g(t_0))$, it holds

$$\begin{aligned}\left(\int v^{2n/(n-2)} d\mu(g(t_0))\right)^{(n-2)/n} &\le A(t_0) \int (|\nabla v|^2 + \frac{1}{4} R v^2) d\mu(g(t_0)) \\ &\quad + B(t_0) \int v^2 d\mu(g(t_0)).\end{aligned}$$

The same also holds for the normalized Ricci flow. Since $t_0 \in [0, T_0)$ is arbitrary, the proof is done.

Case 2. Suppose $\lambda_0 > 0$.

The bound in (6.2.15) is independent of $t_0$. Consider $p_F$ again. The statement in Theorem 4.2.1 about the relation between the constants in the heat kernel bounds and the Sobolev inequality shows:

for $v \in C^\infty(\mathbf{M})$, and a positive constant $A_0$ independent of $t_0$, and a constant $C$,

$$\begin{aligned}\left(\int v^{2n/(n-2)} d\mu(g(t_0))\right)^{(n-2)/n} &\le A_0 \int (|\nabla v|^2 + \frac{1}{4} R v^2) d\mu(g(t_0)) \\ &\quad + C(\sup R^-(x,0) + 1) \int v^2 d\mu(g(t_0)).\end{aligned}$$

By the monotone increasing property of the $F$ entropy mentioned above, we have

$$\int v^2 d\mu(g(t_0)) \le \lambda_0^{-1} \int (4|\nabla v|^2 + R v^2) d\mu(g(t_0)).$$

Consequently, there exists a constant, still denoted by $A_0$ such that

$$\left(\int v^{2n/(n-2)} d\mu(g(t_0))\right)^{(n-2)/n} \le A_0 \int (|\nabla v|^2 + \frac{1}{4} R v^2) d\mu(g(t_0)),$$

i.e. the last statement in part (a) of the theorem is true.

One can also prove the Sobolev imbedding by establishing a Nash type inequality first and using an argument in [BCLS]. □

## 6.3 Critical and local Sobolev inequality

In this section, we state and prove a critical and a local Sobolev inequality under Ricci flow. They are not used in the proof of the Poincaré conjecture.

First we show a uniform critical Sobolev inequality under Ricci flow. This can be regarded as a generalization of the imbedding $W_0^{1,n}(\Omega)$ into certain Orlicz space, which was described in Theorem 2.2.1. A generalization of this Euclidean imbedding to compact manifolds with fixed metrics can be found in [Fo]. The Sobolev constants in our theorem in general are not as nice as the fixed metric case or as those in Theorem 6.2.1.

**Theorem 6.3.1** *Let* $\mathbf{M}$ *be a compact Riemann manifold with dimension* $n \ge 3$ *and the metrics* $g = g(t)$ *evolve by the Ricci flow* $\partial_t g = -2Ric$. *Let* $A$ *and* $B$ *be positive numbers such that the* $L^2$ *Sobolev inequality for* $(\mathbf{M}, g(0))$ *holds, i.e. for any* $v \in W^{1,2}(\mathbf{M})$,

$$\left( \int_{\mathbf{M}} v^{2n/(n-2)}\, d\mu(g(0)) \right)^{(n-2)/n} \le A \int_{\mathbf{M}} |\nabla v|^2\, d\mu(g(0)) + B \int_{\mathbf{M}} v^2\, d\mu(g(0)).$$

*Also define* $\lambda_0$ *be the first eigenvalue of Perelman's* $F$ *entropy. i.e*

$$\lambda_0 = \inf_{\|v\|_2 = 1} \int_{\mathbf{M}} (4|\nabla v|^2 + Rv^2) d\mu(g(0))$$

*Suppose the Ricci flow exists for* $t \in [0, T_0)$ *where* $T_0 \le \infty$ *is the life span. Given* $\alpha > 0$ *and* $\rho > 0$, *define*

$$\Phi(\rho) = e^{\alpha \rho^{n/(n-1)}} - \Sigma_{k=0}^{n-1} \frac{\alpha^k}{k!} \rho^{nk/(n-1)}.$$

*Then for any* $u \in W^{1,n}(\mathbf{M}, g(t))$, *it holds*

$$\int_{\mathbf{M}} \Phi \left( \frac{|u(x)|}{\|\nabla u\|_n + \|\sqrt{\frac{1}{4}R^+ + h}\ u\|_n} \right) d\mu(g(t)) \le C(n, \alpha),$$

*provided that* $\alpha < \frac{b}{[\Theta(\mathbf{M}) h_2 \exp(t h_1)]^{n/(n-1)}}$. *Here, the constants are defined as follows.*

*(1).* $b > 0$ *is a numerical constant;*

*(2).*

$$h_1 = \begin{cases} 0, & \text{if} \quad \lambda_0 > 0; \\ h_1(A, B, \lambda_0, \sup R^-(\cdot, 0)) \ge 0, & \text{if} \quad \lambda_0 < 0; \end{cases}$$

*(3).* $h_2 = h_2(A, B, \lambda_0) > 0$;
*(4).*

$$h = \begin{cases} 1, & \textit{if} \quad \lambda_0 > 0; \\ (h_1 - b_2\lambda_0) + 1, & \textit{for a positive constant} \quad b_2, \quad \textit{if} \quad \lambda_0 < 0; \end{cases}$$

*(5).*

$$\Theta(\mathbf{M}) = \sup_{\beta \in (1, n/(n-1))} \sup_x \times \left( [n - \beta(n-1)] \int_{\mathbf{M}} \frac{\exp(-e_0 d(x,y,t))}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \right)^{1/\beta}$$

*where $e_0 > 0$ is a small positive number.*

The statements in the theorem take particularly succinct forms in two special cases. The first one is when the Ricci curvature is bounded from below by negative constants, due to classical volume comparison theorem. The other is the 3 dimension case, which is presented as a corollary.

**Corollary 6.3.1** *Let $\mathbf{M}$ be a 3 dimensional, orientable compact manifold and the metrics $g = g(t)$ evolve by the Ricci flow $\partial_t g = -2Ric$, with normalized initial metric.*

*Suppose the scalar curvature for $g(0)$ is nonnegative. Then, there exists a positive number $\alpha$, depending only on the initial metric such that the critical Sobolev imbedding holds for all $t$ during the life span of the Ricci flow:*

*for any $u \in W^{1,n}(\mathbf{M}, g(t))$, there exists a positive constant $C(\alpha, g(0))$, depending only on $\alpha$ and $g(0)$, such that*

$$\int_{\mathbf{M}} \Phi \left( \frac{|u(x)|}{\|\nabla u\|_n + \|\sqrt{\frac{1}{4}R + 1}\ u\|_n} \right) d\mu(g(t)) \leq C(\alpha, g(0)).$$

*Here $n = 3$, and as in the theorem,*

$$\Phi(\rho) = e^{\alpha \rho^{n/(n-1)}} - \Sigma_{k=0}^{n-1} \frac{\alpha^k}{k!} \rho^{nk/(n-1)}.$$

Proof. (of the corollary) By the maximum principle, we know that the scalar curvature immediately becomes positive when $t > 0$, unless the manifold is Ricci flat. In this special case the metric is stationary and the result is well known [Fo]. Hence we can just assume that the initial scalar curvature is positive.

Since the initial condition is normalized, we know that the Ricci flow is controlled by the initial value explicitly in a fixed time interval, say $[0, \delta]$. So we only need to prove the corollary when $t \geq \delta$. For convenience we just take $\delta = 1$ and assume $t \geq 1$. Since the scalar curvature is positive, the life span of the flow is finite though.

By the assumption on the initial scalar curvature, we know that $\lambda_0 > 0$, $h = 1$, and $h_1 = 0$ in the theorem. So the corollary follows from the theorem once we can show the following quantity $\Theta(\mathbf{M})$ is uniformly bounded when $t \geq 1$. Here

$$\Theta(\mathbf{M}) = \sup_{\beta \in (1, n/(n-1))} \sup_x J(x, t, \beta),$$

$$J(x,t,\beta) \equiv \left( [n - \beta(n-1)] \int_{\mathbf{M}} \frac{\exp(-e_0 d(x,y,t))}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \right)^{1/\beta}. \tag{6.3.1}$$

To continue, we need to get a little ahead of ourselves by looking at Theorem 7.5.1, which establishes the canonical neighborhood property of 3 dimensional Ricci flow. For a fixed small positive number $\epsilon$ and for the whole finite life span of the flow, let $r_0$ be the parameter in the canonical neighborhood property with accuracy $\epsilon$. For $t \geq 1$, $r_0 > 0$, and $x \in \mathbf{M}$, consider the ball $B(x, r_0, t)$. We have three cases to deal with.

Case 1 is when the scalar curvature is bounded from above by $1/r_0^2$ in the ball.

Then by Hamilton-Ivey pinching Theorem 5.2.4, we know that

$$|Ric(y,t)| \leq 2R(y,t) + C \leq 2/r_0^2 + C, \qquad y \in B(x, r_0, t),$$

for a positive constant $C$. By standard volume comparison theorem, for any $r \in (0, r_0]$, we have

$$|B(x_0, r, t)|_{g(t)} \leq e^{c(r_0^{-1}+1)r} c_n r^n, \quad n = 3.$$

Therefore

$$
\begin{aligned}
&[n-\beta(n-1)]\int_{\mathbf{M}} \frac{\exp(-e_0 d(x,y,t))}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \\
&\le [n-\beta(n-1)][\int_{d(x,y,t)\le r_0} \frac{1}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \\
&\quad + \int_{d(x,y,t)> r_0} \frac{1}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t))] \\
&\le [n-\beta(n-1)]\int_{d(x,y,t)\le r_0} \frac{1}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \\
&\quad + r_0^{-\beta(n-1)} Vol(\mathbf{M}, g(t)).
\end{aligned}
$$

Notice that

$$
\begin{aligned}
&\int_{d(x,y,t)\le r_0} \frac{1}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \\
&= \Sigma_{i=0}^{\infty} \int_{2^{-(i+1)} r_0 \le d(x,y,t) \le 2^{-i} r_0} \frac{1}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \\
&\le \Sigma_{i=0}^{\infty} (2^{2+i}/r_0)^{(n-1)\beta} \int_{d(x,y,t)\le 2^{-i} r_0} d\mu(g(t)) \\
&\le c_n e^{c(r_0^{-1}+1) r_0}\ \Sigma_{i=0}^{\infty} (2^{2+i}/r_0)^{(n-1)\beta} (2^{-i} r_0)^n \\
&\le c_n e^{c(r_0^{-1}+1) r_0} r_0^{n-(n-1)\beta} \frac{1}{n-(n-1)\beta}.
\end{aligned}
$$

Here the second from last step is by the volume upper bound mentioned above. Combining the last two paragraphs, we deduce

$$
[n-\beta(n-1)]\int_{\mathbf{M}} \frac{\exp(-e_0 d(x,y,t))}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \le c(r_0)[1 + vol(\mathbf{M}, g(0))]. \tag{6.3.2}
$$

Here we just used the fact that the volume of $\mathbf{M}$ is decreasing in time since the scalar curvature is nonnegative. Therefore, in this case

$$
J(x,t,\beta) \le C(r_0, g(0)).
$$

By Theorem 7.5.1, $r_0$ depends only on $\epsilon$, the initial metric and the life span. The life span is bounded from above by a quantity involving only $\min R(\cdot, 0)$. So $J(x,t,\beta)$ depends only on the initial value in this case.

Case 2 is when the ball $B(x, r_0, t)$ contains a point $y$ such that $R(y,t) = \frac{1}{r_0^2}$.

According to Perelman's singularity structure theorem [P1] (Theorem 7.5.1 here), the ball $B(x, r_0, t)$ under the re-scaled metric $r_0^{-2} g(t)$ is $\epsilon$ close, in $C^{[\epsilon^{-1}]}$ topology, to the corresponding ball in an (ancient) $\kappa$ solution. Let us use $\tilde{g}$ and $\tilde{d}$ to denote the metric $r_0^{-2} g(t)$ and the corresponding distance. Then

$$\int_{d(x,y,t)\le r_0} \frac{1}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \\ = r_0^{n-(n-1)\beta} \int_{\tilde{d}(x,y,t)\le 1} \frac{1}{\tilde{d}(x,y,t)^{\beta(n-1)}} d\mu(\tilde{g}).$$

The Ricci curvature for $\tilde{g}$ in the unit ball is bounded from below by a negative constant depending only on $\epsilon$. This is so because the near by $\kappa$ solution has nonnegative sectional curvature. Therefore by the same computation as above, we have

$$\int_{d(x,y,t)\le r_0} \frac{1}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \le C(\epsilon) r_0^{n-(n-1)\beta} \frac{1}{n-(n-1)\beta}.$$

Fixing a small $\epsilon$, this shows, as in Case 1,

$$J(x,t,\beta)^\beta = [n - \beta(n-1)] \int_{\mathbf{M}} \frac{\exp(-e_0 d(x,y,t))}{d(x,y,t)^{\beta(n-1)}} d\mu(g(t)) \\ \le c(r_0)[1 + vol(\mathbf{M}, g(0))]. \tag{6.3.3}$$

Case 3 is when every point $y$ in the ball $B(x, r_0, t)$ satisfies $R(y,t) > \frac{1}{r_0^2}$.

According to Theorem 7.5.1, the ball is either a compact manifold with positive curvature or it is contained in an $\epsilon$ horn or capped $\epsilon$ horn. The definitions of these two objects can be found in Definition 8.1.1. If $B(x, r_0, t)$ is a compact manifold with positive curvature, then we again can use the classical volume comparison theorem to prove

$$J(x,t,\beta) \le C(r_0, g(0)).$$

Now we assume that $B(x, r_0, t)$ is contained in an $\epsilon$ horn. We claim that there exists a constant $C > 0$ such that

$$|B(x,r,t)|_{g(t)}/r^3 \le C, \quad r \in [0, r_0].$$

If the ball $B(x,r,t)$ is contained in one $\epsilon$ neck, then the claim is obviously true. Suppose, for a $k \ge 3$, the ball $B(x,r,t)$ contains $k-2$ $\epsilon$

necks and $B(x, r, t)$ is contained in the union of $k$ $\epsilon$ necks. Let $x_i$ be at the center of the $i$-th $\epsilon$ neck, where $i = 1, \ldots, k$. Then

$$r \sim \epsilon^{-1} \Sigma_{i=1}^k R(x_i)^{-1/2}.$$

Here $R(x_i)$ is the scalar curvature at $x_i$ and time $t$. Note the volume of the $i$-th $\epsilon$ neck is $CR(x_i)^{-3/2}\epsilon^{-1}$. Hence

$$|B(x, r, t)|_{g(t)} \sim \epsilon^{-1} \Sigma_{i=1}^k R(x_i)^{-3/2} \le \epsilon^2 \left[ \epsilon^{-1} \Sigma_{i=1}^k R(x_i)^{-1/2} \right]^3 \le Cr^3.$$

This proves the claim.

Now we can just proceed as in Case 1, using the claim to replace the classical volume comparison theorem there to deduce

$$J(x, t, \beta) \le C(r_0, g(0)).$$

Finally, if $B(x, r_0, t)$ is contained in a capped $\epsilon$ horn, a proof of the bound on $J(x, t, \beta)$ can be done similarly.

Thus we have shown in all cases that $J(x, t, \beta)$ has a bound which depends only on the initial metric. By (6.3.1), this bound implies the bound on $\Theta(\mathbf{M})$ and the corollary. □

**Proof of Theorem 6.3.1**

We follow the notations of Theorem 6.2.1. Suppose the Ricci flow exists in the time interval $[0, t_0]$.

*Step 1.* Off diagonal bound for $p$, the heat kernel of $\Delta - R/4$ at time $t_0$.

In this step, we prove a Gaussian type upper bound for $p$, the heat kernel for $\Delta - R/4$ under the metric $g(t_0)$. We emphasize that $\Delta$ and $R$ are with respect to the fixed metric $g(t_0)$. We will again use $dx$, $d(x, y)$ and $B(x, r)$ to denote the volume element, distance and geodesic balls under $g(t_0)$ respectively. If the scalar curvature $R$ were absent, then the Gaussian upper bound follows immediately from the on diagonal bound (6.2.12) and the main theorem in Grigoryan's paper [Gr]. In our case we will couple Grigoryan's method with Perelman's monotonicity for the $F$ entropy to treat the scalar curvature term.

This step is divided into a few substeps.

*Step 1.1.* Monotonicity of certain weighted $L^2$ norms of solutions.

Let $u$ be a positive solution to the equation

$$\Delta u - \frac{1}{4} R u - \partial_s u = 0$$

where the underlying manifold is $(\mathbf{M}, g(t_0))$. Given a weight function $e^{\xi(x,s)}$ to be specified later, we compute

$$\frac{d}{ds}\int_{\mathbf{M}} u^2 e^{\xi} dx = \int_{\mathbf{M}} u^2 e^{\xi} \partial_s \xi dx + \int_{\mathbf{M}} 2u(\Delta u - \frac{1}{4}Ru)e^{\xi} dx. \tag{6.3.4}$$

Note that

$$\begin{aligned}\int_{\mathbf{M}} u\Delta u e^{\xi} dx &= -\int_{\mathbf{M}} \nabla u \nabla (u e^{\xi}) dx = -\int_{\mathbf{M}} \nabla u \nabla (u e^{\xi/2}\ e^{\xi/2}) dx \\ &= -\int_{\mathbf{M}} \nabla u \left[\nabla (u e^{\xi/2})\ e^{\xi/2} + u e^{\xi/2} \nabla e^{\xi/2}\right] dx \\ &= -\int_{\mathbf{M}} |\nabla (u e^{\xi/2})|^2 dx + \int_{\mathbf{M}} u^2 |\nabla e^{\xi/2}|^2 dx.\end{aligned}$$

Substituting this to the right-hand side of (6.3.4), we obtain

$$\begin{aligned}\frac{d}{ds}\int_{\mathbf{M}} u^2 e^{\xi} dx = &\int_{\mathbf{M}} (\partial_s \xi + \frac{1}{2}|\nabla \xi|^2) u^2 e^{\xi} dx \\ &- 2\int_{\mathbf{M}} \left[|\nabla (u e^{\xi/2})|^2 + \frac{1}{4} R (u e^{\xi/2})^2\right] dx.\end{aligned}$$

Let

$$\lambda(t_0) = \inf_{u \neq 0} \frac{\int_{\mathbf{M}} \left[4|\nabla (u e^{\xi/2})|^2 + R(u e^{\xi/2})^2\right] dx}{\int_{\mathbf{M}} u^2 e^{\xi} dx}.$$

Recall that $dx$ is actually the volume form $d\mu(g(t_0))$ where $t_0$ is a moment during the Ricci flow. Therefore $\lambda(t_0)$ is actually the infimum of Perelman's $F$ entropy at time $t_0$. By Perelman [P1] (Theorem 6.1.1 here), it holds

$$\lambda(t_0) \geq \lambda(0) = \lambda_0.$$

Hence

$$\frac{d}{ds}\int_{\mathbf{M}} u^2 e^{\xi} dx \leq \int_{\mathbf{M}} (\partial_s \xi + \frac{1}{2}|\nabla \xi|^2) u^2 e^{\xi} dx - \frac{1}{2}\lambda_0 \int_{\mathbf{M}} u^2 e^{\xi} dx.$$

If we choose $\xi$ so that

$$\partial_s \xi + \frac{1}{2}|\nabla \xi|^2 \leq 0,$$

then we deduce

$$\int_{\mathbf{M}} u^2 e^{\xi} dx\Big|_{s_1} \leq e^{-\lambda_0 (s_1 - s_2)/2} \int_{\mathbf{M}} u^2 e^{\xi} dx\Big|_{s_2} \tag{6.3.5}$$

when $s_2 < s_1$.

*Step 1.2.* Once we have the above monotone formula, we can use the idea in [Gr] to prove the Gaussian upper bound. Since the proof is not identical to that in [Gr], due to the presence of the exponential term in (6.3.5), we will present it here.

Taking a point $x \in \mathbf{M}$ and $s, r > 0$, define

$$I_r(s) = \int_{\mathbf{M}-B(x,r)} u^2(y,s)dy. \tag{6.3.6}$$

We aim to show that $I_r(s)$ has certain exponential decay for $u(y,s) = p(y,s,x)$. Picking two numbers $A$ and $\sigma_0$ such that $A \ge 2$ and $\sigma_0 > s$, we choose

$$\xi = \xi(y,s) = \begin{cases} -\frac{(r-d(x,y))^2}{A(\sigma_0 - s)}, & d(x,y) \le r; \\ 0, \quad d(x,y) > r. \end{cases}$$

Then, for $y \in B(x,r)$,

$$\partial_s \xi + \frac{1}{2}|\nabla \xi|^2 = -\frac{(r-d(x,y))^2}{A(\sigma_0 - s)^2} + \frac{2(r-d(x,y))^2}{A^2(\sigma_0 - s)^2} \le 0.$$

By (6.3.5), we have, for $s_2 < s_1 < s_0$,

$$\int_{\mathbf{M}} u^2 e^{\xi} dx\big|_{s_1} \le e^{-\lambda_0(s_1 - s_2)/2} \int_{\mathbf{M}} u^2 e^{\xi} dx\big|_{s_2}.$$

Since $\xi(y,s) = 0$ when $d(x,y) \ge r$, this implies

$$\begin{aligned} I_r(s_1) = \int_{\mathbf{M}-B(x,r)} u^2(y,s_1)dy &\le \int_{\mathbf{M}} u^2(y,s_1)e^{\xi(y,s_1)}dy \\ &\le \int_{\mathbf{M}} u^2(y,s_2)e^{\xi(y,s_2)}dy \; e^{-\lambda_0(s_1-s_2)/2}. \end{aligned}$$

For a number $\rho < r$, we can write this inequality as

$$\begin{aligned} I_r(s_1) \le \int_{B(x,\rho)} & u^2(y,s_2)e^{\xi(y,s_2)}dy \; e^{-\lambda_0(s_1-s_2)/2} \\ &+ e^{-\lambda_0(s_1-s_2)/2} \int_{\mathbf{M}-B(x,\rho)} u^2(y,s_2)e^{\xi(y,s_2)}dy. \end{aligned}$$

This shows

$$I_r(s_1) \le e^{-\lambda_0(s_1-s_2)/2} \left[ I_\rho(s_2) + e^{-(r-\rho)^2/(A(\sigma_0 - s_2))} \int_{B(x,\rho)} u^2(y,s_2)dy \right].$$

So far the bound holds for all positive solutions to the equation $\Delta u - Ru/4 - \partial_s u = 0$. Now we take $u(y,s) = p(y,s,x)$ the heat kernel. For this $u$, it holds

$$\int_{B(x,\rho)} u^2(y,s_2)dy \le \int_{\mathbf{M}} p^2(y,s_2,x)dy = p(x,2s_2,x) \le \frac{1}{f(2s_2)}.$$

Here $f$ is given by the right-hand side of the on-diagonal bound in (6.2.15) and (6.2.16). We know that there are two positive constants $h_1$ and $h_2$ such that, for all $T>0$, it holds

$$\frac{1}{f(T)} = \frac{h_2 \ \exp[(T+t_0)h_1]}{T^{n/2}}. \tag{6.3.7}$$

where

$$h_2 = h_2(A,B,\lambda_0)$$

and

$$h_1 = \begin{cases} 0, \quad \text{if} \quad \lambda_0 > 0; \\ h_1(A,B,\lambda_0, \sup R^-(\cdot,0)) \ge 0, \ \text{otherwise}; \end{cases} \tag{6.3.8}$$

Thus we reach the inequality

$$I_r(s_1) \le e^{-\lambda_0(s_1-s_2)/2}\left[I_\rho(s_2) + e^{-(r-\rho)^2/(A(\sigma_0-s_2))} \ \frac{1}{f(2s_2)}\right].$$

Observe the above inequality depends on $\sigma_0$ only in the exponential term, which is the parameter in the definition of $\xi$. So we can just take $\sigma_0 = s_1$ to get

$$I_r(s_1) \le e^{-\lambda_0(s_1-s_2)/2}\left[I_\rho(s_2) + e^{-(r-\rho)^2/(A(s_1-s_2))} \ \frac{1}{f(2s_2)}\right] \tag{6.3.9}$$

where $r>\rho$, $s_1>s_2$ and $A\ge 2$.

Now fixing $r,s>0$, we define the sequences, as in [Gr]

$$r_k = \left(\frac{1}{2}+\frac{1}{k+2}\right)r, \qquad s_k = \frac{s}{a^k}, \quad k=0,1,2,\dots$$

where $a>1$ will be chosen later. Applying (6.3.9), we deduce

$$\begin{aligned} I_{r_k}(s_k) \le{}& e^{-\lambda_0(s_k-s_{k+1})/2} \\ &\times, \left[I_{r_{k+1}}(s_{k+1}) + e^{-(r_k-r_{k+1})^2/(A(s_k-s_{k+1}))} \ \frac{1}{f(2s_{k+1})}\right] \end{aligned} \tag{6.3.10}$$

Remember that $I_{r_k}(s_k) = \int_{\mathbf{M}-B(x,r_k)} p^2(y,s_k,x)dy$. When $k \to \infty$, $s_k \to 0$ and $p(y,s_k,x) \to \delta(y,x)$ which is concentrated at the point $x$. Hence $\lim_{k\to\infty} I_{r_k}(s_k) = 0$. This argument can easily be made rigorous by approximating $p$ with regular solutions whose initial value is supported in $B(x,r/2)$.

After iterations of (6.3.10), we obtain

$$I_r(s) = I_{r_0}(s_0) \le \Sigma_{k=0}^{\infty} \frac{1}{f(2s_{k+1})} e^{-(r_k - r_{k+1})^2/(A(s_k - s_{k+1}))} e^{-\lambda_0(s_0 - s_{k+1})/2}.$$

Using the relation

$$r_k - r_{k+1} \ge r/(k+3)^2, \qquad s_k - s_{k+1} = (a-1)s/a^{k+1},$$

we arrive at

$$I_r(s) \le \Sigma_{k=0}^{\infty} \frac{1}{f(2s_{k+1})} \exp\left(-\frac{a^{k+1}\ r^2}{(k+3)^4\ (a-1)\ As}\right)\ e^{-\lambda_0(s-s_{k+1})/2}.$$

Using (6.3.7)

$$\frac{1}{f(2s_{k+1})} \le \frac{a^{(k+1)n/2} h_2 \exp[(s+t_0)h_1]}{s^{n/2}}.$$

Substituting this to the last inequality concerning $I_r(s)$, we deduce, for some constant $c = c(a) > 0$, depending on the sign of $\lambda_0$,

$$\begin{aligned} I_r(s) \le \frac{h_2 \exp[(s+t_0)h_1]}{s^{n/2}} e^{-\lambda_0 s c(a)} \Sigma_{k=0}^{\infty} a^{(k+1)n/2} \\ \times \exp\left(-\frac{a^{k+1}\ r^2}{(k+3)^4\ (a-1)\ As}\right). \end{aligned}$$

By making the constant $a$ sufficiently large, it is easy to check that

$$I_r(s) = \int_{\mathbf{M}-B(x,r)} p^2(y,s,x)dy \le \frac{h_2 \exp[(s+t_0)h_1]}{s^{n/2}} e^{-\lambda_0 s c(a)} e^{-h_3 r^2/s} \tag{6.3.11}$$

for some numerical constant $h_3 > 0$.

Define, for a small positive number $m < h_3$,

$$E_m(s) = \int_{\mathbf{M}} p^2(y,s,x) e^{m d^2(y,x)/s} dy.$$

Take $r = \sqrt{s}$ and split the integral for $E_m(s)$ as

$$\begin{aligned} E_m(s) \le e^m \int_{B(x,r)} p^2(y,s,x)dy \\ + \Sigma_{k=0}^{\infty} \int_{2^k r < d(x,y) \le 2^{k+1} r} p^2(y,s,x) e^{md^2(y,x)/s} dy \\ \le e^m \int_{B(x,r)} p^2(y,s,x)dy \\ + \Sigma_{k=0}^{\infty} e^{m 2^{2(k+1)}} \int_{2^k r < d(x,y) \le 2^{k+1} r} p^2(y,s,x)dy. \end{aligned} \tag{6.3.12}$$

Using integration by parts and the monotonicity of Perelman's $F$ entropy again, it is easy to see that

$$\begin{aligned} \frac{d}{ds} \int_{\mathbf{M}} p^2(y,s,x)dy = 2 \int_{\mathbf{M}} p(\Delta p - Rp/4)dy \\ = -2 \int_{\mathbf{M}} (|\nabla p|^2 + Rp^2/4)dy \\ \le -\frac{\lambda_0}{2} \int_{\mathbf{M}} p^2(y,s,x)dy. \end{aligned}$$

Therefore

$$\begin{aligned} \int_{\mathbf{M}} p^2(y,s,x)dy \le e^{-\lambda_0 s/4} \int_{\mathbf{M}} p^2(y,s/2,x)dy \\ = e^{-\lambda_0 s/4} p(x,s,x) \le e^{-\lambda_0 s/4}/f(s). \end{aligned}$$

Applying this and (6.3.11) to the last two terms of (6.3.12), we have, for a sufficiently small $m > 0$,

$$E_m(s) = \int_{\mathbf{M}} p^2(y,s,x) e^{md^2(y,x)/s} dy \le \frac{c_1 h_2 \exp[(s+t_0)h_1]}{s^{n/2}} e^{-c_2 \lambda_0 s} \tag{6.3.13}$$

where $c_1$ and $c_2$ are positive, numerical constants; the constants $h_1$ and $h_2$ depend only on the initial metric, as given by (6.3.8). It is well known ( [Gr] e.g.) that (6.3.13) implies

$$p(x,s,y) \le b_1 h_2 e^{t_0 h_1} e^{(h_1 - b_2 \lambda_0)s} \frac{e^{-b_3 d^2(x,y)/s}}{s^{n/2}}. \tag{6.3.14}$$

Here $b_1, b_2, b_3$ are positive, numerical constants.

*Step 2.* Integral gradient bound for Green's functions.

Consider $\Gamma = \Gamma(x,y)$, the Green's function of the operator

$$L_h = \Delta - \frac{R^+}{4} - h$$

where $h = (h_1 - b_2\lambda_0)^+ + 1$. Here $h_1, b_2, \lambda_0$ are the same as those in (6.3.14). Denote by $p_h$ the heat kernel of $L_h$. Then (6.3.14) shows

$$p_h \le p e^{-hs} \le b_1 h_2 e^{t_0 h_1} e^{-s} \, \frac{e^{-b_3 d^2(x,y)/s}}{s^{n/2}}. \tag{6.3.15}$$

Therefore, there exits $C > 0$ and $b_4 > 0$ such that

$$\Gamma(x,y) = \int_0^\infty p_h(y,s,x) ds \le C h_2 e^{t_0 h_1} \frac{e^{-b_4 d(x,y)}}{d(x,y)^{n-2}}. \tag{6.3.16}$$

When $y \neq x$, the Green's function $\Gamma(x,y)$, as a function of $y$, satisfies the equation

$$\Delta\Gamma - \frac{R^+}{4}\Gamma - h\Gamma = 0.$$

Applying a standard argument using test functions of the form $\Gamma\phi^2$, we know that

$$\begin{aligned} &\int_{2^k \le d(x,y) \le 2^{k+1}} |\nabla\Gamma(x,y)|^2 dy + \int_{2^k \le d(x,y) \le 2^{k+1}} (\frac{R^+}{4} + h)\Gamma^2(x,y) dy \\ &\qquad \le \frac{c}{2^{2k}} \int_{2^{k-1} \le d(x,y) \le 2^{k+2}} \Gamma^2(x,y) dy. \end{aligned} \tag{6.3.17}$$

In the above, $\phi$ is a cut-off function and $c$ is a positive constant.

For any number $\beta \in (0,2)$, by Hölder's inequality and (6.3.17),

$$\begin{aligned} &\int_{2^k \le d(x,y) \le 2^{k+1}} |\nabla\Gamma(x,y)|^\beta dy \\ &\le \left( \int_{2^k \le d(x,y) \le 2^{k+1}} |\nabla\Gamma(x,y)|^2 dy \right)^{\beta/2} \left( \int_{2^k \le d(x,y) \le 2^{k+1}} dy \right)^{1-(\beta/2)} \\ &\le \left( \frac{c}{2^{2k}} \int_{2^{k-1} \le d(x,y) \le 2^{k+2}} \Gamma^2(x,y) dy \right)^{\beta/2} |B(x,2^{k+1}) - B(x,2^k)|^{1-(\beta/2)}. \end{aligned}$$

Using (6.3.16) on the last term, we deduce

$$\int_{2^k \le d(x,y) \le 2^{k+1}} |\nabla\Gamma(x,y)|^\beta dy \le \frac{ch_2^\beta \exp(t_0 h_1 \beta)\ \exp(-2^{(k-2)}\beta b_4)}{2^{k\beta} 2^{(k-1)(n-2)\beta}}$$
$$\cdot \left(\int_{2^{k-1} \le d(x,y) \le 2^{k+2}} dy\right)^{\beta/2} \quad |B(x,2^{k+1}) - B(x,2^k)|^{1-(\beta/2)}$$
$$\le \frac{ch_2^\beta \exp(t_0 h_1 \beta)\ \exp(-2^{(k-2)}\beta b_4)}{2^{k\beta} 2^{(k-1)(n-2)\beta}}\ |B(x,2^{k+2}) - B(x,2^{k-1})|.$$

Summing up for all integers $k$, we have

$$\int_{\mathbf{M}} |\nabla\Gamma(x,y)|^\beta dy$$
$$\le ch_2^\beta \exp(t_0 h_1 \beta)\ \Sigma_{k=-\infty}^{\infty} \frac{\exp(-2^{(k-2)}\beta b_4)}{2^{k\beta} 2^{(k-1)(n-2)\beta}}\ |B(x,2^{k+2}) - B(x,2^{k-1})|.$$

Since

$$|B(x,2^{k+2}) - B(x,2^{k-1})| = |B(x,2^{k+2}) - B(x,2^{k+1})|$$
$$+ |B(x,2^{k+1}) - B(x,2^k)| + |B(x,2^k) - B(x,2^{k-1})|,$$

we can split the preceding sum to three and shift the indices to get

$$\int_{\mathbf{M}} |\nabla\Gamma(x,y)|^\beta dy \le c \exp(t_0 h_1 \beta)\ \Sigma_{k=-\infty}^{\infty} \left[\frac{\exp(-2^{(k-4)}\beta b_4)}{2^{(k-2)\beta} 2^{(k-3)(n-2)\beta}}\right.$$
$$\left. + \frac{\exp(-2^{(k-3)}\beta b_4)}{2^{(k-1)\beta} 2^{(k-2)(n-2)\beta}} + \frac{\exp(-2^{(k-2)}\beta b_4)}{2^{k\beta} 2^{(k-1)(n-2)\beta}}\right]$$
$$\times |B(x,2^k) - B(x,2^{k-1})|\, h_2^\beta.$$

This tells us

$$\int_{\mathbf{M}} |\nabla\Gamma(x,y)|^\beta dy$$
$$\le ch_2^\beta \exp(t_0 h_1 \beta)\ \Sigma_{k=-\infty}^{\infty} \frac{\exp(-2^{-2} 2^{(k-2)}\beta b_4)}{2^{k\beta} 2^{(k-1)(n-2)\beta}} |B(x,2^k) - B(x,2^{k-1})|$$
$$= ch_2^\beta \exp(t_0 h_1 \beta)\ \Sigma_{k=-\infty}^{\infty} \int_{2^{k-1} \le d(x,y) \le 2^k} \frac{\exp(-2^{-3} 2^{(k-1)}\beta b_4)}{2^{\beta k(n-1) - \beta(n-2)}} dy$$
$$\le ch_2^\beta \exp(t_0 h_1 \beta)\ \Sigma_{k=-\infty}^{\infty} \int_{2^{k-1} \le d(x,y) \le 2^k} \frac{\exp(-2^{-3} d(x,y)\beta b_4)}{d(x,y)^{\beta(n-1)}} dy.$$

Therefore

$$\int_{\mathbf{M}} |\nabla \Gamma(x,y)|^\beta dy \le c h_2^\beta \exp(t_0 h_1 \beta) \int_{\mathbf{M}} \frac{\exp(-2^{-3} d(x,y)\beta b_4)}{d(x,y)^{\beta(n-1)}} dy. \tag{6.3.18}$$

By (6.3.17),

$$\begin{aligned}
&\int_{2^k \le d(x,y) \le 2^{k+1}} [\sqrt{R^+ 4^{-1} + h}\, \Gamma(x,y)]^\beta dy \\
&\le \left( \int_{2^k \le d(x,y) \le 2^{k+1}} [\sqrt{R^+ 4^{-1} + h}\, \Gamma(x,y)]^2 dy \right)^{\beta/2} \\
&\qquad \times \left( \int_{2^k \le d(x,y) \le 2^{k+1}} dy \right)^{1-(\beta/2)} \\
&\le \left( \frac{c}{2^{2k}} \int_{2^{k-1} \le d(x,y) \le 2^{k+2}} \Gamma^2(x,y) dy \right)^{\beta/2} |B(x,2^{k+1}) - B(x,2^k)|^{1-(\beta/2)},
\end{aligned}$$

From here we can use exactly the same argument as in the previous paragraph to show

$$\begin{aligned}
\int_{\mathbf{M}} [\sqrt{R^+ 4^{-1} + h}\, \Gamma(x,y)]^\beta dy &\le c h_2^\beta \exp(t_0 h_1 \beta) \\
&\qquad \int_{\mathbf{M}} \frac{\exp(-2^{-3} d(x,y)\beta b_4)}{d(x,y)^{\beta(n-1)}} dy.
\end{aligned} \tag{6.3.19}$$

*Step 3.* Proof of the Critical Sobolev inequality.

Once we obtain the integral gradient bounds (6.3.18) and (6.3.19), the proof of the theorem can be finished in the standard way, as in Chapter 2.

Let $u$ be a smooth function on $\mathbf{M}$. Then

$$\Delta u - (\frac{1}{4} R^+ + h) u = \Delta u - (\frac{1}{4} R^+ + h) u.$$

Since $\Gamma$ is the fundamental solution, it holds

$$u = -\int_{\mathbf{M}} \Gamma(x,y) [\Delta u - (\frac{1}{4} R^+ + h) u](y) dy.$$

Upon integration by parts, it follows

$$u = \int_{\mathbf{M}} \nabla \Gamma(x,y) \nabla u(y) dy + \int_{\mathbf{M}} \Gamma(x,y) (\frac{1}{4} R^+ + h) u(y) dy.$$

For any $q > n$, applying Young's inequality with the parameters satisfying

$$\frac{1}{n} + \frac{1}{\beta} = 1 + \frac{1}{q},$$

we deduce

$$\begin{aligned}\|u\|_q \le \sup_x \|\nabla\Gamma(\cdot,x)\|_\beta \ \|\nabla u\|_n \\ + \sup_x \|\Gamma(\cdot,x)\sqrt{\frac{1}{4}R^+ + h}\|_\beta \ \|\sqrt{\frac{1}{4}R^+ + h}u\|_n.\end{aligned}$$

According to (6.3.18) and (6.3.19), this implies

$$\begin{aligned}\|u\|_q \le ch_2 \exp(t_0 h_1) \ \sup_x \left( \int_{\mathbf{M}} \frac{\exp(-2^{-3}d(x,y)\beta b_4)}{d(x,y)^{\beta(n-1)}} dy \right)^{1/\beta} \\ \times \left[ \|\nabla u\|_n + \|\sqrt{\frac{1}{4}R^+ + h} \ u\|_n \right].\end{aligned}$$

Note that $\beta \in (1, n/(n-1))$ and so $\beta \le 2$. Hence, there exists a positive number $e_0$, such that

$$\begin{aligned}\|u\|_q \le ch_2 \exp(t_0 h_1) \ \sup_x \left( \int_{\mathbf{M}} \frac{\exp(-e_0 d(x,y))}{d(x,y)^{\beta(n-1)}} dy \right)^{1/\beta} \\ \times (\|\nabla u\|_n + \|\sqrt{\frac{1}{4}R^+ + h}u\|_n).\end{aligned} \tag{6.3.20}$$

Recall the quantity

$$\Theta(\mathbf{M}) = \sup_{\beta\in(1,n/(n-1))} \sup_x \left( [n - \beta(n-1)] \int_{\mathbf{M}} \frac{\exp(-e_0 d(x,y))}{d(x,y)^{\beta(n-1)}} dy \right)^{1/\beta}. \tag{6.3.21}$$

Then (6.3.20) implies

$$\|u\|_q \le c\Theta(\mathbf{M}) h_2 \exp(t_0 h_1) [n-\beta(n-1)]^{-1/\beta} \ (\|\nabla u\|_n + \|\sqrt{\frac{1}{4}R^+ + h}\, u\|_n). \tag{6.3.22}$$

Given integers $k = n, (n+1), \ldots,$ we take $q = nk/(n-1)$. Then the above becomes

$$\begin{aligned}\int_{\mathbf{M}} \left( \frac{|u|(y)}{(\|\nabla u\|_n + \|\sqrt{\frac{1}{4}R^+ + h}u\|_n)} \right)^{nk/(n-1)} dy \\ \le [c\Theta(\mathbf{M}) h_2 \exp(t_0 h_1)]^{nk/(n-1)} \left( \frac{k+1}{n} \right)^{k+1}.\end{aligned}$$

For a number $\alpha > 0$, this inequality shows

$$\int_{\mathbf{M}} \Sigma_{k=n}^{\infty} \frac{\alpha^k}{k!} \left( \frac{|u|(y)}{(\|\nabla u\|_n + \|\sqrt{\frac{1}{4}R^+ + h}u\|_n)} \right)^{nk/(n-1)} dy$$
$$\leq \Sigma_{k=n}^{\infty} \frac{\alpha^k}{k!} [c\Theta(\mathbf{M}) h_2 \exp(t_0 h_1)]^{nk/(n-1)} \left( \frac{k+1}{n} \right)^{k+1}.$$

By Sterling's formula, there exists a constant $b = b(n)$ only, such that the right-hand side of the last inequality is convergent, provided that

$$\alpha < \frac{b}{[\Theta(\mathbf{M}) h_2 \exp(t_0 h_1)]^{n/(n-1)}}. \tag{6.3.23}$$

□

The next theorem is a localized version of Theorem 6.2.1 on the evolution of Sobolev imbedding along Ricci flow. It shows that Sobolev imbedding in a metric ball propagates to a larger ball in future time provided that the curvature tensor is bounded in a portion of the space time. Perelman's no local collapsing theorem II (Theorem 8.2 [P1]) is a special case by Theorem 4.1.2.

**Theorem 6.3.2** *For any $A > 0$, let $g = g(t)$ be a Ricci flow defined for $t \in [0, r_0^2]$ for some $r_0 > 0$. Suppose that $|B(x_0, r_0, 0)|_{g(0)} \geq A^{-1} r_0^n$ and that $|Rm| \leq 1/(n r_0^2)$ for all $(x,t) \in B(x_0, r_0, 0) \times [0, r_0^2]$. Then the following Sobolev imbedding holds in $B(x_0, A r_0, r_0^2)$.*

*For all $v \in W_0^{1,2}(B(x_0, A r_0, r_0^2))$, there exists one $A_2 = C(A, n)$ such that, at $t = r_0^2$,*

$$\left( \int v^{2n/(n-2)} d\mu(g(t)) \right)^{(n-2)/n} \leq A_2 \int (4|\nabla v|^2 + R v^2) d\mu(g(t))$$
$$+ \frac{A_2}{r_0^2} \int v^2 d\mu(g(t)).$$

Proof. We divide the proof into several steps.

*Step 1.* By scaling invariance, we can just take $r_0 = 1$.

Given $\sigma > 0$, define

$$\Lambda = \Lambda_{\sigma^2}(g(1))$$
$$= \inf \{ \int_{B(x_0, A, 1)} [\sigma^2 (4|\nabla v|^2 + R v^2) - v^2 \ln v^2] \, d\mu(g(1)) - n \ln \sigma$$
$$| \; v \in C_0^\infty(B(x_0, A, 1)), \; \|v\|_2 = 1 \}. \tag{6.3.24}$$

We aim to find a lower bound for $\Lambda$ when $\sigma \in [0,1]$. So, without loss of generality we assume $\Lambda \le 0$. Let $v_1$ be a minimizer of the functional in (6.3.24). Recall that the existence and smoothness of the minimizer of the above functional in the whole manifold case were proven in [Ro]. The current case under Dirichlet boundary condition can be dealt with similarly. The proof is left as an exercise. Consequently, $v_1$ is smooth and positive in $B(x_0,1,A)$ and it obeys the equation

$$\begin{cases} \sigma^2(4\Delta v_1 - Rv_1) + 2v_1 \ln v_1 + \Lambda v_1 + n(\ln\sigma)v_1 = 0; \\ v_1(x) = 0, \qquad x \in \partial B(x_0, A, 1). \end{cases} \tag{6.3.25}$$

Here every term is relative to the metric $g(1)$.

Next we define

$$P(v_1) = [\sigma^2(-4\Delta v_1 + Rv_1) - 2v_1 \ln v_1 - n(\ln\sigma)v_1]v_1 = \Lambda v_1^2. \tag{6.3.26}$$

We regard $v_1$ and $P(v_1)$ as functions on $(\mathbf{M}, g(1))$ by assigning zero value outside of their support. The function $v_1$ is Hölder continuous and lies in $W^{1,\infty}(\mathbf{M})$. This fact can be proven by the method in [Ro] using the property that $\partial B(x_0,1,A)$ is Lipschitz. The basic idea is to divide the ball into two parts $D_1$ and $D_2$. In $D_1$, it holds $v_1 < 1$. Then $v_1 \ln v_1$ is bounded. We treat it as a bounded inhomogeneous term in the Laplace equation. Then we know that $v_1$ is Lipschitz. On the other part $D_2$, it holds $v_1 \ge 1$. By Jensen's inequality and $\|v_1\|_2 = 1$, we know that $\ln v_1 \in L^p(D_2)$ for any $p \ge 1$. The standard elliptic theory also shows $v_1$ is Lipschitz. We leave the detailed proof as an exercise again.

Let $u$ be the solution of the conjugate heat equation on the whole manifold and $t \in (0,1)$,

$$\begin{cases} H^*u \equiv \Delta u - Ru + \partial_t u = 0 \\ u(x,1) = v_1^2(x). \end{cases} \tag{6.3.27}$$

Define, for $v = \sqrt{u}$, Perelman's quantity

$$\begin{aligned} P(v) = P(v)(x,t) = [(\sigma^2 + 1 - t)(-4\Delta v + Rv) - 2v \ln v \\ - \frac{n}{2}(\ln(\sigma^2 + 1 - t))v]v. \end{aligned} \tag{6.3.28}$$

This is a smooth function when $t > 0$.

*Step 2.* We prove $P(v) \le 0$.

According to Proposition 9.1, in [P1] (Proposition 6.1.3 here), we know that

$$\Delta P(v) - RP(v) + \partial_t P(v) \geq 0. \tag{6.3.29}$$

Since $P(v)|_{t=1} = P(v_1) = \Lambda v_1^2 \leq 0$, we claim that

$$P(v) \leq 0 \tag{6.3.30}$$

for all $x \in \mathbf{M}$ and $t \in [0,1]$, by the maximum principle.

We mention that the behavior of $P(v)$ when $t$ is near 1 can be tricky, since $\Delta v_1$ is not continuous across the boundary. Therefore, a careful proof of the claim is necessary. In terms of the solution $u$ of the conjugate heat equation, we can write

$$P(v)(x,t) = (\sigma^2+1-t)(-2\Delta u+\frac{|\nabla u|^2}{u}+Ru)-u\ln u-\frac{n}{2}(\ln(\sigma^2+1-t))u. \tag{6.3.31}$$

We will show that $P(v)$ is a proper sub-solution to (6.3.29) such that:

$$\lim_{t\to 1}\int_{\mathbf{M}} P(v)(x,t)\theta(x)d\mu(g(t)) = \int_{\mathbf{M}} P(v_1)(x)\theta(x)d\mu(g(1)) \leq 0$$

for any smooth function $\theta \geq 0$. Then we can apply the maximum principle to prove the claim.

For a positive smooth function $\theta = \theta(x)$ and $t > 0$,

$$\begin{aligned}
&\int_{\mathbf{M}} P(v)(x,t)\theta(x)d\mu(g(t)) \\
&= \int_{\mathbf{M}} [(\sigma^2+1-t)(-2u\Delta\theta(x) + \frac{|\nabla u|^2}{u}\theta(x) + Ru\theta(x)) \\
&\qquad - \theta(x)u\ln u - \frac{n}{2}(\ln(\sigma^2+1-t))u\theta(x)]d\mu(g(t)).
\end{aligned}$$

Recall that the final value of $u$ is $u(x,1) = v_1^2(x)$, where $v_1$ solves (6.3.25). This function, defined on $\mathbf{M}$ after extension by zero value, is Hölder continuous and lies in $W^{1,\infty}(\mathbf{M})$. Therefore $u(x,t) \to v_1^2(x)$ when $t \to 1$. Hence

$$\begin{aligned}
&\lim_{t\to 1}\int_{\mathbf{M}} P(v)(x,t)\theta(x)d\mu(g(t)) \\
&= \int_{\mathbf{M}} [\sigma^2(-2v_1^2\Delta\theta(x) + Rv_1^2\theta(x)) - \theta(x)v_1^2\ln v_1^2 \\
&\quad - \frac{n}{2}(\ln\sigma^2)v_1^2\theta(x)]d\mu(g(1)) + \lim_{t\to 1}\int_{\mathbf{M}} \frac{|\nabla u|^2}{u}\theta(x)d\mu(g(t)).
\end{aligned} \tag{6.3.32}$$

Notice that $\lim_{t\to 1} \frac{|\nabla u|^2}{u}(x,t) = 4|\nabla v_1(x)|^2$ when $x$ is not on the boundary of the ball $B(x_0, A, 1)$. In order to take the limit inside the integral, we need further justification, which comes from the fact that $\frac{|\nabla u|^2}{u}(x,t)$ is bounded for $(x,t) \in \mathbf{M} \times [0,1)$. To prove this boundedness, we recall Proposition 6.1.2 states:

$$H^*\left(\frac{|\nabla u|^2}{u} + Ru\right) = \frac{2}{u}\left(u_{ij} - \frac{u_i u_j}{u}\right)^2 + 2\nabla R \nabla u + \frac{4}{u} Ric(\nabla u, \nabla u) \\ + 2|Ric|^2 u + 2\nabla R \nabla u + 2u\Delta R.$$

Hence

$$H^*\left(\frac{|\nabla u|^2}{u} + Ru\right) \geq -K_1\left(|\nabla u| + |u| + \frac{|\nabla u|^2}{u}\right),$$

where the constant $K_1 (\geq 0)$ depends on the supremum of $|\nabla R|$, $|\Delta R|$ and the lower bound of $Ric$. Since $|\nabla u| \leq \frac{|\nabla u|^2}{u} + u$, we deduce

$$H^*\left(\frac{|\nabla u|^2}{u} + Ru\right) \geq -K_1\left(\frac{|\nabla u|^2}{u} + Ru\right) - K_2, \tag{6.3.33}$$

where $K_2$ depends on $K_1$, the supremum of $|R|$ and $u$. We mention that all the curvatures involved here are bounded since the Ricci flow is smooth by assumption. The function $u$ is bounded by the maximum principle since $u(x,1) = v_1^2$ is bounded. Recall that, when $t = 1$,

$$\frac{|\nabla u|^2}{u} = 4|\nabla v_1|^2$$

which is bounded. Applying the maximum principle on (6.3.33), we find that $\frac{|\nabla u|^2}{u}$ is bounded for all $t \in [0,1]$. This process can be made rigorous by considering $\{u_k\}$ which is an approximation sequence of $u$, defined by

$$H^* u_k = 0, \qquad u_k(x,1) = v_1^2 + k^{-1}, \qquad k = 2, 3, \ldots.$$

Now that we know $\frac{|\nabla u|^2}{u}$ is bounded, we can take limit for (6.3.32) to deduce

$$\lim_{t\to 1} \int_{\mathbf{M}} P(v)(x,t)\theta(x) d\mu(g(t)) = \int_{\mathbf{M}} P(v_1)(x)\theta(x) d\mu(g(1)) < 0. \tag{6.3.34}$$

Note that $u$ is a bounded function. Also, by differentiating (6.3.27), it is easy to see that $\nabla u$ is bounded. By the format of $P(v)$ in (6.3.31), we know that $P(v) \in L^\infty([0,1], W^{-1,2}(\mathbf{M}))$. Applying an integral form the maximum principle to (6.3.29), we conclude that $P(v) \le 0$, i.e. (6.3.30) holds. Here is why. Pick $t_2, t_1 \in [0,1)$ such that $t_2 > t_1$. Let $\theta = \theta(x)$ be a positive, smooth function. Let $f = f(x,t)$ be the solution of the forward heat equation

$$\begin{cases} \Delta f - \partial_t f = 0, \quad t \in [t_1, t_2],\, x \in \mathbf{M} \\ f(x, t_1) = \theta(x),\, x \in \mathbf{M}. \end{cases}$$

Then $f$ is smooth and positive. Performing integration by parts, we obtain

$$\begin{aligned} \frac{d}{dt}\int_{\mathbf{M}} P(v) f d\mu(g(t)) &= \int_{\mathbf{M}} [(\partial_t P(v) - RP(v))f + P(v)\partial_t f] d\mu(g(t)) \\ &= \int_{\mathbf{M}} (\partial_t P(v) - RP(v) + \Delta P(v)) f d\mu(g(t)) \ge 0 \end{aligned}$$

Letting $t_2 \to 1$, we deduce, by using (6.3.34) (with $\theta(x)$ replaced by $f(x,1)$),

$$\int_{\mathbf{M}} P(v)\theta(x) d\mu(g(t_1)) \le \int_{\mathbf{M}} P(v) f(x,1) d\mu(g(1)) < 0$$

Since $\theta$ and $t_1$ are arbitrary, we know that $P(v) \le 0$ throughout.

*Step 3.* Proving a local monotonicity formula with suitable cut-off function.

Let $h$ be the solution to the heat equation with initial Dirichlet condition in the region

$$\{(x,t) \mid d(x_0, x, t) \le 1 + (2A-1)t,\ t \in [0,1]\}.$$

$$\begin{cases} \Delta h - \partial_t h = 0, \\ h = 0 \quad \text{on the sides}, \\ h(x,0) = h_0(x),\ d(x_0, x, 0) \le 1. \end{cases} \tag{6.3.35}$$

Here $h_0$ is a nonnegative function to be chosen later.

We write

$$J(t) = \int_{\mathbf{M}} P(v) h(x,t) d\mu(g(t)). \tag{6.3.36}$$

Then

$$J'(t) = \int_{\mathbf{M}} (\partial_t P(v) - RP(v)) h d\mu(g(t)) + \int_{\mathbf{M}} P(v)\Delta h d\mu(g(t)). \tag{6.3.37}$$

Note $P(v) \le 0$ and $\frac{\partial h}{\partial n} \le 0$ on $\partial B(x_0, 1+(2A-1)t, t)$ where $n$ is the exterior normal. We have

$$\begin{aligned}\int_{\mathbf{M}} P(v)\Delta h d\mu(g(t)) &= \int_{\partial B(x_0,1+(2A-1)t,t)} P(v)\frac{\partial h}{\partial n} dS \\ &- \int_{\partial B(x_0,1+(2A-1)t,t)} \frac{\partial P(v)}{\partial n} h dS + \int_{\mathbf{M}} h\Delta P(v) d\mu(g(t)) \\ &\ge \int_{\mathbf{M}} h\Delta P(v) d\mu(g(t)).\end{aligned}$$

Substituting this to (6.3.37) and using (6.3.29), we deduce

$$J'(t) \ge 0. \tag{6.3.38}$$

Next, we establish a lower bound for $h$. Let $\lambda = \lambda(\cdot)$ be a decreasing real valued function to be specified later. Define

$$\phi = \lambda[d(x_0, x, t) - (1 + (2A-1)t)]. \tag{6.3.39}$$

Then, it is clear that, in the distribution sense,

$$(\Delta - \partial_t)\phi = \lambda'(\Delta - \partial_t)d + (2A-1)\lambda' + \lambda''.$$

Using Lemma 8.3 [P1] (Proposition 5.1.5 here) for those $x$ such that $d(x_0, x, 0) \ge 1$ and the standard Laplacian comparison theorem for other $x$, we have

$$(\Delta - \partial_t)d(x_0, x, t) \le s_0$$

where $s_0$ is a positive number depending on $r_0$ (chosen as 1 here) and $n$ only. Recall that $\lambda$ is nonincreasing. Hence

$$(\Delta - \partial_t)\phi \ge \lambda' s_0 + (2A-1)\lambda' + \lambda''. \tag{6.3.40}$$

We will make the function $\lambda$ so that

$$\lambda(s) = 1,\ s \le -1/2; \lambda(s) = 0, s \ge 0; \quad \lambda''(s) \le 0,\ s \in [-1/2, -1/4];$$

$$\lambda''(s) \ge 0,\ s \in [-1/4, 0]; \quad \|\lambda'\|_\infty, \quad \|\lambda''\|_\infty \le 8.$$

For $s \in [-1/8, 0]$, we take

$$\lambda(s) = a_0[e^{-(s_0+2A-1)s} - 1]^2$$

where $a_0$ is chosen so that $\lambda(-1/8) < \lambda(-1/4)$ and that all the above properties for $\lambda$ hold.

By this choice of $\lambda$, it is clear that

$$(\Delta - \partial_t)\phi(x,t) \geq \begin{cases} 0, & d(x_0,x,t) - [1+(2A-1)t] \leq -1/2, \\ -c(1+s_0+A), & -1/2 < d(x_0,x,t) \\ & \quad -[1+(2A-1)t] \leq -1/8, \\ 0, & d(x_0,x,t) - [1+(2A-1)t] > -1/8. \end{cases}$$

Here $c$ is an absolute constant. Therefore, there exists a constant $Q = Q(s_0, A) > 0$ such that

$$(\Delta - \partial_t)\phi \geq -Q\phi.$$

This shows, via the maximum principle

$$h(x,t) \geq e^{-Qt}\phi(x,t)$$

provided that $h_0 = \phi(x,0)$.

Note that the above shows $h(x,1) \geq e^{-Q}$ when $d(x_0,x,1) \leq A$. Recall also that $P(v)(x,1) = \Lambda v_1^2$ while $v_1$ is supported in $B(x_0,A,1)$. From the definition in (6.3.24), we have, since $\Lambda < 0$,

$$J(1) = \Lambda \int v_1^2 h(x,1) d\mu(g(1)) \leq \Lambda e^{-Q}. \tag{6.3.41}$$

Therefore the monotonicity of $J(t)$ (6.3.38) shows

$$\Lambda \geq e^Q J(0). \tag{6.3.42}$$

Next we find a lower bound for $J(0)$. By definition

$$\begin{aligned} J(0) &= \int P(v)\phi(x,0)d\mu(g(0)) \\ &= \int [(\sigma^2+1)(-4\Delta v + Rv) - 2v\ln v - \frac{n}{2}(\ln(\sigma^2+1))v]v\phi(x,0)d\mu(g(0)). \end{aligned}$$

Using integration by parts, it is easy to see that

$$\begin{aligned} J(0) = \int \Big[ &(\sigma^2+1)(4|\nabla(v\sqrt{\phi})|^2 + R(v\sqrt{\phi})^2) - (v\sqrt{\phi})^2 \ln(v\sqrt{\phi})^2 \\ &- \frac{n}{2}(v\sqrt{\phi})^2 \ln(\sigma^2+1)\Big] d\mu(g(0)) \\ &- 4(\sigma^2+1)\int |\nabla\sqrt{\phi}|^2 v^2 d\mu(g(0)) + \int v^2(\sqrt{\phi})^2 \ln \sqrt{\phi}^2 d\mu(g(0)). \end{aligned}$$

It is clear that we can choose $\lambda$ in the definition of $\phi$ so that

$$|\nabla \sqrt{\phi(x,0)}|^2 \le C, \qquad \sqrt{\phi} \ln \sqrt{\phi}(x,0) \ge -e.$$

Here $C$ is an absolute positive constant. Since the $L^2$ norm of $v_1$ is 1 and that $u = v^2$ is the solution of the conjugate heat equation with final value $v_1^2$, we know that the $L^2$ norm of $v = v(\cdot, t)$ is always 1. Hence

$$\begin{aligned} J(0) \ge \int & \left[(\sigma^2+1)(4|\nabla(v\sqrt{\phi})|^2 + R(v\sqrt{\phi})^2) - (v\sqrt{\phi})^2 \ln(v\sqrt{\phi})^2 \right. \\ & \left. - \frac{n}{2}(v\sqrt{\phi})^2 \ln(\sigma^2+1)\right] d\mu(g(0)) - C. \end{aligned}$$

Write $w = v\sqrt{\phi(x,0)}/\|v\sqrt{\phi(\cdot,0)}\|_2$. The $L^2$ norm of $w$ is 1 and

$$\begin{aligned} J(0) \ge \|v\sqrt{\phi(\cdot,0)}\|_2^2 \int & [(\sigma^2+1)(4|\nabla w|^2 + Rw^2) \\ & - w^2 \ln w^2 - \frac{n}{2} w^2 \ln(\sigma^2+1)] d\mu(g(0)) \\ & - \|v\sqrt{\phi(\cdot,0)}\|_2^2 \ln \|v\sqrt{\phi(\cdot,0)}\|_2^2 - C. \end{aligned}$$

Minimizing the right-hand side, we deduce

$$J(0) \ge \|v\sqrt{\phi(\cdot,0)}\|_2^2 \Lambda_{\sigma^2+1}(g(0)) - C.$$

Therefore

$$\Lambda = \Lambda_{\sigma^2}(g(1)) \ge e^Q[\|v\sqrt{\phi(\cdot,0)}\|_2^2 \Lambda_{\sigma^2+1}(B(x_0,1,0), g(0)) - C]. \tag{6.3.43}$$

Here

$$\begin{aligned} \Lambda_{\sigma^2+1}(B(x_0,1,0), g(0)) = \inf \Bigg\{ & \int_{B(x_0,1,0)} [(\sigma^2+1)(4|\nabla v|^2 + Rv^2) \\ & - v^2 \ln v^2]\ d\mu(g(0)) - \frac{n}{2}\ln(\sigma^2+1) \\ & \mid v \in C_0^\infty(B(x_0,1,0)),\ \|v\|_2 = 1 \Bigg\}. \end{aligned}$$

*Step 4*. Completion of the proof.

With the curvature and volume assumption of the theorem at time $t = 0$, a Sobolev inequality holds for functions in $W_0^{1,2}(B(x_0,1,0))$, i.e. for all $v \in W_0^{1,2}(B(x_0, 1, 0))$, there exist constants $S_i = S_i(A, n)$, $i = 1, 2$, such that

$$\left( \int v^{2n/(n-2)} d\mu(g(0)) \right)^{(n-2)/n} \leq S_1 \int (4|\nabla v|^2 + R v^2) d\mu(g(0)) + S_2 \int v^2 d\mu(g(0)).$$

This is a version of the Sobolev inequality in [Au], which is fitted for bounded domains. The proof of this is similar to the full manifold version, as given in Theorem 4.1.1. We leave it as an exercise.

By Theorem 4.2.1, a log Sobolev inequality in the form of (II) in that theorem and perturbed by the scalar curvature term holds for functions in $W_0^{1,2}(B(x_0, 1, 0))$. The constants depend only on dimension. This means that

$$\Lambda_{\sigma^2+1}(B(x_0, 1, 0), g(0)) \geq -c_1(\sigma + 1)^2 - c_2$$

where $c_1$ and $c_2$ are constants depending on $n$ and $A$. Using (6.3.43), we know that, for all $\sigma > 0$,

$$\Lambda_{\sigma^2}(g(1)) \geq -c_3\sigma^2 - c_4$$

where $c_3$ and $c_4$ are positive constants depending only on $n$ and $A$. Using Theorem 4.2.1, we can proceed as in the end of the proof of Theorem 6.2.1 to show that this family of log Sobolev inequalities for $W_0^{1,2}(B(x_0, A, 1))$ induces the desired Sobolev inequality. □

## 6.4 A differential Harnack inequality for the conjugate heat equation

In this section we prove a differential Harnack inequality for all positive solutions of the conjugate heat equation. This can be regarded as a generalization of the main results in [AB] and [LY] to the Ricci flow setting. Perelman [P1] (Proposition 6.1.3 here) actually proved a differential Harnack inequality for the fundamental solution of the conjugate heat equation. However, there is one place where some improvement is still desirable, namely Perelman's differential Harnack inequality does not apply to all positive solutions. For instance, for the Ricci flat manifold $S^1 \times S^1$. The constant 1, as a solution to the conjugate heat equation is a counterexample. Here we present a result in [KZ] which holds for all positive solutions. It remains to be seen whether a sharper gradient estimate exists. Recently similar results have appeared in [CaH] and [Cx2].

The main result of this section is

**Theorem 6.4.1** *Let $(M, g(t))$ be a smooth Ricci flow, where $M$ is a closed manifold and $t \in [0, T)$. Let $u : M \times [0, T) \mapsto (0, \infty)$ be a positive $C^{2,1}$ solution to the conjugate heat equation $H^* u = \triangle u + u_t - Ru = 0$. Let $u = \frac{e^{-f}}{(4\pi\tau)^{\frac{n}{2}}}$ and $\tau = T - t$. Then the following inequalities are true:*
*(i) if the scalar curvature $R \geq 0$, then for all $t \in (0, T)$ and all points,*

$$2\triangle f - |\nabla f|^2 + R \leq \frac{2n}{\tau}; \tag{6.4.1}$$

*(ii) without assuming the non-negativity of $R$, then for $t \in [\frac{T}{2}, T)$ and all points,*

$$2\triangle f - |\nabla f|^2 + R \leq \frac{3n}{\tau}. \tag{6.4.2}$$

**Remark 6.4.1** *Since $f = -\ln u - \frac{n}{2}\ln(4\pi\tau)$, if we replace $f$ by $u$ in the above inequalities, then we get*

$$\begin{aligned} &\frac{|\nabla u|^2}{u^2} - 2\frac{u_\tau}{u} - R \leq \frac{2n}{\tau}, \quad \textit{if} \quad R \geq 0; \\ &\frac{|\nabla u|^2}{u^2} - 2\frac{u_\tau}{u} - R \leq \frac{3n}{\tau}, \quad \textit{if } R \textit{ changes sign and} \quad t \geq T/2. \end{aligned} \tag{6.4.3}$$

*It is similar to the Li-Yau gradient estimate for the heat equation on manifolds with nonnegative Ricci curvature, i.e.*

$$\frac{|\nabla u|^2}{u^2} - \frac{u_t}{u} \leq \frac{n}{2t}$$

*for positive solutions of $\Delta u - \partial_t u = 0$.*

**Remark 6.4.2** *Some related gradient estimates with various dependence on the Ricci and other curvatures can be found in [Gro] and [Ni].*

Proof. (of Theorem 6.4.1). By a standard approximation argument as in [Cetc] Vol. 2 e.g., we can assume without loss of generality that $g = g(t)$ is smooth in the closed time interval $[0, T]$ and that $u$ is strictly positive everywhere. Indeed, by Theorem A.23 in [Cetc] Vol. 2 (due to W.X. Shi), the curvature tensor is uniformly bounded in the time interval $[0, T - \delta]$ with the bound depending only on the initial data and $\delta$, a positive number. Moreover the lower bound of the scalar curvature is nondecreasing since the scalar curvature $R$ satisfies (c.f. p209 [CK])

$$\Delta R - \partial_t R + \frac{2}{n} R^2 \leq 0.$$

Therefore, we can just work on the interval $[0, T - \delta]$ first. In the proof, it will be clear that all constants are independent of the curvature tensor. They only depend on the lower bound of the scalar curvature, which is nondecreasing with time. Hence we can take $\delta$ to zero to get the desired result on $[0, T)$.

(i) By standard computation (one can consult various sources for more details ([CK] or Proposition 5.1.1 here e.g.)),

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \triangle\right)(\triangle f) &= \triangle \frac{\partial f}{\partial t} + 2\langle Ric, Hess(f)\rangle + \triangle(\triangle f)\\ &= \triangle\left(-\triangle f + |\nabla f|^2 - R + \frac{n}{2\tau}\right) + 2\langle Ric, Hess(f)\rangle\\ &\quad + \triangle(\triangle f)\\ &= 2\langle Ric, Hess(f)\rangle + \triangle\left(|\nabla f|^2 - R\right).\end{aligned}$$

Also using the evolution equation of $g$,

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \triangle\right)|\nabla f|^2 &= 2Ric(\nabla f, \nabla f) + 2\langle \nabla f, \nabla \frac{\partial f}{\partial t}\rangle + \triangle|\nabla f|^2\\ &= 2Ric(\nabla f, \nabla f) + 2\langle \nabla f, \nabla(-\triangle f + |\nabla f|^2 - R)\rangle\\ &\quad + \triangle|\nabla f|^2.\end{aligned}$$

Notice also

$$\left(\frac{\partial}{\partial t} + \triangle\right) R = 2\triangle R + 2|Ric|^2. \tag{6.4.4}$$

Combining these three expressions, we deduce

$$\begin{aligned}&\left(\frac{\partial}{\partial t} + \triangle\right)(2\triangle f - |\nabla f|^2 + R)\\ &= 4\langle Ric, Hess(f)\rangle + \triangle|\nabla f|^2 - 2Ric(\nabla f, \nabla f)\\ &\quad - 2\langle \nabla f, \nabla(-\triangle f + |\nabla f|^2 - R)\rangle + 2|Ric|^2.\end{aligned} \tag{6.4.5}$$

Denote

$$q(x,t) = 2\triangle f - |\nabla f|^2 + R.$$

By Bochner's identity, writing $Hess\, f = f_{ij}$ in a local coordinates,

$$\triangle|\nabla f|^2 = 2|f_{ij}|^2 + 2\nabla f \nabla(\triangle f) + 2R_{ij} f_i f_j,$$

the above equation becomes

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\triangle\right)q &= 4R_{ij}f_{ij}+\big(2|f_{ij}|^2+2\nabla f\,\nabla(\triangle f)+2R_{ij}f_if_j\big)-2R_{ij}f_if_j\\ &\quad -2\nabla f\,\nabla(-\triangle f+|\nabla f|^2-R)+2R_{ij}^2\\ &= 4R_{ij}f_{ij}+2|f_{ij}|^2+2R_{ij}^2+2\nabla f\,\nabla\big(2\triangle f-|\nabla f|^2+R\big)\\ &= 2|R_{ij}+f_{ij}|^2+2\nabla f\,\nabla q\end{aligned}$$

that is,

$$\left(\frac{\partial}{\partial t}+\triangle\right)q-2\nabla f\,\nabla q=2|R_{ij}+f_{ij}|^2\geq\frac{2}{n}(R+\triangle f)^2\,. \tag{6.4.6}$$

(We note that equation (6.4.6) was also shown in [Cetc].)

Since

$$q=2\triangle f-|\nabla f|^2+R=2(\triangle f+R)-|\nabla f|^2-R,$$

and hence

$$R+\triangle f=\frac{1}{2}(q+|\nabla f|^2+R),$$

we have

$$\left(\frac{\partial}{\partial t}+\triangle\right)q-2\nabla f\,\nabla q\geq\frac{1}{2n}\big(q+|\nabla f|^2+R\big)^2\,. \tag{6.4.7}$$

By direct computation, we also have, for any $\epsilon>0$

$$\left(\frac{\partial}{\partial t}+\triangle\right)\frac{2n}{T-t+\epsilon}-2\nabla f\,\nabla\Big(\frac{2n}{T-t+\epsilon}\Big)=\frac{1}{2n}\Big(\frac{2n}{T-t+\epsilon}\Big)^2.$$

Combining the above two expressions, we get

$$\begin{aligned}&\left(\frac{\partial}{\partial t}+\triangle\right)\Big(q-\frac{2n}{T-t+\epsilon}\Big)-2\nabla f\,\nabla\Big(q-\frac{2n}{T-t+\epsilon}\Big)\\ &\geq\frac{1}{2n}\left(q+\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R\right)\left(q-\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R\right).\end{aligned} \tag{6.4.8}$$

We deal with the above inequality in two cases:

**Case 1.** At a point $(x,t)$, $q+\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R\leq 0$, then also

$$q-\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R\leq 0$$

thus,

$$\left(\frac{\partial}{\partial t}+\triangle\right)\big(q-\frac{2n}{T-t+\epsilon}\big)-2\nabla f\,\nabla\big(q-\frac{2n}{T-t+\epsilon}\big)\geq 0.$$

**Case 2.** At a point $(x,t)$, $q+\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R>0$, then the inequality (6.4.8) can be transformed to

$$\begin{aligned}&\left(\frac{\partial}{\partial t}+\triangle\right)\big(q-\frac{2n}{T-t+\epsilon}\big)-2\nabla f\,\nabla\big(q-\frac{2n}{T-t+\epsilon}\big)\\ &-\frac{1}{2n}\left(q+\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R\right)\left(q-\frac{2n}{T-t+\epsilon}\right)\\ &\geq\frac{1}{2n}(|\nabla f|^2+R)\left(q+\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R\right)\geq 0\,.\end{aligned}$$

Defining a potential term by

$$V=V(x,t)=\begin{cases}0;\\ \quad\text{if } q+\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R\leq 0 \text{ at } (x,t)\\ \frac{1}{2n}\big(q+\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R\big);\\ \quad\text{if } q+\frac{2n}{T-t+\epsilon}+|\nabla f|^2+R\geq 0 \text{ at } (x,t)\,.\end{cases}\qquad(6.4.9)$$

We know $V$ is continuous; further, by the above two cases, we conclude

$$\begin{aligned}&\big(\frac{\partial}{\partial t}+\triangle\big)\big(q-\frac{2n}{T-t+\epsilon}\big)-2\nabla f\nabla\big(q-\frac{2n}{T-t+\epsilon}\big)\\ &\quad-V\big(q-\frac{2n}{T-t+\epsilon}\big)\geq 0.\end{aligned}$$

Since we assumed that the Ricci flow is smooth in $[0,T]$ and that $u(x,t)$ is a positive $C^{2,1}$ solution to the conjugate heat equation, thus

$$q=2\triangle f-|\nabla f|^2+R=\frac{|\nabla u|^2}{u^2}-\frac{2\triangle u}{u}+R$$

is bounded for $t\in[0,T]$. If we choose $\epsilon$ sufficiently small, then $q(x,T)\leq\frac{2n}{\epsilon}$, thus by the maximum principle ( [CK], e.g.), for all $t\in[0,T]$, $q(x,t)\leq\frac{2n}{T-t+\epsilon}$. Let $\epsilon\to 0$, we have for all $t\in[0,T]$,

$$q(x,t)\leq\frac{2n}{T-t}\,.$$

Recall $q = 2\triangle f - |\nabla f|^2 + R$, $\tau = T - t$, then we have

$$2\triangle f - |\nabla f|^2 + R \leq \frac{2n}{\tau}\,. \tag{6.4.10}$$

Further, $f = -\ln u - \frac{n}{2}(4\pi\tau)$, then the above yields

$$\frac{|\nabla u|^2}{u^2} - \frac{2u_\tau}{u} - R \leq \frac{2n}{\tau}. \tag{6.4.11}$$

**Proof of (ii).** Next we prove the gradient estimate without the non-negativity assumption for the scalar curvature $R$. Let $c \geq 2n$ be a constant to be determined later; denote

$$B = |\nabla f|^2 + R\,.$$

Similar to the inequality (6.4.8), we also have,

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} + \triangle\right)(q - \frac{c}{T-t+\epsilon}) - 2\nabla f\,\nabla(q - \frac{c}{T-t+\epsilon})\\
&\geq \frac{1}{2n}(q+B)^2 - \frac{c}{(T-t+\epsilon)^2}\\
&= \frac{1}{2n}[(q+B)^2 - \frac{c^2}{(T-t+\epsilon)^2} + \frac{c^2}{(T-t+\epsilon)^2} - \frac{2cn}{(T-t+\epsilon)^2}]\\
&= \frac{1}{2n}[(q - \frac{c}{T-t+\epsilon} + B)(q + \frac{c}{T-t+\epsilon} + B) + \frac{c(c-2n)}{(T-t+\epsilon)^2}].
\end{aligned} \tag{6.4.12}$$

We deal with the previous inequality at a given point $(x,t)$ in three cases:

**Case 1.** $B \geq 0$, and $q + \frac{c}{T-t+\epsilon} + B \leq 0$, then also

$$q - \frac{c}{T-t+\epsilon} + B \leq 0$$

thus,

$$\left(\frac{\partial}{\partial t} + \triangle\right)(q - \frac{c}{T-t+\epsilon}) - 2\nabla f\,\nabla(q - \frac{c}{T-t+\epsilon}) \geq 0.$$

**Case 2.** $B \geq 0$, and $q + \frac{c}{T-t+\epsilon} + B > 0$, then the inequality (6.4.12) can be changed to

$$\begin{aligned}&\left(\frac{\partial}{\partial t}+\triangle\right)(q-\frac{c}{T-t+\epsilon})-2\nabla f\,\nabla(q-\frac{c}{T-t+\epsilon})\\&-\frac{1}{2n}\left(q+\frac{c}{T-t+\epsilon}+B\right)\left(q-\frac{c}{T-t+\epsilon}\right)\\&\geq\frac{1}{2n}B\left(q+\frac{c}{T-t+\epsilon}+B\right)\geq 0.\end{aligned}$$

**Case 3.** $B \leq 0$, then the inequality (6.4.12) can be changed to

$$\begin{aligned}&\left(\frac{\partial}{\partial t}+\triangle\right)(q-\frac{c}{T-t+\epsilon})-2\nabla f\,\nabla(q-\frac{c}{T-t+\epsilon})\\&\geq\frac{1}{2n}\left(q+\frac{c}{T-t+\epsilon}+B\right)\left(q-\frac{c}{T-t+\epsilon}\right)\\&\quad+\frac{1}{2n}B\left(q-\frac{c}{T-t+\epsilon}\right)+\frac{1}{2n}\left(\frac{2Bc}{T-t+\epsilon}+\frac{c(c-2n)}{(T-t+\epsilon)^2}\right).\end{aligned}$$

To continue, we need the following estimate of the scalar curvature $R$ under the Ricci flow i.e.

$$R \geq -\frac{n}{2(t+\epsilon)} \tag{6.4.13}$$

for some $\epsilon > 0$ depending on the initial value of $R$. This lower bound is a result of the weak maximum principle applied on differential inequality $\frac{\partial R}{\partial t} \geq \triangle R + \frac{2}{n}R^2$ (c.f. [CK], e.g.). Thus

$$B = |\nabla f|^2 + R \geq R \geq -\frac{n}{2(t+\epsilon)} \geq -\frac{n}{2(T-t+\epsilon)}$$

for $t \geq \frac{T}{2}$ because $t \geq \frac{T}{2} \Rightarrow t \geq T-t \Rightarrow t+\epsilon \geq T-t+\epsilon \Rightarrow \frac{1}{t+\epsilon} \leq \frac{1}{T-t+\epsilon}$. Consequently

$$\begin{aligned}&\frac{1}{2n}\left(\frac{2Bc}{T-t+\epsilon}+\frac{c(c-2n)}{(T-t+\epsilon)^2}\right)\\&\geq\frac{1}{2n}\left(-\frac{n}{2(T-t+\epsilon)}\,\frac{2c}{T-t+\epsilon}+\frac{c(c-2n)}{(T-t+\epsilon)^2}\right)\\&=\frac{1}{2n}\left(\frac{c(c-3n)}{(T-t+\epsilon)^2}\right).\end{aligned}$$

Therefore

$$\begin{aligned}&\left(\frac{\partial}{\partial t}+\triangle\right)\left(q-\frac{c}{T-t+\epsilon}\right)-2\nabla f\,\nabla\left(q-\frac{c}{T-t+\epsilon}\right)\\&-\frac{1}{2n}\left(q+\frac{c}{T-t+\epsilon}+2B\right)\left(q-\frac{c}{T-t+\epsilon}\right)\\&\geq\frac{c\,(c-3n)}{2n(T-t+\epsilon)^2}.\end{aligned}$$

Taking $c=3n$, we have,

$$\begin{aligned}&\left(\frac{\partial}{\partial t}+\triangle\right)\left(q-\frac{2n}{T-t+\epsilon}\right)-2\nabla f\nabla\left(q-\frac{2n}{T-t+\epsilon}\right)\\&\qquad-V\left(q-\frac{2n}{T-t+\epsilon}\right)\geq 0\end{aligned}$$

where $V=V(x,t)$ is a bounded function defined by

$$V=\begin{cases}0; & \text{if } B\geq 0,\ q+\frac{2n}{T-t+\epsilon}+B\leq 0 \text{ at } (x,t)\\ \frac{1}{2n}\left(q+\frac{2n}{T-t+\epsilon}+B\right); & \text{if } B\geq 0,\ q+\frac{2n}{T-t+\epsilon}+B>0 \text{ at } (x,t)\\ \frac{1}{2n}\left(q+\frac{2n}{T-t+\epsilon}+2B\right); & \text{if } B<0 \text{ at } (x,t)\,.\end{cases} \tag{6.4.14}$$

Follow the similar argument for the inequality (6.4.1), by the maximum principle again, we have, after letting $\epsilon\to 0$,

$$2\triangle f-|\nabla f|^2+R\leq\frac{3n}{\tau}\quad\textit{and}\quad\frac{|\nabla u|^2}{u^2}-\frac{2u_\tau}{u}-R\leq\frac{3n}{\tau},\quad t\geq T/2. \tag{6.4.15}$$

$\square$

An immediate consequence of the above theorem is:

**Corollary 6.4.1 (Differential Harnack Inequality)** *Given a smooth Ricci flow on a closed manifold $M$, let $u: M\times[0,T)\mapsto(0,\infty)$ be a positive $C^{2,1}$ solution to the conjugate heat equation.*

*(a). Suppose the scalar curvature $R\geq 0$ for $t\in[0,T)$. Then for any two points $(x,t_1)$, $(y,t_2)$ in $\mathbf{M}\times(0,T)$ such that $t_1<t_2$, it holds*

$$u(y,t_2)\leq u(x,t_1)\left(\frac{\tau_1}{\tau_2}\right)^n\exp\frac{\int_0^1[\,4|\gamma'(s)|^2+(\tau_1-\tau_2)^2\,R\,]\,ds}{2(\tau_1-\tau_2)}.$$

*Here $\tau_i=T-t_i$, $i=1,2$, and $\gamma(s):[0,1]\to M$ is a smooth curve from $x$ to $y$.*

*(b). Without assuming the nonnegativity of the scalar curvature $R$, then for $t_2 > t_1 \geq T/2$, it holds*

$$u(y,t_2) \leq u(x,t_1)\left(\frac{\tau_1}{\tau_2}\right)^{3n/2} \exp\frac{\int_0^1 [\,4|\gamma'(s)|^2 + (\tau_1-\tau_2)^2 R\,]\,ds}{2(\tau_1-\tau_2)}.$$

*Here $R = R(\gamma(s), T-\tau)$ with $\tau = \tau_2 + (1-s)(\tau_1-\tau_2)$ and $|\gamma'(s)|^2 = g_{(T-\tau)}(\gamma'(s),\gamma'(s))$.*

Proof. We will only prove (a) since the proof of (b) is similar.

Denote $\tau(s) := \tau_2 + (1-s)(\tau_1-\tau_2)$, $0 \leq \tau_2 < \tau_1 \leq T$, define

$$\ell(s) := \ln\, u(\gamma(s), T-\tau(s))$$

where $\ell(0) = \ln\, u(x,t_1)$, $\ell(1) = \ln\, u(y,t_2)$.
By direct computation,

$$\begin{aligned}
\frac{\partial \ell(s)}{\partial s} = \frac{du/ds}{u} &= \frac{\nabla u}{u}\frac{\partial\gamma}{\partial s} - \frac{u_\tau(\tau_1-\tau_2)}{u}\\
&= (\tau_1-\tau_2)\left(\frac{\nabla u}{\sqrt{2}u}\cdot\frac{\sqrt{2}\,\gamma'(s)}{\tau_1-\tau_2} - \frac{u_\tau}{u}\right)\\
&\leq (\tau_1-\tau_2)\left(\frac{2|\gamma'(s)|^2}{(\tau_1-\tau_2)^2} + \frac{|\nabla u|^2}{2u^2} - \frac{u_\tau}{u}\right)\\
&= \frac{2|\gamma'(s)|^2}{(\tau_1-\tau_2)} + \frac{\tau_1-\tau_2}{2}\left(\frac{|\nabla u|^2}{u^2} - \frac{2u_\tau}{u}\right).
\end{aligned}$$

By our gradient estimate, if $R \geq 0$, then

$$\frac{|\nabla u|^2}{u^2} - \frac{2u_\tau}{u} \leq R + \frac{2n}{\tau} \tag{6.4.16}$$

where $\tau = \tau_2 + (1-s)(\tau_1-\tau_2)$. Therefore

$$\frac{\partial \ell(s)}{\partial s} \leq \frac{2|\gamma'(s)|^2}{(\tau_1-\tau_2)} + \frac{\tau_1-\tau_2}{2}\left(R + \frac{2n}{\tau}\right).$$

Integrating with respect to $s$ on $[0,1]$, we have

$$\ell(1) - \ell(0) \leq \frac{2\int_0^1 |\gamma'(s)|^2\,ds}{(\tau_1-\tau_2)} + \frac{(\tau_1-\tau_2)\int_0^1 R\,ds}{2} + n\ln\frac{\tau_1}{\tau_2}.$$

Recall $\ell(0) = \ln\, u(x,t_1)$, $\ell(1) = \ln\, u(y,t_2)$, then

$$\ln\frac{u(y,t_2)}{u(x,t_1)} \leq \frac{\int_0^1 [\,4|\gamma'(s)|^2 + (\tau_1-\tau_2)^2 R\,]\,ds}{2(\tau_1-\tau_2)} + \ln\left(\frac{\tau_1}{\tau_2}\right)^n.$$

Therefore, given any two points $(x, t_1)$, $(y, t_2)$ in the space-time, we have

$$u(y, t_2) \le u(x, t_1) \left(\frac{\tau_1}{\tau_2}\right)^n \exp \frac{\int_0^1 [\,4|\gamma'(s)|^2 + (\tau_1 - \tau_2)^2 \, R\,] \, ds}{2(\tau_1 - \tau_2)} .$$

□

## 6.5 Pointwise bound on the fundamental solutions of heat type equations

In this section, we specialize to the case of nonnegative Ricci curvature. We establish a certain Gaussian type upper bound for the fundamental solution of the conjugate heat equation. We will begin with the traditional method of establishing a mean value inequality via Moser's iteration and a weighted estimate in the spirit of Davies [Da]. However, there is some difficulty in applying this method directly due to the lack of control of the time derivative of the distance function. The new idea to overcome this difficulty is to use the interpolation result of Theorem 6.5.1.

Let us recall and define some notations to be used in this section. We will use $B(x, r; t)$ to denote the geodesic ball centered at $x$ with radius $r$ under the metric $g(t)$; $|B(x, r; t)|_s$ to denote the volume of $B(x, r; t)$ under the metric $g(s)$. In this section we will use $dg(x, t)$ to denote the volume element of the metric $g(t)$ at the point $x$.

The main result of this section is Theorem 6.5.2 below, which was first proven in [Z1]. Note that the theorem is qualitatively sharp in general since it matched the well-known Gaussian upper bound for the fixed metric case [LY]. Also there is no assumption on the comparability of metrics at different times. In this theorem, we assume the manifold is compact. This accounts for the extra 1 on the Gaussian upper bound. Even in the case of fixed metric, the heat kernel converges to a positive constant for large time. The theorem still holds for certain noncompact manifolds under suitable assumptions. In this case the extra 1 in the upper bound should be replaced by 0.

**Remark 6.5.1** *In the case of Ricci $\ge -k$ with $k > 0$, then certain integral Gaussian bound can still be proven by the same method. However, so far we are not able to derive a pointwise Gaussian upper bound without an exponentially growing term $e^{kt}$. This is due to a lack of an* *efficient mean value inequality for the second entries of the fundamental solution, which satisfies the forward heat equation after a time reversal.*

First we state and prove Theorem 3.3 in [Z1], which will also be used at the end of Step 2 in the proof of Theorem 7.2.1. See also [CaH].

**Theorem 6.5.1** *Let* $\mathbf{M}$ *be a compact or complete noncompact Riemann manifold with bounded curvature and equipped with a family of Riemann metric evolving under the forward Ricci flow* $\partial_t g = -2Ric$ *with* $t \in [0, T]$. *Suppose* $u$ *is any positive solution to* $\Delta u - \partial_t u = 0$ *in* $\mathbf{M} \times [0, T]$. *Then, it holds*

$$\frac{|\nabla u(x,t)|}{u(x,t)} \le \sqrt{\frac{1}{t}} \sqrt{\ln \frac{M}{u(x,t)}}$$

*for* $M = \sup_{\mathbf{M}\times[0,T]} u$ *and* $(x, t) \in \mathbf{M} \times [0, T]$.

*Moreover, the following interpolation inequality holds for any* $\delta > 0$, $x, y \in \mathbf{M}$ *and* $0 < t \le T$*:*

$$u(y,t) \le c_1 u(x,t)^{1/(1+\delta)} M^{\delta/(1+\delta)} e^{c_2 d(x,y,t)^2/t}.$$

*Here* $c_1, c_2$ *are positive constants depending only on* $\delta$.

PROOF. This is almost the same as that of Theorem 1.1 in [Ha5], except for a cancelation effect induced by the Ricci flow. By direct calculation

$$\Delta(u \ln \frac{M}{u}) - \partial_t(u \ln \frac{M}{u}) = -\frac{|\nabla u|^2}{u}, \tag{6.5.1}$$

$$(\Delta - \partial_t)(\frac{|\nabla u|^2}{u}) = \frac{2}{u}\left|\partial_i\partial_j u - \frac{\partial_i u \partial_j u}{u}\right|^2 \ge 0. \tag{6.5.2}$$

The first inequality in the statement of the theorem follows immediately from the maximum principle since

$$t\frac{|\nabla u|^2}{u} - u \ln \frac{M}{u}$$

is a subsolution of the heat equation.

To prove the second inequality, we set

$$l(x,t) = \ln(M/u(x,t)).$$

Then the first inequality implies

$$|\nabla \sqrt{l(x,t)}| \le 1/\sqrt{t}.$$

Fixing two points $x$ and $y$, we can integrate along a geodesic to reach

$$\sqrt{\ln(M/u(x,t))} \le \sqrt{\ln(M/u(y,t))} + \frac{d(x,y,t)}{\sqrt{t}}.$$

The result follows by squaring both sides. □

**Exercise 6.5.1** *Prove (6.5.2).*

**Theorem 6.5.2** *Assume that the conjugate heat equation (after reversal of time)*

$$\begin{cases} \Delta u - Ru - \partial_t u = 0, \\ \partial_t g = 2Ric \end{cases} \tag{6.5.3}$$

*has a smooth solution in the time interval* $[0,T]$ *and let* $G$ *be the fundamental solution. Suppose that Ricci* $\ge 0$ *and that the injectivity radius is bounded from below by a positive constant* $i$ *throughout. Then the following statement holds.*

*For any* $s, t \in (0,T)$, $s < t$, *and* $x, y \in \mathbf{M}$, *there exist a dimensional constant* $c_n$, *a dimensionless constant* $c$ *and a constant* $A$ *depending only on* $i$ *such that*

$$G(x,t;y,s) \le c_n A \left( 1 + \frac{1}{(t-s)^{n/2}} + \frac{1}{|B(x,\sqrt{t-s},t)|_s} \right) e^{-cd(x,y,s)^2/(t-s)}.$$

**Remark 6.5.2** *By Perelman's local noncollapsing theorem, the condition on the injectivity lower bound can be replaced by scalar curvature upper bound.*

Proof. We divide the proof into two steps.

*Step 1.* Proving the on diagonal bound, i.e. the one without the exponential term.

First we use Moser's iteration to prove a mean value inequality. The only new factor is a cancelation effect induced by the backward Ricci flow. So we will be brief in the presentation at this part of the proof.

Let $u$ be a positive solution to (6.5.3) in the region

$$Q_{\sigma r}(x,t) \equiv \{(y,s) \mid z \in \mathbf{M}, t-(\sigma r)^2 \le s \le t,\ d(y,x,s) \le \sigma r\}.$$

Here $r > 0, 2 \ge \sigma \ge 1$. Given any $p \ge 1$, it is clear that

$$\Delta u^p - pRu^p - \partial_t u^p \ge 0. \tag{6.5.4}$$

Let $\phi : [0,\infty) \to [0,1]$ be a smooth function such that $|\phi'| \le 2/((\sigma-1)r)$, $\phi' \le 0$, $\phi \ge 0$, $\phi(\rho) = 1$ when $0 \le \rho \le r$, $\phi(\rho) = 0$ when $\rho \ge \sigma r$. Let $\eta : [0,\infty) \to [0,1]$ be a smooth function such that $|\eta'| \le 2/((\sigma-1)r)^2$, $\eta' \ge 0$, $\eta \ge 0$, $\eta(s) = 1$ when $t - r^2 \le s \le t$, $\eta(s) = 0$ when $s \le t - (\sigma r)^2$. Define a cut-off function

$$\psi = \psi(y,s) = \phi(d(x,y,s))\eta(s).$$

Writing $w = u^p$ and using $w\psi^2$ as a test function on (6.5.4), we deduce

$$\begin{aligned}\int \nabla(w\psi^2)\nabla w dg(y,s)ds + p\int Rw^2\psi^2 dg(y,s)ds \\ \le -\int(\partial_s w)w\psi^2 dg(y,s)ds.\end{aligned} \tag{6.5.5}$$

By direct calculation

$$\int \nabla(w\psi^2)\nabla w dg(y,s)ds = \int |\nabla(w\psi)|^2 dg(y,s)ds - \int |\nabla\psi|^2 w^2 dg(y,s)ds. \tag{6.5.6}$$

Next we estimate the right-hand side of (6.5.5). Here we will use the backward Ricci flow.

$$\begin{aligned}-\int(\partial_s w)w\psi^2 dg(y,s)ds &= \int w^2\psi\partial_s\psi dg(y,s)ds \\ &+ \frac{1}{2}\int(w\psi)^2 R dg(y,s)ds - \frac{1}{2}\int(w\psi)^2 dg(y,t).\end{aligned}$$

Observe that

$$\partial_s\psi = \eta(s)\phi'(d(y,x,s))\partial_s d(y,x,s) + \phi(d(y,x,s))\eta'(s) \le \phi(d(y,x,s))\eta'(s).$$

This is so because $\phi' \le 0$ and $\partial_s d(y,x,s) \ge 0$ under the backward Ricci flow with nonnegative Ricci curvature (see Proposition 5.1.1 (3)). Hence

$$\begin{aligned}-\int(\partial_s w)w\psi^2 dg(y,s)ds &\le \int w^2\psi\phi(d(y,x,s))\eta'(s)dg(y,s)ds \\ &+ \frac{1}{2}\int(w\psi)^2 R dg(y,s)ds - \frac{1}{2}\int(w\psi)^2 dg(y,t).\end{aligned} \tag{6.5.7}$$

Combining (6.5.5) to (6.5.7), we obtain, in view of $p \ge 1$ and $R \ge 0$,

$$\begin{aligned}\int |\nabla(w\psi)|^2 dg(y,s)ds + \frac{1}{2}\int(w\psi)^2 dg(y,t) \\ \le \frac{c}{(\sigma-1)^2 r^2}\int_{Q_{\sigma r(x,t)}} w^2 dg(y,s)ds.\end{aligned} \tag{6.5.8}$$

By Hölder's inequality

$$\int (\psi w)^{2(1+(2/n)} dg(y,s) \le \Big( \int (\psi w)^{2n/(n-2)} dg(y,s) \Big)^{(n-2)/n} \times \Big( \int (\psi w)^2 dg(y,s) \Big)^{2/n}. \tag{6.5.9}$$

Let us assume that $B(x, \sigma r, s)$ is a proper sub-domain of $\mathbf{M}$. In this case, for manifolds with nonnegative Ricci curvature, it is well-known that the following Sobolev imbedding holds (see [Sal] e.g.)

$$\Big( \int (\psi w)^{2n/(n-2)} dg(y,s) \Big)^{(n-2)/n} \le \frac{c_n \sigma^2 r^2}{|B(x,\sigma r, s)|_s^{2/n}} \int [|\nabla (\psi w)|^2 + r^{-2} (\psi w)^2] dg(y,s). \tag{6.5.10}$$

For $s \in [t - (\sigma r)^2, t]$, by the assumption that the Ricci curvature is nonnegative, it holds

$$B(x, \sigma r, s) \supset B(x, \sigma r, t); \qquad |B(x, \sigma r, s)|_s \ge |B(x, \sigma r, t)|_{t-(\sigma r)^2}.$$

Therefore we have, for $s \in [t - (\sigma r)^2, t]$,

$$\Big( \int (\psi w)^{2n/(n-2)} dg(y,s) \Big)^{(n-2)/n} \le \frac{c_n \sigma^2 r^2}{|B(x,\sigma r, t)|_{t-(\sigma r)^2}^{2/n}} \int [|\nabla (\psi w)|^2 + r^{-2} (\psi w)^2] dg(y,s). \tag{6.5.11}$$

Substituting (6.5.9) and (6.5.11) to (6.5.8), we arrive at the estimate

$$\int_{Q_r(x,t)} w^{2\theta} dg(y,s) ds \le c_n \frac{r^2}{|B(x,\sigma r, t)|_{t-(\sigma r)^2}^{2/n}} \Big( \frac{1}{(\sigma-1)^2 r^2} \int_{Q_{\sigma r}(x,t)} w^2 dg(y,s) ds \Big)^{\theta},$$

with $\theta = 1 + (2/n)$. Now we apply the above inequality with the parameters

$$\sigma_0 = 2, \sigma_i = 2 - \Sigma_{j=1}^i 2^{-j}, \quad p_i = \theta^i.$$

This shows a $L^2$ mean value inequality

$$\sup_{Q_{r/2}(x,t)} u^2 \le \frac{c_n}{r^2|B(x,r,t)|_{t-r^2}} \int_{Q_r(x,t)} u^2 dg(y,s)ds. \tag{6.5.12}$$

From here, by a generic trick of Li and Schoen [LS] e.g., applicable here since it uses only the doubling property of the metric balls, we arrive at the $L^1$ mean value inequality, for $r > 0$,

$$\sup_{Q_{r/2}(x,t)} u \le \frac{c_n}{r^2|B(x,r,t)|_{t-r^2}} \int_{Q_r(x,t)} u dg(z,\tau)d\tau. \tag{6.5.13}$$

Fixing $y \in \mathbf{M}$ and $s < t$, we apply (6.5.13) on $u = G(\cdot, \cdot; y, s)$ with $r = \sqrt{t-s}/2$. Note that $\int_{\mathbf{M}} u(z,\tau)dg(z,\tau) = 1$. The doubling property of the geodesic balls show that

$$G(x,t;y,s) \le \frac{c_n}{|B(x,\sqrt{t-s},t)|_s} \tag{6.5.14}$$

when $|B(x,\sqrt{t-s},s)|$ is a proper subdomain of $\mathbf{M}$.

Next we claim that, for some constant $A > 0$,

$$G(x,t;y,s) \le \frac{A}{(t-s)^{n/2}}, \tag{6.5.15}$$

when $|B(x,\sqrt{t-s},s)|$ is a proper subdomain of $\mathbf{M}$. The proof is almost a rerun of the proof of (6.5.14), except that we replace the Sobolev imbedding (6.5.10), by the following one:

$$\left(\int (\psi w)^{2n/(n-2)} dg(y,s)\right)^{(n-2)/n} \le s_0 \int [|\nabla(\psi w)|^2 + (\psi w)^2] dg(y,s).$$

This well-known inequality is valid under the assumption that $Ricc \ge 0$ and the injectivity radius is bounded from below by a positive constant. See [Heb1] e.g.

If $B(x,\sqrt{t-s},s)$ is not a proper subdomain of $\mathbf{M}$, then we can just carry out the argument in the previous paragraph without using cut off functions to prove

$$G(x,t;y,s) \le A.$$

*Step 2.* Proving the full bound.

It is obvious that we only have to deal with the case that $B(x, 2\sqrt{t-s}, s)$ is a proper subdomain of $\mathbf{M}$. Otherwise, $2\sqrt{t-s} \geq d(x, y, s)$ for any $x, y \in \mathbf{M}$. So the exponential term is mute.

Without loss of generality, we take $s = 0$. We begin by using a modified version of the exponential weight method due to Davies [Da]. Pick a point $x_0 \in \mathbf{M}$, a number $\lambda < 0$ and a function $f \in L^2(\mathbf{M}, g(0))$. Consider the functions $F$ and $u$ defined by

$$\begin{aligned} F(x,t) &\equiv e^{\lambda d(x,x_0,t)} u(x,t) \\ &\equiv e^{\lambda d(x,x_0,t)} \int G(x,t;y,0) e^{-\lambda d(y,x_0,0)} f(y) dg(y,0). \end{aligned}$$

It is clear that $u$ is a solution of (6.5.3). By direct computation, we have

$$\begin{aligned} \partial_t \int F^2(x,t) dg(x,t) &= \partial_t \int e^{2\lambda d(x,x_0,t)} u^2(x,t) dg(x,t) \\ &= 2\lambda \int e^{2\lambda d(x,x_0,t)} \partial_t d(x,x_0,t) u^2(x,t) dg(x,t) \\ &\quad + \int e^{2\lambda d(x,x_0,t)} u^2(x,t) R(x,t) dg(x,t) \\ &\quad + 2 \int e^{2\lambda d(x,x_0,t)} [\Delta u - Ru] u(x,t) dg(x,t). \end{aligned}$$

By the assumption that $Ricci \geq 0$ and $\lambda < 0$, we know as before

$$\lambda \partial_t d(x,x_0,t) \leq 0.$$

Hence the above shows

$$\partial_t \int F^2(x,t) dg(x,t) \leq 2 \int e^{2\lambda d(x,x_0,t)} u \Delta u dg(x,t).$$

Using integration by parts, we turn this inequality into

$$\begin{aligned} \partial_t \int F^2(x,t) dg(x,t) \leq &-4\lambda \int e^{2\lambda d(x,x_0,t)} u \nabla d(x,x_0,t) \nabla u dg(x,t) \\ &- 2 \int e^{2\lambda d(x,x_0,t)} |\nabla u|^2 dg(x,t). \end{aligned}$$

Observe also

$$\begin{aligned} &\int |\nabla F(x,t)|^2 dg(x,t) = \int |\nabla (e^{\lambda d(x,x_0,t)} u(x,t))|^2 dg(x,t) \\ &\quad = \int e^{2\lambda d(x,x_0,t)} |\nabla u|^2 dg(x,t) + 2\lambda \int e^{2\lambda d(x,x_0,t)} u \nabla d(x,x_0,t) \nabla u dg(x,t) \\ &\qquad + \lambda^2 \int e^{2\lambda d(x,x_0,t)} |\nabla d|^2 u^2 dg(x,t). \end{aligned}$$

Combining the last two expressions, we deduce

$$\partial_t \int F^2(x,t)dg(x,t) \leq -2\int |\nabla F(x,t)|^2 dg(x,t) + \lambda^2 \int e^{2\lambda d(x,x_0,t)}|\nabla d|^2 u^2 dg(x,t).$$

By the definition of $F$ and $u$, this shows

$$\partial_t \int F^2(x,t)dg(x,t) \leq \lambda^2 \int F(x,t)^2 dg(x,t).$$

Upon integration, we derive the following $L^2$ estimate

$$\int F^2(x,t)dg(x,t) \leq e^{\lambda^2 t}\int F^2(x,0)dg(x,0) = e^{\lambda^2 t}\int f(x)^2 dg(x,0). \tag{6.5.16}$$

Recall that $u$ is a solution to (6.5.3). Therefore, by the mean value inequality (6.5.12), the following holds

$$u(x,t)^2 \leq \frac{c_n}{t|B(x,\sqrt{t/2},t)|_{t/2}} \int_{t/2}^{t}\int_{B(x,\sqrt{t/2},\tau)} u^2(z,\tau)dg(z,\tau)d\tau.$$

By the definition of $F$ and $u$, it follows that

$$u(x,t)^2 \leq \frac{c_n}{t|B(x,\sqrt{t/2},t)|_{t/2}} \int_{t/2}^{t}\int_{B(x,\sqrt{t/2},\tau)} e^{-2\lambda d(z,x_0,\tau)}F^2(z,\tau)dg(z,\tau)d\tau.$$

In particular, this holds for $x = x_0$. So, for $z \in B(x_0,\sqrt{t/2},\tau)$, there holds $d(z,x_0,\tau) \leq \sqrt{t/2}$. Therefore, by the assumption that $\lambda < 0$,

$$u(x_0,t)^2 \leq \frac{c_n e^{-2\lambda\sqrt{t/2}}}{t|B(x_0,\sqrt{t/2},t)|_{t/2}} \int_{t/2}^{t}\int_{B(x_0,\sqrt{t/2},\tau)} F^2(z,\tau)dg(z,\tau)d\tau.$$

This combined with (6.5.16) shows that

$$u(x_0,t)^2 \leq \frac{c_n e^{\lambda^2 t-\lambda\sqrt{2t}}}{|B(x_0,\sqrt{t/2},t)|_{t/2}} \int f(y)^2 dg(y,0).$$

i.e.

$$\left(\int G(x_0,t;z,0)e^{-\lambda d(z,x_0,0)}f(z)dg(z,0)\right)^2 \le \frac{c_n e^{\lambda^2 t-\lambda\sqrt{2t}}}{|B(x_0,\sqrt{t/2},t)|_{t/2}} \times \int f(y)^2 dg(y,0). \tag{6.5.17}$$

Now, we fix $y_0$ such that $d(y_0,x_0,0)^2 \ge 4t$. Then it is clear that, by $\lambda < 0$ and the triangle inequality,

$$-\lambda d(z,x_0,0) \ge -\frac{\lambda}{2} d(x_0,y_0,0)$$

when $d(z,y_0,0) \le \sqrt{t}$. In this case, (6.5.17) implies

$$\left(\int_{B(y_0,\sqrt{t},0)} G(x_0,t;z,0)f(z)dg(z,0)\right)^2 \le \frac{c_n e^{\lambda d(x_0,y_0,0)+\lambda^2 t-\lambda\sqrt{2t}}}{|B(x_0,\sqrt{t/2},t)|_{t/2}} \times \int f(y)^2 dg(y,0). \tag{6.5.18}$$

Now we take

$$\lambda = -\frac{d(x_0,y_0,0)}{bt}.$$

with $b > 0$ sufficiently large. Then (6.5.18) shows, for some $c > 0$,

$$\int_{B(y_0,\sqrt{t},0)} G^2(x_0,t;z,0)dg(z,0) \le \frac{c_n e^{-cd(x_0,y_0,0)^2/t}}{|B(x_0,\sqrt{t/2},t)|_{t/2}}.$$

Hence, there exists $z_0 \in B(y_0,\sqrt{t},0)$ such that

$$G^2(x_0,t;z_0,0) \le \frac{c_n e^{-cd(x_0,y_0,0)^2/t}}{|B(x_0,\sqrt{t/2},t)|_{t/2}|B(x_0,\sqrt{t},0)|_0}.$$

By the doubling property of the geodesic balls, it implies

$$G^2(x_0,t;z_0,0) \le \frac{c_n e^{-cd(x_0,y_0,0)^2/t}}{|B(x_0,\sqrt{t},t)|_0|B(x_0,\sqrt{t},0)|_0}. \tag{6.5.19}$$

Finally, let us remind ourself that $G(x_0,t;\cdot,\cdot)$ is a solution to the conjugate equation of (6.5.3), i.e.

$$\Delta_z G(x,t;z;\tau) + \partial_\tau G(x,t;z,\tau) = 0.$$

Therefore Theorem 6.5.1 can be applied to it after a reversal in time. Consequently, for $\delta > 0, C > 0$,

$$G(x_0, t; y_0, 0) \leq CG^{1/(1+\delta)}(x_0, t, z_0, 0)M^{\delta/(1+\delta)}, \tag{6.5.20}$$

where $M = \sup_{M\times[0,t/2]} G(x_0, t, \cdot, \cdot)$. By (6.5.15), there exists a constant $A > 0$, depending only on the lower bound of the injectivity radius such that

$$M \leq A \max\{\frac{1}{t^{n/2}}, 1\}.$$

This, (6.5.19) and (6.5.20) show, with $\delta = 1$, that

$$G(x_0, t; y_0, 0)^2 \leq \max\{\frac{c_n}{t^{n/2}}, 1\} \frac{Ae^{-cd(x_0,y_0,0)^2/t}}{\sqrt{|B(x_0, \sqrt{t}, t)|_0 |B(x_0, \sqrt{t}, 0)|_0}}.$$

By the assumption that the Ricci curvature is nonnegative, we have

$$|B(x_0, \sqrt{t}, t)|_0 \leq |B(x_0, \sqrt{t}, 0)|_0.$$

Therefore

$$G^2(x_0, t; y_0, 0) \leq \max\{\frac{c_n}{t^{n/2}}, 1\} \frac{Ae^{-cd(x_0,y_0,0)^2/t}}{|B(x_0, \sqrt{t}, t)|_0}.$$

Consequently

$$G(x_0, t; y_0, 0) \leq c_n A \left(1 + \frac{1}{t^{n/2}} + \frac{1}{|B(x_0, \sqrt{t}, t)|_0}\right) e^{-cd(x_0,y_0,0)^2/t}.$$

Since $x_0$ and $y_0$ are arbitrary, the proof is done. □

# Chapter 7

# Properties of ancient $\kappa$ solutions and singularity analysis for 3-dimensional Ricci flow

The main objective of this chapter is to show, with the help of Perelman's $\kappa$ noncollapsing Theorem 6.1.2 and certain blow-up procedure, that blow-up limit of a finite time singularity in a Ricci flow is an ancient ancient $\kappa$ solution. In order to understand the structure of singularity, the first logical step is to study the structure of ancient $\kappa$ solutions.

## 7.1 Preliminaries

In this section we list a few basic results and concepts of various types, which are necessary for the main results in this chapter. Most of the proofs will be omitted.

**Lemma 7.1.1** *(point picking lemma) Let* $\mathbf{M}$ *be a Riemann manifold and* $R$ *be the scalar curvature. Suppose* $B(x, 5r)$, $r > 0$, *is a proper subset of* $\mathbf{M}$. *Then there exists a ball* $B(y, \rho) \subset B(x, 5r)$, $\rho \leq r$, *such that*

*(i).* $R(z) \leq 2R(y)$ *for all* $z \in B(y, \rho)$;

*(ii).* $R(y)\rho^2 \geq R(x)r^2$.

PROOF. We define sequences of points $x_i \in B(x, 5r)$ and numbers $r_i > 0$ by the following induction. Let $x_1 = x$ and $r_1 = r$. For $i > 1$, define $x_{i+1} = x_i$ and $r_{i+1} = r_i$ if $R(z) \le 2R(x_i)$ for all $z \in B(x_i, r_i)$. Otherwise, we pick $x_{i+1}$ as a point in $B(x_i, r_i)$ such that $R(x_{i+1}) > 2R(x_i)$ and pick $r_{i+1} = r_i/\sqrt{2}$. In either case it holds $R(x_{i+1})r_{i+1}^2 \ge R(x_i)r_i^2$. Since $d(x, x_i) \le r\Sigma_{k=0}^{\infty}(1/\sqrt{2})^k < 4r$, we know that $B(x_i, r_i) \subset B(x, 5r)$. Due to the boundedness of $R$ in the ball $B(x, 5r)$, there exists a positive integer $j$ such that $x_i = x_j$ and $r_i = r_j$ for all $i \ge j$. The lemma is proven by taking $y = x_j$ and $\rho = r_j$. □

**Theorem 7.1.1** *(Classical splitting theorem) 1. Toponogov Splitting theorem [Topo]: If an $n$-dimensional complete Riemann manifold of nonnegative sectional curvature contains a geodesic line, then the manifold is a Riemannian product of* $\mathbf{R}$ *and an* $(n-1)$*-dimensional Riemann manifold of nonnegative sectional curvature.*

*2. Cheeger-Gromoll Splitting theorem [CG]: The above conclusion still holds for complete Riemann manifold of nonnegative Ricci curvature.*

Here a geodesic line means a distance minimizing geodesic without ends and with infinite length. The theorem allows one to factor certain high dimensional manifold into the product of lower dimensional ones, which may be easier to understand. Of course this is in the same spirit as factoring a large integer into the product of smaller ones. One can also consult Section 9.3.2 of [Pet] for a proof of this theorem.

**Theorem 7.1.2** *(Soul theorem, Cheeger-Gromoll-Meyer) If* $\mathbf{M}$ *is a complete, noncompact, $n$-dimensional Riemann manifold with nonnegative sectional curvature, then* $\mathbf{M}$ *contains a soul* $S \subset \mathbf{M}$. *The soul is a closed totally convex submanifold, such that* $\mathbf{M}$ *is diffeomorphic to the normal bundle over $S$. In addition, if the sectional curvature is positive everywhere, the soul is a point and* $\mathbf{M}$ *is diffeomorphic to* $\mathbf{R^n}$.

A proof of this theorem can be found in [CG2]. See also [Pet] for a proof suited for students and an informative discussion on the history of the theorem.

**Remark 7.1.1** *G. Perelman [P4] gave a strikingly short proof of Cheeger-Gromoll's conjecture that the above* $\mathbf{M}$ *is diffeomorphic to* $\mathbf{R^n}$ *if the sectional curvature is nonnegative everywhere and positive at one point.*

The next result dates back to the classical splitting theorem of Cheeger-Gromoll and Toponogov.

**Proposition 7.1.1** *(dimension reduction) Let* $(\mathbf{M}, g)$ *be a complete manifold with nonnegative sectional curvature. Let* $\{P_k\} \subset \mathbf{M}$ *and* $\{\lambda_k\} \subset (0, \infty)$ *be two sequences such that* $d(P_1, P_k) \to \infty$ *and* $\lambda_k d(P_1, P_k) \to \infty$ *when* $k \to \infty$. *Suppose the marked manifolds* $(\mathbf{M}, \lambda_k^2 g, P_k)$ *converge in* $C^\infty_{loc}$ *topology to a Riemann manifold* $\mathbf{M}_\infty$. *Then* $\mathbf{M}_\infty$ *splits as a metric product* $\mathbf{M}_\infty = N \times \mathbf{R}$ *where* $N$ *is a Riemann manifold with nonnegative sectional curvature.*

PROOF. The idea is to show that $\mathbf{M}_\infty$ contains a minimizing geodesic line, i.e. a minimizing geodesic without ends, which also has infinite length. Then the classical splitting theorem mentioned above yields the desired splitting.

Let $\gamma_k$ and $\sigma_k$ be minimizing geodesics connecting $P_1$ with $P_k$ and $P_k$ with $P_{k+1}$ respectively. By choosing a subsequence if necessary, we can assume the following properties hold:

$$d(P_1, P_{k+1}) \geq 2d(P_1, P_k) + 1 \tag{7.1.1}$$

$$\delta_k \equiv \angle(\gamma_k(0), \gamma_{k+1}(0)) \leq 1/k.$$

Here $\angle(\gamma_k(0), \gamma_{k+1}(0))$ is the angle between the tangent vectors of $\gamma_k$ and $\gamma_{k+1}$ at the point $P_1$. From the hypotheses and a suitable choice of subsequence, we can assume that the marked sequence $(\mathbf{M}, \lambda_k^2 g, P_k)$ converges in $C^\infty_{loc}$ sense to a marked manifold $(\tilde{M}, \tilde{g}, \tilde{P})$ which has nonnegative sectional curvature. Moreover the geodesics $\gamma_k$ and $\sigma_k$ converge to geodesic rays $\tilde{\gamma}$ and $\tilde{\sigma}$.

Pick an arbitrary point $\tilde{A} \in \tilde{\gamma}$ and another one $\tilde{B} \in \tilde{\sigma}$. Let $a = \tilde{d}(\tilde{A}, \tilde{P})$, $b = \tilde{d}(\tilde{B}, \tilde{P})$ and $c = \tilde{d}(\tilde{A}, \tilde{B})$. Here $\tilde{d}$ is the distance under the limit metric $\tilde{g}$. We claim that

$$a + b = c \tag{7.1.2}$$

which will show that $\tilde{\gamma} \cup \tilde{\sigma}$ is a geodesic line. By definition of $\tilde{M}$, there exist two sequences of points $A_k \in \gamma_k$ and $B_k \in \sigma_k$ such that, as $k \to \infty$,

$$\lambda_k d(A_k, P_k) \to a, \quad \lambda_k d(B_k, P_k) \to b, \quad \lambda_k d(A_k, B_k) \to c.$$

Consider the geodesic triangles $\Delta P_k P_1 P_{k+1}$, $\Delta P_k A_k B_k$ in $\mathbf{M}$ and their comparison triangles $\Delta \bar{P}_k \bar{P}_1 \bar{P}_{k+1}$, $\Delta \bar{P}_k \bar{A}_k \bar{B}_k$ in $\mathbf{R^2}$. Notice that

$$d(P_k, P_1) = |\bar{P}_k \bar{P}_1|, \quad d(P_k, P_{k+1}) = |\bar{P}_k \bar{P}_{k+1}|, \quad d(P_1, P_{k+1}) = |\bar{P}_1 \bar{P}_{k+1}|,$$

$$d(P_k, A_k) = |\bar{P}_k \bar{A}_k|, \quad d(P_k, B_k) = |\bar{P}_k \bar{B}_k|, \quad d(A_k, B_k) = |\bar{A}_k \bar{B}_k|.$$

Using the Toponogov comparison theorem (see [CE] e.g.), we have

$$\angle \bar{A}_k \bar{P}_k \bar{B}_k \geq \angle \bar{P}_1 \bar{P}_k \bar{P}_{k+1},$$

$$\angle \bar{P}_k \bar{P}_1 \bar{P}_{k+1} \leq \angle P_k P_1 P_{k+1} = \delta_k \leq 1/k.$$

Also note that

$$\angle \bar{P}_k \bar{P}_{k+1} \bar{P}_1 \leq \angle \bar{P}_k \bar{P}_1 \bar{P}_{k+1} \leq 1/k$$

which is due to (7.1.1). Thus

$$\angle \bar{A}_k \bar{P}_k \bar{B}_k \geq \pi - (2/k).$$

The Euclidean law of cosine shows

$$|\bar{A}_k \bar{B}_k|^2 \geq |\bar{A}_k \bar{P}_k|^2 + |\bar{P}_k \bar{B}_k|^2 - 2|\bar{A}_k \bar{P}_k| \; |\bar{P}_k \bar{B}_k| \; \cos(\pi - (2/k)).$$

i.e.

$$\begin{aligned} d(A_k, B_k)^2 \geq & d(A_k, P_k)^2 + d(P_k, B_k)^2 \\ & - 2d(A_k, P_k) d(P_k, B_k) \; \cos(\pi - (2/k)). \end{aligned}$$

Multiplying this inequality by $\lambda_k^2$ and taking $k \to \infty$, we arrive at

$$c \geq a + b \geq c.$$

This shows $\tilde{\gamma} \cup \tilde{\sigma}$ is a geodesic line. The classical splitting theorem tells us that $\tilde{M} = N \times \mathbf{R}$ as stated. □

As mentioned earlier, the study of singularity of Ricci flow relies on the study of $\kappa$ solutions (Definition 5.4.1). If a three dimensional $\kappa$ solution $M$ can be split as $N \times \mathbf{R}$, then $N$ is a 2 dimensional ancient $\kappa$ solution. Interestingly there are only two such 2 dimensional solutions.

**Theorem 7.1.3** *(2-dimensional $\kappa$ solution) The only 2-dimensional $\kappa$ solutions are the round $S^2$ and the round $RP^2$, the real projective plane.*

This theorem was proven by Hamilton in Section 26 of [Ha7]. See also p. 369 of [CZ] for a different proof. Clearly this theorem and splitting theorem are instrumental in the understanding of higher dimensional $\kappa$ solutions.

When analyzing singularity of Ricci flow, one often encounters some special manifolds. Here we give one of them.

**Definition 7.1.1** *($\epsilon$ necks and their center points) Given $\epsilon > 0$, an $\epsilon$ neck in a 3 manifold $(\mathbf{M}, g)$ centered at a point $x$ is an open set $U \subset \mathbf{M}$, satisfying the conditions:*

*(i) $x \in U$.*

*(ii) There exists a diffeomorphism $\phi$ from the cylinder $S^2 \times (-\epsilon^{-1}, \epsilon^{-1})$ onto $U$ such that $\phi^{-1}(x) \in S^2 \times \{0\}$.*

*(iii) The normalized pullback metric $R(x)\phi^* g$ lies within $\epsilon$ of the standard round metric on the cylinder in $C^{[\epsilon^{-1}]}$ topology. Here $R(x)$ is the scalar curvature at $x$.*

*(iv) The set $\phi(S^2 \times \{0\})$ is called the center (set) or central 2 sphere of the $\epsilon$ neck and every point in the central 2 sphere is called a center of the $\epsilon$ neck.*

**Remark 7.1.2** *Let $z$ be a number in $(-\epsilon^{-1}, \epsilon^{-1})$ and $\theta \in S^2$, $(\theta, z)$ is a parametrization of $U$ via the diffeomorphism $\phi$. We can identify the metric on $U$ with its pull back on the round neck by $\phi$ in this manner.*

**Proposition 7.1.2** *(no arbitrarily small $\epsilon$ necks) Let $M$ be any complete 3 manifold with nonnegative sectional curvature. There exists $\epsilon_0 > 0$ such that for any $\epsilon \in (0, \epsilon_0]$, $M$ does not contain $\epsilon$ necks of arbitrarily small radius.*

Proof. This result seems to have been known by experts in Riemann geometry for a while. Intuitively, if $M$ has an arbitrarily small $\epsilon$ neck, then it looks like a cusp. Thus some sectional curvatures have to be negative somewhere. A proof of the proposition can be found on p357 in [CZ] and a different proof in Chapter 2 of [MT]. The former proof uses comparison theorem and Buseman function. The later proof uses the soul theorem and Buseman function.

We will present a condensed proof following [MT] where great details can be found. The details that we are skipping are relatively straightforward and intuitively clear.

If the proposition is false, then one can find a sequence of $\epsilon$ necks $N_i$ whose radii converge to 0. Consider two disjoint $\epsilon$ necks $N_1$ and $N_i$ in the sequence. By the Soul Theorem 7.1.2, if $\epsilon$ is sufficiently small, one can prove that the central 2 spheres of $N_1$ and $N_i$ are the boundary components of a region called $X$ which is diffeomorphic to $S^2 \times [-1, 1]$. For a proof of this statement, see Section 2.5 of [MT].

Let $c = c(t)$ be a minimal geodesic ray, starting from a point $p$ outside of $X$, which traverses $N_1$ and $N_i$. We also let $t$ be the arclength.

Let $B$ be the associated Buseman function, i.e.

$$B(x) = \lim_{t\to\infty} (d(c(t), x) - t). \tag{7.1.3}$$

Since the curvature is nonnegative, it is well known that $\Delta B \le 0$ in the weak sense, i.e. $B$ is super harmonic (see Chapter 1 of [SY] e.g.).

Next we construct a Lipschitz cut off function $\lambda$ with the following properties:

(1) $supp\lambda \subset X$; for points in $N_1, N_i$, $\lambda = \lambda(z)$ where $z$ is the longitudinal parameters of the $\epsilon$ necks, described in Remark 7.1.2.

(2) Let $\phi_1$ and $\phi_i$ be the diffeomorphisms in the definition of $\epsilon$ necks $N_1$ and $N_i$. We make $\lambda = 0$ to the left $\{z = 0\}$, the central 2 sphere of $N_1$, linear for $z \in [0, 1]$ and becomes 1 to the right $\phi_1(S^2 \times \{1\})$ all the way to $\phi_i(S^2 \times \{-1\})$. Here and later the words "left" and "right" mean the direction of the negative and positive $z$ axis.

(3) $\lambda$ is linear from $\phi_i(S^2 \times \{-1\})$ to the central 2 sphere of $N_i$ and becomes 0 to the right of the central 2 sphere.

When $p$, the starting point of the geodesic ray, is sufficiently far from the $\epsilon$ necks $N_1$ and $N_2$, one can prove that $\nabla B$ is $\epsilon$ close to $\nabla z$ (see Section 2.5 of [MT] e.g.). Since $B$ is super harmonic, we have, since $\nabla\lambda = \lambda'\nabla z$

$$\begin{aligned} 0 \le \int_M < \nabla B, \nabla\lambda > d\mu &= \int_{supp\nabla\lambda} < \nabla B, \nabla\lambda > d\mu \\ &= (a_2 R(x_i)^{-1} - a_1 R(x_1)^{-1})|S^2|. \end{aligned}$$

Here $a_1$ and $a_i$ are constants which converge to a positive number $a_0$ when $\epsilon \to 0$. Also $x_1$ and $x_i$ are in the centers of $N_1$ and $N_i$ respectively. Therefore $R(x_i) \le 2R(x_1)$ when $\epsilon$ is sufficiently small. This shows that the radius of $N_i$, comparable to $R(x_i)^{-1/2}$, can not be arbitrarily small.

□

**Definition 7.1.2** *(cone) A (open) cone over a Riemann manifold* $(N, g)$ *is the manifold* $J(N) \equiv N \times (0, \infty)$ *equipped with the metric*

$$\tilde{g}(x, s) = s^2 g(x) + ds^2.$$

**Proposition 7.1.3** *(curvature of a cone) Let* $N$ *be a Riemann manifold of dimension* $n$ *and* $(x^1, \ldots, x^n)$ *be a local coordinate. Let* $(x^1, \ldots, x^n, x^0)$ *be a local coordinate for the cone* $J(N)$*. Denote by* $Rm_g$ *the curvature tensor for* $(N, g)$ *and by* $\tilde{Rm}_{\tilde{g}}$ *the curvature tensor for the* *cone. Then there hold:*

$$\tilde{R}m_{\tilde{g}}(\partial_i, \partial_j)\partial_0 = 0, \quad 0 \le i, j \le n,$$
$$\tilde{R}m_{\tilde{g}}(\partial_i, \partial_j)\partial_i = Rm_g(\partial_i, \partial_j) + g_{ij}\partial_j - g_{ji}\partial_i, \quad 1 \le i, j \le n.$$

*Moreover let $\lambda_k$, $k = 1, \ldots, n(n-1)/2$ be the eigenvalues of $Rm_g$ at a point $p \in N$. Then for any $s > 0$, there are $n$ zero eigenvalues of $\tilde{R}m_{\tilde{g}}$ at the point $(p, s) \in J(N)$. The other eigenvalues of $\tilde{R}m_{\tilde{g}}$ are $s^{-2}(\lambda_k - 1)$ where $k = 1, \ldots, n(n-1)/2$.*

**Exercise 7.1.1** *Prove Proposition 7.1.3.*

## 7.2 Bounds for the heat kernel of conjugate heat equation on $\kappa$ solutions

In this section, we establish a certain Gaussian type upper bound for the heat kernel of the conjugate heat equation associated with 3-dimensional ancient $\kappa$ solutions to the Ricci flow. The material is an expanded version of [Z5].

The proof follows the framework in the last section of Chapter 6. There is an upper bound in the case of Ricci flow with nonnegative Ricci curvature was given. In the current situation, the ancient $\kappa$ solutions provide better control on curvature and volume. These allow us to find a better Gaussian upper bound for the heat kernel.

Using this heat kernel bound, in the next section, we show that the $W$ entropy associated with the heat kernel is uniformly bounded from below after certain scaling. After this is done, we use Perelman's monotonicity formula for the $W$ entropy to prove the backward limit is a gradient shrinking Ricci soliton.

One more notation that will be used in the section is that for the space time region

$$P(x_0, t_0, r, -r^2) \equiv \{(x,t) \ |d(x, x_0, t) < r, \ t_0 - r^2 < t < t_0\}.$$

Here $r > 0$ and $(x_0, t_0)$ is a point in space time.

Let us recall the concept of $\kappa$ solutions (Definition 5.4.1).

**Definition 7.2.1** *A solution to the Ricci flow $\partial_t g = -2Ric$ in a $n$ dimensional manifold* **M** *is a $\kappa$ solution or ancient $\kappa$ solution if it satisfies the following properties.*

*1. It is complete (compact or noncompact) and defined on an ancient time interval $(-\infty, T_0]$, $T_0 \ge 0$.*

*2. It has nonnegative curvature operator and bounded curvature at each time level.*

*3. It is $\kappa$ noncollapsed on all scales for some positive constant $\kappa$, i.e.*

*for any $x_0 \in \mathbf{M}$, $t_0 \in (-\infty, T_0]$ and any $r > 0$, if $|Rm| \le r^{-2}$ on $P(x_0, t_0, r, -r^2)$, then $|B(x_0, r, t_0)|_{t_0} \ge \kappa r^n$.*

*4. It is nonflat.*

Occasionally we will also use the term nonflat $\kappa$ solutions to stress that the solution is not flat.

For convenience, we take the final time $T_0$ of the ancient solution to be 0 throughout the section. The conjugate heat equation is

$$\Delta u - Ru - \partial_\tau u = 0. \tag{7.2.1}$$

Here and always $\tau = -t$. $\Delta$ and $R$ are the Laplace-Beltrami operator and the scalar curvature with respect to $g(t)$. This equation, coupled with the initial value $u_{\tau=0} = u_0$ is well posed if $\mathbf{M}$ is compact or the curvature is bounded, and if $u_0$ is bounded [Gro].

We use $G = G(x, \tau; x_0, \tau_0)$ to denote the heat kernel (fundamental solution) of (7.2.1). Here $\tau > \tau_0$ and $x, x_0 \in \mathbf{M}$. Heat kernel estimate is a traditionally active area of research with many applications. For the conjugate heat equation, existence of $G$ was established in [Gro]. Various bounds for $G$ were proven in [Gro], [P1] Section 9, [Ni], [Cetc] and [Z1]. The main technical result of the section is

**Theorem 7.2.1** *(i) Let $(\mathbf{M}, g(t))$ be a $n$ dimensional ancient $\kappa$ solution of the Ricci flow. Suppose also that $R(x,t) \le \frac{D_0}{1+|t|}$ for some $D_0 > 0$ and for $t \in [-T, 0]$. Here $T$ is any positive number or $T = \infty$. Then exist positive numbers $a$ and $b$ depending only on $n$, $\kappa$ and $D_0$ such that the following holds.*

*For all $x, x_0 \in \mathbf{M}$,*

$$G(x, \tau; x_0, \tau_0) \le \frac{a}{(\tau - \tau_0)^{n/2}},$$

*and*

$$G(x, \tau; x_0, \tau_0) \le \frac{a}{|B(x, \sqrt{\tau - \tau_0}, t_0)|_{t_0}} e^{-bd^2(x, x_0, t_0)/(\tau - \tau_0)},$$

*where $\tau = -t$, $\tau_0 = -t_0$, $\tau > \tau_0 \ge 0$ and $t \in [-T, 0]$.*

*(ii) In particular, if $R(x,t) \le \frac{D_0}{1+|t|}$ for all $t \le 0$, namely $(\mathbf{M}, g(t))$ is a Type I ancient $\kappa$ solution, there exist positive numbers $a_1$ and $b_1$*

*depending only on $\kappa$, $n$, and $D_0$ such that the following holds. For all $x, x_0 \in \mathbf{M}$, and all $\tau = -t > 0$,*

$$\frac{1}{a_1 \tau^{n/2}} e^{-d^2(x,x_0,t)/(b_1 \tau)} \le G(x, \tau; x_0, \tau/2) \le \frac{a_1}{\tau^{n/2}} e^{-b_1 d^2(x,x_0,t)/\tau}.$$

**Remark 7.2.1** *A natural question is whether the bounds in part (ii) holds when $\tau/2$ is replaced by $0$.*

Proof. We divide the proof into three steps. The first two are for part (i) of the theorem. We always assume that all the time variables involved are not smaller than $-T$, so that the condition $R(\cdot.t) \le \frac{D_0}{1+|t|}$ holds. The proof of this theorem is similar to that of Theorem 6.5.2. Comparing with that case, we have two new ingredients coming from ancient $\kappa$ solutions. One is the noncollapsing condition on all scales. The other is the bound on the scalar curvature. These allow us to prove a better bound. Without loss of generality we assume $\tau_0 = 0$ in $G(x, \tau; x_0, \tau_0)$. It is convenient to work with the reversed time $\tau$. Note that the Ricci flow is a backward flow with respect to $\tau$ and the conjugate heat equation is a forward heat equation with a potential term.

*Step 1.* Since *Ricci* $\ge 0$, it is well known (see Theorem 3.7 [Heb2] e.g.) the following Sobolev inequality holds:

Let $B(x, r, t)$ be a proper subdomain for $(\mathbf{M}, g(t))$. For all $v \in W^{1,2}(B(x, r, t))$, there exists $c_n > 0$ depending only on the dimension $n$ such that

$$\left( \int v^{2n/(n-2)} dg(t) \right)^{(n-2)/n} \le \frac{c_n r^2}{|B(x, r, t)|_t^{2/n}} \int \left[ |\nabla v|^2 + r^{-2} v^2 \right] dg(t). \tag{7.2.2}$$

For our purpose, we only need to take $r = c\sqrt{|t|}$, for $c \le 1$. By the assumption that $R(x, t) \le \frac{D_0}{1+|t|}$ and the $\kappa$ noncollapsing property, we have

$$|B(x, \sqrt{|t|}, t)|_t \ge \kappa D_0^{-n} |t|^{n/2}.$$

Therefore the above Sobolev inequality becomes

$$\left( \int v^{2n/(n-2)} dg(t) \right)^{(n-2)/n} \le \frac{c_n D_0^2}{\kappa^{2/n}} \int \left[ |\nabla v|^2 + |t|^{-1} v^2 \right] dg(t) \tag{7.2.3}$$

for all $v \in W^{1,2}(B(x, \sqrt{|t|}, t))$.

Before moving forward, we would like to clarify a technical point in the definition of Perelman's $\kappa$ noncollapsing as given in Definition

7.2.1. The issue is whether the metric balls $B(x,r,t)$ in the definition are required to be a proper subdomain of the manifold $\mathbf{M}$. When $\mathbf{M}$ is noncompact, $B(x,r,t)$ is always a proper subdomain so this issue is moot. Now one assumes that $\mathbf{M}$ is compact. Without requiring $B(x,r,t)$ being a proper subdomain, if $r$ is larger than the diameter of $\mathbf{M}$, then $B(x,r,t)$ is the whole manifold. In this case $|B(x,r,t)|_t$ can not be greater than $\kappa r^n$ for large $r$. So to be $\kappa$ noncollapsed, at some point in the parabolic ball $|Rm|$ is greater than $1/r^2$. In other words, if $|Rm| \le 1/r^2$ in the parabolic ball, then the volume of the manifold is at least $\kappa r^n$. If the Ricci curvature is nonnegative, then by standard volume comparison theorem, the diameter of the manifold at time $t$ is at least $cr$.

Here, we take this explanation for Perelman's $\kappa$ noncollapsing, i.e. $B(x,r,t)$ in the definition of $\kappa$ ancient solutions is not required to be a proper subdomain. This seems to be the prevailing view in the literature. That is why the Sobolev imbedding (7.2.3) holds without requiring that $B(x,\sqrt{|t|},t)$ is a proper subdomain of $\mathbf{M}$. A natural question is: What happens when $B(x,r,t)$ is implicitly assumed as a proper subdomain in the definition of $\kappa$ solutions? Then we have to make this extra assumption throughout. However either way does not affect the application for the Poincaré conjecture. The reason is that one can always consider the product manifold $\mathbf{M} \times \mathbf{R}$ if $\mathbf{M}$ is compact. See the proof of Theorem 7.3.1, especially Case 4.

Next we show that, under the assumptions of the theorem, $(\mathbf{M}, g(t))$ possess a space time doubling property: the distance between two points at times $t_1$ and $t_2$ are comparable if $t_1$ and $t_2$ are comparable. The proof is very simple. Given $x_1, x_2 \in \mathbf{M}$, let $\mathbf{r}$ be a shortest geodesic connecting the two. Then

$$\partial_t d(x_1,x_2,t) = -\int_{\mathbf{r}} Ric(\partial_r,\partial_r) ds.$$

Since the sectional curvature is nonnegative, it holds

$$|Ric(x,t)| \le R(x,t) \le \frac{D_0}{1+|t|}.$$

Therefore

$$-\frac{D_0}{1+|t|} d(x_1,x_2,t) \le \partial_t d(x_1,x_2,t) \le 0.$$

After integration, we arrive at:

$$(|t_1/|t_2|)^{D_0} \le d(x_1,x_2,t_1)/d(x_1,x_2,t_2) \le 1 \qquad (7.2.4)$$

for all $t_2 < t_1 < 0$. Note that the above inequality is of local nature. If the distance is not smooth, then one can just shift one point, say $x_1$, slightly and then obtain the same integral inequality by taking limits.

Similarly, we have

$$0 \ge \partial_t \int_{B(x,\sqrt{|t_1|},t_1)} dg(t) = -\int_{B(x,\sqrt{|t_1|},t_1)} R(y,t)dg(t) \\ \ge -\frac{D_0}{1+|t|}\int_{B(x,\sqrt{|t_1|},t_1)} dg(t).$$

Upon integration, we know that the volume of the balls

$$|B(x,\sqrt{|t_3|},t_4)|_{t_5} \tag{7.2.5}$$

are all comparable for $t_3, t_4, t_5 \in [t_2, t_1]$, provided that $t_1$ and $t_2$ are comparable.

Let $u$ be a positive solution to (7.2.1) in the region

$$Q_{\sigma r}(x,\tau) \equiv \{(y,s) \mid y \in \mathbf{M}, \tau-(\sigma r)^2 \le s \le \tau,\ d(y,x,-s) \le \sigma r\}.$$

Here $r = \sqrt{|t|}/8 > 0, 2 \ge \sigma \ge 1$. Given any $p \ge 1$, it is clear that

$$\Delta u^p - pRu^p - \partial_\tau u^p \ge 0. \tag{7.2.6}$$

Let $\phi : [0,\infty) \to [0,1]$ be a smooth function such that $|\phi'| \le 2/((\sigma-1)r)$, $\phi' \le 0$, $\phi(\rho) = 1$ when $0 \le \rho \le r$, $\phi(\rho) = 0$ when $\rho \ge \sigma r$. Let $\eta : [0,\infty) \to [0,1]$ be a smooth function such that $|\eta'| \le 2/((\sigma-1)r)^2$, $\eta' \ge 0$, $\eta \ge 0$, $\eta(s) = 1$ when $\tau - r^2 \le s \le \tau$, $\eta(s) = 0$ when $s \le \tau-(\sigma r)^2$. Define a cut-off function $\psi = \phi(d(x,y,-s))\eta(s)$.

Writing $w = u^p$ and using $w\psi^2$ as a test function on (7.2.6), we deduce

$$\int \nabla(w\psi^2)\nabla w dg(y,-s)ds + p\int Rw^2\psi^2 dg(y,-s)ds \\ \le -\int (\partial_s w)w\psi^2 dg(y,-s)ds. \tag{7.2.7}$$

Here and later in this section we also use $dg(\cdot,\cdot)$ to denote the volume element at the space time point $(\cdot,\cdot)$. By direct calculation

$$\int \nabla(w\psi^2)\nabla w dg(y,-s)ds = \int |\nabla(w\psi)|^2 dg(y,-s)ds \\ - \int |\nabla\psi|^2 w^2 dg(y,-s)ds.$$

Next we estimate the right-hand side of (7.2.7).

$$
\begin{aligned}
-\int(\partial_s w) w\psi^2 dg(y,-s)ds \\
&= \int w^2\psi\partial_s\psi dg(y,-s)ds + \frac{1}{2}\int (w\psi)^2 R dg(y,-s)ds \\
&\quad - \frac{1}{2}\int (w\psi)^2 dg(y,-\tau).
\end{aligned}
$$

Observe that

$$
\begin{aligned}
\partial_s\psi =& \eta(s)\phi'(d(y,x,-s))\partial_s d(y,x,-s) + \phi(d(y,x,-s))\eta'(s) \\
&\le \phi(d(y,x,-s))\eta'(s).
\end{aligned}
$$

This is so because $\phi' \le 0$ and $\partial_s d(y,x,-s) \ge 0$ under the Ricci flow with nonnegative Ricci curvature. Hence

$$
\begin{aligned}
&-\int(\partial_s w) w\psi^2 dg(y,-s)ds \\
&\le \int w^2\psi\phi(d(y,x,-s))\eta'(s) dg(y,-s)ds + \frac{1}{2}\int (w\psi)^2 R dg(y,-s)ds \\
&\quad - \frac{1}{2}\int (w\psi)^2 dg(y,-\tau).
\end{aligned} \tag{7.2.8}
$$

Combining (7.2.7) with (7.2.8), we obtain, in view of $p \ge 1$ and $R \ge 0$,

$$
\begin{aligned}
&\int |\nabla(w\psi)|^2 dg(y,-s)ds + \frac{1}{2}\int (w\psi)^2 dg(y,-\tau) \\
&\quad \le \frac{c}{(\sigma-1)^2 r^2}\int_{Q_{\sigma r(x,\tau)}} w^2 dg(y,-s)ds.
\end{aligned} \tag{7.2.9}
$$

By Hölder's inequality

$$
\begin{aligned}
\int (\psi w)^{2(1+(2/n))} dg(y,-s) \le& \left(\int (\psi w)^{2n/(n-2)} dg(y,-s)\right)^{(n-2)/n} \\
&\times \left(\int (\psi w)^2 dg(y,-s)\right)^{2/n}.
\end{aligned} \tag{7.2.10}
$$

By the $\kappa$ noncollapsing assumption, $|B(x,\sqrt{|t|},t)|_t \ge \kappa c_2 r^n$. Recall that $r = \sqrt{|t|}/8$. Since $\mathbf{M}$ has nonnegative Ricci curvature, the diameter of $\mathbf{M}$ at time $t$ is at least a constant multiple of $c\sqrt{|t|}$ for some $c = c_n > 0$. Therefore by the distance doubling property (7.2.4), $B(x, \sigma r, -s)$ is a proper subdomain of $\mathbf{M}$, $s \in [\tau - (\sigma r)^2, \tau]$. Here we just take the number 8 for simplicity. If it is not large enough, we just replace it by a sufficiently large number $D$ and consider $r = \sqrt{|t|}/D$ instead. By the Sobolev inequality (7.2.3), it holds

$$\left(\int (\psi w)^{2n/(n-2)} dg(y,-s)\right)^{(n-2)/n} \leq c(\kappa, D_0) \int [|\nabla(\psi w)|^2 + r^{-2}(\psi w)^2] dg(y,-s),$$

for $s \in [t - (\sigma r)^2, t]$. Substituting this and (7.2.9) to (7.2.10), we arrive at the estimate

$$\int_{Q_r(x,\tau)} w^{2\theta} dg(y,-s)ds \leq c(\kappa, D_0)\left(\frac{1}{(\sigma-1)^2 r^2}\int_{Q_{\sigma r}(x,\tau)} w^2 dg(y,-s)ds\right)^{\theta},$$

with $\theta = 1 + (2/n)$. Now we apply the above inequality repeatedly with the parameters $\sigma_0 = 2, \sigma_i = 2 - \Sigma_{j=1}^{i} 2^{-j}$ and $p_i = \theta^i$. This shows a $L^2$ mean value inequality

$$\sup_{Q_{r/2}(x,\tau)} u^2 \leq \frac{c(\kappa, D_0)}{r^{n+2}} \int_{Q_r(x,\tau)} u^2 dg(y,-s)ds. \tag{7.2.11}$$

This technique is often called Moser's iteration.

This inequality clearly also holds if one replaces $r$ by any positive number $r' < r$ since $|B(x,r',t)| \geq kc_n |B(x,r,t)|(r'/r)^n \geq cr'^n$ by the doubling condition for manifolds with nonnegative Ricci curvature. Then one can just rerun the above Moser's iteration.

From here, by a trick of Li and Schoen [LS], applicable here since it uses only the doubling property of the metric balls, we arrive at the $L^1$ mean value inequality

$$\sup_{Q_{r/2}(x,\tau)} u \leq \frac{c(\kappa, D_0)}{r^{n+2}} \int_{Q_r(x,\tau)} u dg(y,-s)ds.$$

We remark that the doubling constant is uniform since the metrics have nonnegative Ricci curvature.

Now we take $u(x,\tau) = G(x,\tau; x_0, 0)$. Note that $\int_{\mathbf{M}} u(z,s)dg(z,-s) = 1$ and $r = \sqrt{t}/8$. Thus:

$$G(x,\tau; x_0, 0) \leq \frac{c(\kappa, D_0)}{|t|^{n/2}}. \tag{7.2.12}$$

*Step 2.* Proof of the Gaussian upper bound.

We begin by using a modified version of the exponential weight method due to Davies [Da]. Pick a point $x_0 \in \mathbf{M}$, a number $\lambda < 0$ and a function $f \in C_0^\infty(\mathbf{M}, g(0))$. Consider the functions $F$ and $u$ defined by

$$\begin{aligned} F(x,\tau) &\equiv e^{\lambda d(x,x_0,t)} u(x,\tau) \\ &\equiv e^{\lambda d(x,x_0,t)} \int G(x,\tau;y,0) e^{-\lambda d(y,x_0,0)} f(y) dg(y,0). \end{aligned} \tag{7.2.13}$$

Here and always $\tau = -t$. It is clear that $u$ is a solution of (7.2.1). By direct computation,

$$\begin{aligned} \partial_\tau \int F^2(x,\tau) dg(x,t) &= \partial_\tau \int e^{2\lambda d(x,x_0,t)} u^2(x,\tau) dg(x,t) \\ &= 2\lambda \int e^{2\lambda d(x,x_0,t)} \partial_\tau d(x,x_0,t) u^2(x,\tau) dg(x,t) \\ &\quad + \int e^{2\lambda d(x,x_0,t)} u^2(x,\tau) R(x,t) dg(x,t) \\ &\quad + 2 \int e^{2\lambda d(x,x_0,t)} [\Delta u - R(x,t) u(x,\tau)] u(x,\tau) dg(x,t). \end{aligned}$$

By the assumption that *Ricci* $\geq 0$ and $\lambda < 0$, the above shows

$$\partial_\tau \int F^2(x,\tau) dg(x,t) \leq 2 \int e^{2\lambda d(x,x_0,t)} u \Delta u(x,\tau) dg(x,t).$$

Using integration by parts, we turn the above inequality into

$$\begin{aligned} \partial_\tau \int F^2(x,\tau) dg(x,t) \leq &-4\lambda \int e^{2\lambda d(x,x_0,t)} u \nabla d(x,x_0,t) \nabla u dg(x,t) \\ &- 2 \int e^{2\lambda d(x,x_0,t)} |\nabla u|^2 dg(x,t). \end{aligned}$$

Observe also

$$\begin{aligned} &\int |\nabla F(x,\tau)|^2 dg(x,t) = \int |\nabla (e^{\lambda d(x,x_0,t)} u(x,\tau))|^2 dg(x,t) \\ &= \int e^{2\lambda d(x,x_0,t)} |\nabla u|^2 dg(x,t) + 2\lambda \int e^{2\lambda d(x,x_0,t)} u \nabla d(x,x_0,t) \nabla u dg(x,t) \\ &\quad + \lambda^2 \int e^{2\lambda d(x,x_0,t)} |\nabla d|^2 u^2 dg(x,t). \end{aligned}$$

Combining the last two expressions, we deduce

$$\begin{aligned}\partial_\tau \int F^2(x,\tau)dg(x,t) \le &-2\int |\nabla F(x,\tau)|^2 dg(x,t)\\ &+\lambda^2 \int e^{2\lambda d(x,x_0,t)}|\nabla d|^2 u^2 dg(x,t).\end{aligned}$$

By the definition of $F$ and $u$, this shows

$$\partial_\tau \int F^2(x,\tau)dg(x,t) \le \lambda^2 \int F(x,\tau)^2 dg(x,t).$$

Upon integration, we derive the following $L^2$ estimate

$$\int F^2(x,\tau)dg(x,t) \le e^{\lambda^2\tau}\int F^2(x,0)dg(x,0) = e^{\lambda^2\tau}\int f(x)^2 dg(x,0). \tag{7.2.14}$$

Recall that $u$ is a solution to (7.2.1). Therefore, by the mean value inequality (7.2.11), the following holds

$$u(x,\tau)^2 \le \frac{c(\kappa,D_0)}{\tau^{1+n/2}}\int_{\tau/2}^{\tau}\int_{B(x,\sqrt{|t|/2},-s)} u^2(z,s)dg(z,-s)ds.$$

By the definition of $F$ and $u$, it follows that

$$u(x,\tau)^2 \le \frac{c(\kappa,D_0)}{\tau^{1+n/2}}\int_{\tau/2}^{\tau}\int_{B(x,\sqrt{|t|/2},-s)} e^{-2\lambda d(z,x_0,-s)}F^2(z,s)dg(z,-s)ds.$$

In particular, this holds for $x = x_0$. In this case, for $z \in B(x_0,\sqrt{|t|/2},-s)$, there holds $d(z,x_0,-s) \le \sqrt{|t|/2}$. Therefore, by the assumption that $\lambda < 0$,

$$u(x_0,\tau)^2 \le \frac{c(\kappa,D_0)}{\tau^{1+n/2}}e^{-\lambda\sqrt{2|t|}}\int_{\tau/2}^{\tau}\int_{B(x_0,\sqrt{|t|/2},-s)} F^2(z,s)dg(z,-s)ds.$$

This combined with (7.2.14) shows that

$$u(x_0,\tau)^2 \le \frac{c(\kappa,D_0)}{\tau^{n/2}}e^{\lambda^2\tau-\lambda\sqrt{2|t|}}\int f(y)^2 dg(y,0).$$

i.e.

$$\begin{aligned}\left(\int G(x_0,\tau;z,0)e^{-\lambda d(z,x_0,0)}f(z)dg(z,0)\right)^2 \le &\frac{c(\kappa,D_0)}{\tau^{n/2}}e^{\lambda^2\tau-\lambda\sqrt{2|t|}}\\ &\times\int f(y)^2 dg(y,0).\end{aligned}$$

Now, we fix $y_0$ such that $d(y_0, x_0, 0)^2 \geq 4|t|$. Then it is clear that, by $\lambda < 0$ and the triangle inequality,

$$-\lambda d(z, x_0, 0) \geq -\frac{\lambda}{2} d(x_0, y_0, 0)$$

when $d(z, y_0, 0) \leq \sqrt{|t|}$. In this case, the above integral inequality implies

$$\left( \int_{B(y_0, \sqrt{|t|}, 0)} G(x_0, \tau; z, 0) f(z) dg(z, 0) \right)^2 \leq \frac{c(\kappa, D_0) e^{\lambda d(x_0, y_0, 0) + \lambda^2 \tau - \lambda \sqrt{2|t|}}}{\tau^{n/2}} \int f(y)^2 dg(y, 0).$$

Now we take, after freezing $t$ and $\tau$,

$$\lambda = -\frac{d(x_0, y_0, 0)}{\beta \tau}$$

with $\beta > 0$ sufficiently large. Since $f$ is arbitrary, this shows, for some $b > 0$,

$$\int_{B(y_0, \sqrt{|t|}, 0)} G^2(x_0, \tau; z, 0) dg(z, 0) \leq \frac{c(\kappa, D_0) e^{-b d(x_0, y_0, 0)^2 / \tau}}{\tau^{n/2}}.$$

Hence, there exists $z_0 \in B(y_0, \sqrt{|t|}, 0)$ such that

$$G^2(x_0, \tau; z_0, 0) \leq \frac{c(\kappa, D_0)}{\tau^{n/2} \ |B(x_0, \sqrt{|t|}, 0)|_0} e^{-b d(x_0, y_0, 0)^2 / \tau}.$$

In order to get the upper bound for all points, let us consider the function

$$v = v(z, l) \equiv G(x_0, \tau; z, l).$$

This is a solution to the conjugate of the conjugate equation (7.2.1), i.e.

$$\Delta_z G(x, \tau; z; l) + \partial_l G(x, \tau; z, l) = 0, \quad \partial_l g = 2Ric.$$

Therefore, we can use Theorem 6.5.1, after a reversal in time. Note this theorem was stated only for compact manifolds. However, it is valid in the noncompact case whenever the maximum principle for the heat equation holds. This is the case for $\kappa$ solutions since the curvature is nonnegative and bounded.

Consequently, for $\delta > 0, C > 0$,

$$G(x_0, \tau; y_0, 0) \le C G^{1/(1+\delta)}(x_0, \tau, z_0, 0) M^{\delta/(1+\delta)},$$

where $M = \sup_{\mathbf{M}\times[0,\tau/2]} G(x_0, \tau, \cdot, \cdot)$. By Step 1, there exists a constant $c(\kappa, D_0) > 0$, such that

$$M \le \frac{c(\kappa, D_0)}{\tau^{n/2}}.$$

Consequently

$$\begin{aligned} G^2(x_0, \tau; y_0, 0) \le & \frac{c(\kappa, D_0)}{\tau^{n/2}|B(x_0, \sqrt{|t|}, 0)|_0} \ e^{-bd(x_0,y_0,0)^2/t} \\ \le & \frac{c(\kappa, D_0)}{|B(x_0, \sqrt{|t|}, 0)|_0^2} \ e^{-b \ d(x_0,y_0,0)^2/t}. \end{aligned}$$

The last step holds since the Ricci curvature is nonnegative.

Since $x_0$ and $y_0$ are arbitrary, the proof of part (i) is done.

*Step 3.* In this step, we prove the upper and lower bound for $G(x, \tau; x_0, \tau/2)$ in the case of type I ancient $\kappa$ solution. The upper bound is already proven in view the distance and volume comparison result (7.2.4), (7.2.5) and the fact that $|B(x, \sqrt{|t|}, t)|_t \ge c(\kappa, D_0)|t|^{n/2}$. So we just need to prove the lower bound.

For a number $\beta > 0$ to be fixed later, the upper bound implies

$$\begin{aligned} & \int_{B(x_0,\sqrt{\beta|t|},t)} G^2(x, \tau; x_0, \tau/2) dg(x,t) \\ & \ge \frac{1}{|B(x_0, \sqrt{\beta|t|}, t)|_t} \left( \int_{B(x_0,\sqrt{\beta|t|},t)} G(x, \tau; x_0, \tau/2) dg(x,t) \right)^2 \\ & = \frac{1}{|B(x_0, \sqrt{\beta|t|}, t)|_t} \left( 1 - \int_{B(x_0,\sqrt{\beta|t|},t)^c} G(x, \tau; x_0, \tau/2) dg(x,t) \right)^2 \\ & \ge \frac{1}{|B(x_0, \sqrt{\beta|t|}, t)|_t} \left( 1 - \int_{B(x_0,\sqrt{\beta|t|},t)^c} \frac{c(\kappa, D_0)}{\tau^{n/2}} \ e^{-b \ d(x_0,x,t)^2/t} dg(x,t) \right)^2. \end{aligned}$$

Since the Ricci curvature is nonnegative, one can use the volume-doubling property to compute that

$$\int_{B(x_0,\sqrt{\beta|t|},t)^c} \frac{c(\kappa, D_0)}{\tau^{n/2}} \ e^{-b \ d(x_0,x,t)^2/t} dg(x,t) \le 1/2$$

provided that $\beta$ is sufficiently large. Here we stress that all constants are independent of $t$. Since $|B(x_0, \sqrt{\beta|t|}, t)|_t \le c_n(\beta|t|)^{n/2}$ by standard volume comparison theorem, this shows

$$\int_{B(x_0,\sqrt{\beta|t|},t)} G^2(x,\tau;x_0,\tau/2)dg(x,t) \ge \frac{c(\kappa, D_0)}{|t|^{n/2}}.$$

Hence there exists $x_1 \in B(x_0, \sqrt{\beta|t|}, t)$ such that

$$G(x_1,\tau;x_0,\tau/2) \ge \frac{c(\kappa, D_0)}{|t|^{n/2}}.$$

An inspection of the proof shows that actually for any $\lambda \in [3/4, 4]$, it holds, for some $x_\lambda \in B(x_0, \sqrt{\beta|t|}, t)$,

$$G(x_\lambda,\lambda\tau;x_0,\tau/2) \ge \frac{c(\kappa, D_0)}{|t|^{n/2}}.$$

It is well known that such a lower bound implies the full Gaussian lower bound if one has a suitable Harnack inequality. Now we can apply the Harnack inequality for the heat kernel in Section 9 of [P1] (Corollary 6.1.1 here; see also Corollary 2.1 (a) in [KZ] and [CaH]),

$$\begin{aligned} &G(x_{3/4}, \frac{3}{4}\tau; x_0, \tau/2) \\ &\qquad \le G(x,\tau;x_0,\tau/2)\left(\frac{\tau}{\tau 3/4}\right)^n \exp\frac{\int_0^1 [\,4|\gamma'(s)|^2 + (\tau/4)^2\, R\,]\, ds}{2(\tau/4)}, \end{aligned}$$

where $\gamma$ is a smooth curve on $\mathbf{M}$ such that $\gamma(0) = x_{3/4}$ and $\gamma(1) = x$. Also $|\gamma'(s)|^2 = g_{(-l)}(\gamma'(s), \gamma'(s))$, and $l = 3\tau/4 + s\tau/4$.

This inequality together with the decay property of $R$ and compatibility of distances to conclude

$$G(x,\tau;x_0,\tau/2) \ge \frac{c(\kappa, D_0)}{|t|^{n/2}} e^{-b_1 d(x,x_0,t)^2/\tau}.$$

This finishes the proof of the theorem. □

## 7.3 Backward limits of $\kappa$ solutions

In this section we use Theorem 7.2.1 and the $W$ entropy associated with the heat kernel to give a different and shorter proof of Perelman's classification of backward limits of these ancient solutions. This part of the argument resembles that in [Cx] and [Se1] where forward convergence results for normalized Ricci flow were proven. The current section together with Chapter 8 and a different proof of universal noncollapsing due to Chen and Zhu [ChZ1] lead to a simplified proof of the Poincaré conjecture without using reduced distance and reduced volume. This point will be made clear in Chapter 9. The method seems to have the potential to work for high dimensional cases, especially for type I ancient $\kappa$ solutions.

**Theorem 7.3.1** *(Perelman [P1]) Let $(\mathbf{M}, g(\cdot, t))$ with $t \in (-\infty, 0]$ be a nonflat, 3 dimensional ancient $\kappa$ solution for some $\kappa > 0$. Then there exist sequences of points $\{q_k\} \subset \mathbf{M}$ and times $t_k \to -\infty$, $k = 1, 2, \ldots$, such that the scaled metrics*

$$g_k(x, s) \equiv R(q_k, t_k) g(x, t_k + sR^{-1}(q_k, t_k))$$

*around $q_k$ converge to a nonflat gradient shrinking soliton in $C^\infty_{loc}$ topology.*

Proof. We divide the proof into several cases.

Case 1 is when the section curvature is zero somewhere and $\mathbf{M}$ is noncompact. Then Hamilton's strong maximum principle for tensors show that the universal cover $\tilde{\mathbf{M}} = \mathbf{M}_2 \times \mathbf{R}$ where $\mathbf{M}_2$ is a 2 dimensional, nonflat ancient $\kappa$ solution. See p249 of [CLN] for a detailed explanation. According to Hamilton [Ha3], $\mathbf{M_2}$ is either $S^2$ or $RP^2$. Since $\tilde{\mathbf{M}}$ is simply connected, the only choice is $\mathbf{M}_2 = S^2$. So the theorem is already proven in this case. This case can also be covered in Case 4 below together.

Case 2 is when the section curvature is zero somewhere and $\mathbf{M}$ is compact.

Then, again using maximum principle, Hamilton (see Theorem 6.64 in [CLN] e.g.) showed that $\mathbf{M}$ is the metric quotient of $\mathbf{R^3}$ with the flat metric or that of $S^2 \times \mathbf{R}$. So the theorem is also proven in this case.

Case 3 is when the sectional curvature is positive everywhere and $\mathbf{M}$ is a type II ancient solution, i.e. $\sup_{t<0} |t|\ R(\cdot, t) = \infty$.

The situation is already understood before 2002 when Perelman's paper [P1] appeared. In fact Hamilton [Ha4] showed by a scaling argument and his matrix Harnack inequality (Theorem 5.3.3 here) that the backward limit is a steady gradient soliton. See also Proposition 9.29 in [CLN], in which a proof is given for the noncompact case. However the compact case can be proven in the same way with the $\kappa$ noncollapsing assumption. It is known that a suitable scaling limit of such a steady gradient soliton is a gradient shrinking soliton. See Theorem 9.66 in [CLN] e.g. Thus the theorem in this case is also proven.

The details, adapted from Proposition 9.29 in [CLN], go as follows. We choose a time sequence $T_i \to -\infty$ and positive numbers $\epsilon_i \to 0$. Let $(x_i, t_i) \in M \times [T_i, 0]$ be space time points such that

$$|t_i|(t_i - T_i)R(x_i, t_i) \geq (1 - \epsilon_i) \sup_{M \times [T_i, 0]} |t|(t - T_i)R(x, t). \tag{7.3.1}$$

Write

$$R_i \equiv R(x_i, t_i), \quad a_i \equiv (T_i - t_i)R_i, \quad b_i = -t_i R_i.$$

Observe that, when $i \to \infty$,

$$\begin{aligned} \frac{1}{-a_i^{-1} + b_i^{-1}} &= \frac{|t_i|(t_i - T_i)R_i}{|T_i|} \\ &\geq (1 - \epsilon_i)|T_i^{-1}| \sup_{M \times [T_i, 0]} |t|(t - T_i)R(x, t) \\ &\geq (1 - \epsilon_i)|T_i^{-1}| \sup_{M \times [T_i/2, 0]} |t|(t - T_i)R(x, t) \\ &\geq \frac{1 - \epsilon_i}{2} \sup_{M \times [T_i/2, 0]} |t|R(x, t) \to \infty. \end{aligned}$$

The last inequality is due to the assumption of type II ancient solution in this case. Hence

$$\lim_{i \to \infty} b_i = - \lim_{i \to \infty} a_i = \infty.$$

Next we consider the scaled metrics

$$g^{(i)}(\cdot, s) \equiv R_i g(\cdot, t_i + R_i^{-1} s)$$

which are defined on the time interval $(-\infty, b_i]$. For $s \in [a_i, b_i]$, we have, from (7.3.1),

$$R_{g^{(i)}}(\cdot, s) \leq \frac{a_i b_i}{(1 - \epsilon_i)(a_i - s)(b_i - s)}.$$

Here $R_{g^{(i)}}$ is the scalar curvature under $g^{(i)}$. In particular there holds

$$\limsup_{i \to \infty} R_{g^{(i)}}(\cdot, s) \leq 1$$

on any compact time interval. Note also

$$R_{g^{(i)}}(\cdot, 0) \le \frac{1}{1-\epsilon_i}.$$

By the $\kappa$ noncollapsing property, we know the injectivity radius at $s = 0$ is bounded below by a positive constant. Now we can apply Hamilton's compactness Theorem 5.3.5 to deduce that a marked (sub)sequence $(\mathbf{M}, g^{(i)}, x_i)$, converges in $C_0^\infty$ topology to a limit solution of the Ricci flow $(\mathbf{M}_\infty, g_\infty(s), x_\infty)$. The limit solution has scalar curvature $R_{g_\infty}$ bounded between 0 and 1. Moreover

$$R_{g_\infty}(x_\infty, 0) = \lim_{i\to\infty} R_{g^{(i)}}(x_i, 0) = 1.$$

By Theorem 5.3.4, it is well known that such $\mathbf{M}_\infty$ is a steady gradient soliton. According to Theorem 9.66 in [CLN], certain scaling limit of $\mathbf{M}_\infty$ is a gradient shrinking soliton. Hence the original flow $(\mathbf{M}, g(t))$ has a backward limit as a gradient shrinking soliton too.

If the ancient $\kappa$ solution arises from the blow up of finite time type II singularity, then Hamilton [Ha4] even proved that $\mathbf{M}$ is a steady gradient soliton. If $\mathbf{M}$ is compact, then it is well known that $\mathbf{M}$ is an Einstein manifold, i.e. $Ric = \lambda g$ for a constant $\lambda$. A proof of this statement can be found in Proposition 1.1.1 of [CZ] e.g. Since the curvature is positive, $\mathbf{M}$ has to be $S^3$.

So there is only one case left.

Case 4: $\mathbf{M}$ has positive sectional curvature and is of type I ancient solution.

In the special case that $\mathbf{M}$ is compact and is obtained as a scaling limit of a Type I maximal solution, N. Sesum already proved the theorem in this case [Se1]. Actually she proved a stronger result, namely, $\mathbf{M}$ is a gradient shrinking soliton. See also p 302 [CZ] and the work of X.D. Cao [Cx]. For the noncompact case, a similar result was proven by A. Naber [Nab]. However it is not clear if the noncompact gradient soliton is flat.

The following proof works in both compact and noncompact cases. If $\mathbf{M}$ is compact, we just consider the product flow $\mathbf{M} \times \mathbf{R}$.

By the $\kappa$ noncollapsing assumption and the bound $R(\cdot, t) \le \frac{D_0}{1+|t|}$, we can use Theorem 5.3.5 to find a sequence of time $t_k$ such that $\tau_k \equiv |t_k| \to \infty$ and that the following statement holds:

for any fixed point $x_0 \in \mathbf{M}$, the marked manifolds $(\mathbf{M}, g_k, x_0)$ with the metric

$$g_k \equiv \tau_k^{-1} g(\cdot, -s\tau_k)$$

converge, in $C^\infty_0$ sense, to a marked manifold $(\mathbf{M}_\infty, g_\infty(\cdot, s), y_\infty)$. Here $s > 0$.

We aim to prove that $g_\infty$ is a gradient, shrinking Ricci soliton. Note that we are scaling by $\tau_k^{-1}$. But this is equivalent to scaling by the scalar curvatures. The reason is that we are dealing with type I $\kappa$ solution and the limit is nonflat. We define, for $x \in \mathbf{M}$ and $s \ge 1$, the functions

$$u_k = u_k(x, s) \equiv \tau_k^{n/2}\ G(x, s\tau_k; x_0, 0).$$

Here $G$ is the heat kernel of the conjugate heat equation and $x_0$ is a fixed point. We choose $y_k = x_0$ in the scaled metrics above. Here $n$ is taken as 3. However the proof is also valid for any $n \ge 3$. By Theorem 7.2.1 (actually (7.2.12) is sufficient), we know that

$$u_k(x, s) \le U_0 \tag{7.3.2}$$

uniformly for all $k = 1, 2, \ldots$, $x \in \mathbf{M}$ and $s$ in a compact interval. Here $U_0$ is a positive constant. Note that $u_k$ is a positive solution of the conjugate heat equation under the metric on $(\mathbf{M}, g_k(s))$ i.e.

$$\Delta_{g_k} u_k - R_{g_k} u_k - \partial_s u_k = 0.$$

We have seen that $u_k$ and $R_{g_k}$ are uniformly bounded on compact intervals of $s$ in $(0, \infty)$, and also the Ricci curvature is nonnegative and the curvature tensors are uniformly bounded. The standard parabolic theory shows that $u_k$ is Hölder continuous uniformly with respect to $g_k$. Hence we can extract a subsequence, still called $\{u_k\}$, which converges in $C^\alpha_{loc}$ sense, modulo diffeomorphism, to a $C^\alpha_{loc}$ function $u_\infty$ on $(\mathbf{M}_\infty, g_\infty(s), y_\infty)$.

Using integration by parts, it is easy to see that $u_\infty$ is a weak solution of the conjugate heat equation on $(\mathbf{M}_\infty, g_\infty(s))$, i.e.

$$\int\int (u_\infty \Delta\phi - R_\infty u_\infty \phi + u_\infty \partial_s \phi)\, dg_\infty(s) ds = 0$$

for all $\phi \in C^\infty_0(\mathbf{M}_\infty \times (-\infty, 0])$. Here $R_\infty$ is the scalar curvature of the limit manifold. By standard parabolic theory, the function $u_\infty$, being bounded on compact time intervals, is a smooth solution of the conjugate heat equation on $(\mathbf{M}_\infty, g_\infty(s), y_\infty)$. We need to show that $u_\infty$ is not zero.

Let $u = u(x, \tau) = G(x, \tau; x_0, 0)$. We claim that for a constant $a > 0$ and all $\tau \ge 1$,

$$u(x_0, \tau) \ge \frac{a}{\tau^{n/2}}.$$

Here is the proof. Define $f$ by

$$(4\pi\tau)^{-n/2}e^{-f} = u.$$

By Corollary 9.4 in [P1] (Corollary 6.1.1 here), which is a consequence of his differential Harnack inequality for fundamental solutions (Proposition 6.1.3 here), we have, for $\tau = -t$,

$$-\partial_t f(x_0, t) \le \frac{1}{2}R(x_0, t) - \frac{1}{2\tau}f(x_0, t).$$

Since $R(x_0, t) \le c/\tau$,

$$\partial_t(\sqrt{\tau}f(x_0, t)) = \sqrt{\tau}\partial_t f(x_0, t) - \frac{1}{2\sqrt{\tau}}f(x_0, t) \ge -\frac{c}{2\sqrt{\tau}}.$$

We can integrate the above from $\tau = 1$ to get

$$f(x_0, \tau) \le c + \frac{f(x_0, 1)}{\tau} \le C.$$

Here we have used the fact that $f(x_0, 1)$ is bounded, by the standard short time bounds for $G = G(x_0, 1; x_0, 0)$. See [Gro] e.g. This proves the claim. By definition of $u_k$ as a scaling of $u$, we know that $u_k(x_0, s) \ge b > 0$ for $s \in [1, 4]$. Here $b$ is independent of $k$. Therefore $u_\infty(x_0, s) \ge b > 0$. The maximum principle shows $u_\infty$ is positive everywhere.

Let us recall that Perelman's $W$ entropy for each $u_k$ is

$$W_k(s) = W(g_k, u_k, s) = \int \left[s(|\nabla f_k|^2 + R_k) + f_k - n\right] u_k dg_k(s)$$

where $f_k$ is determined by the relation

$$(4\pi s)^{-n/2}e^{-f_k} = u_k;$$

and $R_k$ is the scalar curvature under $g_k$. By the uniform upper bound for $u_k$ in (7.3.2), we know that there exist $c_0 > 0$ such that

$$f_k = -\ln u_k - \frac{n}{2}\ln(4\pi s) \ge -c_0 \tag{7.3.3}$$

for all $k = 1, 2, \ldots$ and $s \in [1, 3]$. Here the choice of this interval for $s$ is just for convenience. Any finite time interval also works.

Since $\mathbf{M}$ is noncompact, one needs to justify the integral in $W_k(s)$ is finite. For fixed $k$, $u_k$ has a generic Gaussian upper and lower bound with coefficients depending on $\tau_k$ and curvature tensor and their derivatives, as shown in [Gro]. The manifold has nonnegative Ricci curvature and bounded curvature. So the term $f_k u_k$ which is essentially $-u_k \ln u_k$ is integrable. The term $|\nabla f_k|^2 u_k = |\nabla u_k|^2/u_k$ is also integrable. A proof is outlined as follows. First one proves the result at one fixed time level using an inequality similar to the one in Theorem 6.5.1. The result for the rest of time is done by using inequality (6.3.33) and the maximum principle. It is left as an exercise. One can also consult the paper [CTY] and the book [Cetc] for a proof. These together imply that $W_k(s)$ is well defined.

Since $\int_{\mathbf{M}} u_k dg_k = 1$, by (7.3.3) we know that

$$W_k(s) \geq -c_0 - n \tag{7.3.4}$$

for all $k = 1, 2 \ldots$ and $s \in [1, 3]$.

There is an alternative proof of the lower bound for $W_k$. Actually $W_k(s)$ is uniformly bounded from below if $u_k$ is replaced by any $v \in W^{1,2}$ such that $\|v\|_2 = 1$. This can be seen since $(\mathbf{M}, g_k(s), y_k)$, $s \in [1, 3]$ has uniformly bounded curvature operator and are $\kappa$ noncollapsed. Therefore, a uniform Sobolev inequality holds, which implies the lower bound of $W_k(s)$. The later is nothing but a lower bound on the best constants of log Sobolev inequalities.

By scaling it is easy to see that

$$W_k(s) = W(g, u, s\tau_k),$$

where $u = u(x, l) = G(x, l, x_0, 0)$. According to [P1],

$$\frac{dW_k(s)}{ds} = -2s \int |Ric_{g_k} + Hess_{g_k} f_k - \frac{1}{2s} g_k|^2 u_k dg_k(s) \leq 0. \tag{7.3.5}$$

Note that the integral on the right-hand side is finite by a similar argument as in the case of $W_k(s)$. So, for a fixed $s$, $W_k(s) = W(g, u, s\tau_k)$ is a nonincreasing function of $k$. Using the lower bound on $W_k(s)$ (7.3.4), we can find a function $W_\infty(s)$ such that

$$\lim_{k \to \infty} W_k(s) = \lim_{k \to \infty} W(g, u, s\tau_k) = W_\infty(s).$$

Now we pick $s_0 \in [1, 2]$. Clearly we can find a subsequence $\{\tau_{n_k}\}$, tending to infinity, such that

$$W(g, u, s_0\tau_{n_k}) \geq W(g, u, (s_0 + 1)\tau_{n_k}) \geq W(g, u, s_0\tau_{n_{k+1}}).$$

Since

$$\lim_{k\to\infty} W(g,u,s_0\tau_{n_k}) = \lim_{k\to\infty} W(g,u,s_0\tau_{n_{k+1}}) = W_\infty(s_0),$$

we know that

$$\lim_{k\to\infty}[W(g,u,s_0\tau_{n_k}) - W(g,u,(s_0+1)\tau_{n_k})] = 0.$$

That is

$$\lim_{k\to\infty}[W_{n_k}(s_0) - W_{n_k}(s_0+1)] = 0.$$

Integrating (7.3.5) from $s_0$ to $s_0+1$, we use the above to conclude that

$$\lim_{k\to\infty}\int_{s_0}^{s_0+1}\int s|Ric_{g_{n_k}} + Hess_{g_{n_k}} f_{n_k} - \frac{1}{2s} g_{n_k}|^2 u_{n_k} dg_{n_k}(s)ds = 0.$$

Therefore

$$Ric_\infty + Hess_\infty f_\infty - \frac{1}{2s} g_\infty = 0.$$

So the backward limit is a gradient shrinking Ricci soliton.

Finally we need to show the soliton is nonflat. We can assume the original ancient $\kappa$ solution is not a gradient shrinking soliton. Otherwise there is nothing to prove. Hence, we know that

$$W_k(s) < W_k(0) = W_0 = 0$$

where $W_0$ is the Euclidean $W$ entropy with respect to the standard Gaussian. By the upper bound of $u_k$, i.e. (7.3.2), we know that the integrand in $W_k(s)$, $s \in [1,3]$, is bounded from below by a negative constant. By replacing the constant $-n$ in the definition of $W_k(s)$ by a sufficiently large constant $a_0$, we can make the integrand nonnegative. This allows us to use Fatou's lemma. Applying Fatou's lemma on a sequence of exhausting domains, we find that

$$W(g_\infty, u_\infty, s) \le W_k(s) < W_0 = 0. \tag{7.3.6}$$

If the gradient shrinking soliton $g_\infty$ is flat, it has to be $\mathbf{R^3}$. See [MT] e.g. Indeed, if $g_\infty$ is flat then

$$Hess_\infty f_\infty = \frac{1}{2s} g_\infty$$

The universal cover of $(\mathbf{M}_\infty, g_\infty)$ is isometric to $\mathbf{R^3}$. The lifting of $f$ to the universal cover, called $\tilde{f}$, then satisfies $\partial_i\partial_j\tilde{f} = \frac{1}{2s}\delta_{ij}$. Here all derivatives are Euclidean ones in a standard orthonormal system of $\mathbf{R^3}$. Therefore $\tilde{f} - \frac{1}{4s}|x|^2$ is an affine linear function. Since $\tilde{f}$ has a unique minimal point, it is not invariant under any free action of a nontrivial group. Recall that a group $G$ acts on a set $X$ freely if $g(x) = x$ for some $x \in X$ and $g \in G$, then $g = e$. Hence $\mathbf{R^3}$, as the trivial universal cover, is just $\mathbf{M}_\infty$.

Note $\int_{\mathbf{M}_\infty} u_\infty \le 1$ by Fatou's lemma again. It is known that $W_\infty(s) \ge 0$ on the Euclidean space for such a $u_\infty$. Indeed, let $\hat{u} = u_\infty / \|u_\infty\|_1$. In the Euclidean space, the best constant of the log Sobolev inequality is achieved by the Gaussian. This means that the $W$ entropy associated with a Gaussian is 0. Therefore the $W$ entropy associated with $\hat{u}$ is nonnegative. Namely

$$W(g_\infty, \hat{u}, s) = \int [s \frac{|\nabla \hat{u}|^2}{\hat{u}} - \hat{u} \ln \hat{u} - \frac{n}{2}(\ln 4\pi s)\hat{u} - n\hat{u}]dx \ge 0.$$

Now, $u_\infty = \|u_\infty\|_1 \, \hat{u}$ and

$$W(g_\infty, u_\infty, s) = \int [s \frac{|\nabla u_\infty|^2}{u_\infty} - u_\infty \ln u_\infty - \frac{n}{2}(\ln 4\pi s) u_\infty - n u_\infty]dx.$$

Then

$$W(g_\infty, u_\infty, s) = \|u_\infty\|_1 W(g_\infty, \hat{u}, s) - \|u_\infty\|_1 \ln \|u_\infty\|_1 \ge 0.$$

We have reached a contradiction with (7.3.6), which implies $g_\infty$ is not flat. □

**Remark 7.3.1** *Case 4 with positive curvature tensor can also be dealt with by the method in [CL]. There Chow and Lu actually constructed an embedded region of the flow, which is close to $S^2 \times \mathbf{R}$. They even do not need to assume the soliton is $\kappa$ noncollapsed on all scales.*

## 7.4 Qualitative properties of $\kappa$ solutions

In this section we present a number of structural properties on $\kappa$ solutions, culminating in the canonical neighborhood theorem. All these results were proven by Perelman in [P1] and [P2]. Many proofs below follow the presentation in [Tao], which is based on [MT] Chapter 9.

The first result says that the volume of large balls in a $\kappa$ solution grows slower than the Euclidean ones. In other words, the asymptotic volume ratio is 0.

**Proposition 7.4.1** *Let* $(\mathbf{M}, g(t))$ *be a 3-dimensional* $\kappa$ *solution. Then for any time* $t$ *and* $p \in \mathbf{M}$,

$$\lim_{r\to\infty} \frac{|B(p,r,t)|_t}{r^3} = 0.$$

PROOF. By Perelman's classification result Proposition 5.4.2, all 3-dimensional gradient shrinking solitons satisfy the stated property. Using Theorem 7.3.1, for any $\epsilon > 0$, we can find arbitrarily negative times $t_k$, points $x_k$ and radii $r_k$ such that $|B(x_k, r_k, t_k)|/r_k^3 \le \epsilon$. Note the ratio is scaling invariant. According to the classical volume comparison Theorem 3.5.1, this ratio is a nonincreasing function of the radii. Therefore

$$\lim_{r\to\infty} |B(x_k, r, t_k)|_{t_k}/r^3 \le \epsilon,$$

which implies, via the triangle inequality

$$\lim_{r\to\infty} |B(p, r, t_k)|_{t_k}/r^3 \le \epsilon.$$

Fixing a time $t$, we pick one $t_k$ such that $t_k < t$. By Proposition 5.1.5 and the fact that sectional curvatures are bounded between 0 and a positive constants, we know that

$$d(p,x,t_k) - c|t - t_k| \le d(p,x,t) \le d(p,x,t_k).$$

Hence $B(p,r,t) \subset B(p, r + c|t - t_k|, t_k)$. This shows

$$|B(p,r,t)|_t \le |B(p, r + c|t-t_k|, t_k)|_t \le |B(p, r + c|t-t_k|, t_k)|_{t_k}.$$

Here we have used the property that the volume element is decreasing in time, due to the positivity of scalar curvature (Proposition 5.1.1, (2)). Therefore

$$\lim_{r\to\infty} \frac{|B(p,r,t)|_t}{r^3} \le \epsilon.$$

The proposition follows since $\epsilon$ is arbitrary. □

The next proposition claims that the scalar curvature decays slower than inverse square of the distance near infinity.

**Proposition 7.4.2** *Let* $(\mathbf{M}, g(t))$ *be a 3-dimensional, noncompact* $\kappa$ *solution. Then for any time* $t$ *and* $p \in \mathbf{M}$,

$$\limsup_{d(x,p,t)\to\infty} R(x,t) d^2(p,x,t) = \infty.$$

PROOF. Suppose for the sake of contradiction that the result is not true. Then there exists $p$ and $t$ such that $R(x,t) \le c/d(p,x,t)^2$ for all $x \in \mathbf{M}$ and some $c > 0$. Since $\partial_t R(x,t) \ge 0$, due to Hamilton's trace Harnack inequality for the scalar curvature, it holds $R(x,s) \le c/d(p,x,t)^2$ for all previous time $s$. From the $\kappa$ noncollapsing property, we know that $|B(x,d(p,x,t),t)| \ge c\kappa d(p,x,t)^3$. This contradicts with the previous proposition when $x$ is far from $p$. □

The last two propositions induce the next result, which is a sort of converse of the $\kappa$ noncollapsing property.

**Proposition 7.4.3** *(volume noncollapsing implies curvature bound) Let $(\mathbf{M}, g(t))$ be a 3-dimensional $\kappa$ solution. Suppose for a positive constant $b > 0$, it holds $|B(x_0,r,t)|_t \ge br^3$. Then for every $A > 0$, there exists a positive constant $C = C(b,A)$, depending only on $b, A$ such that*

$$R(x,t) \le C(b,A)r^{-2}, \quad x \in B(x_0, Ar, t).$$

PROOF. We take, without loss of generality $t = 0$. Note that $B(x_0,r,0) \subset B(x,(A+1)r,0)$ for each $x \in B(x_0,r,0)$, and hence $|B(x,(A+1)r,0)|_0 \ge br^3$. Thus we can just work on the ball $B(x,(A+1)r,0)$ from the very beginning. Therefore, without loss of generality we can just take $A = 1$ and prove the bound when $x = x_0$ in the statement of the proposition.

If the proposition were not true, then there exists a sequence of marked $\kappa$ solutions $(\mathbf{M}_k, g_k(t), x_k)$ and $r_k \to \infty$ such that

$$r_k^2 R_k(x_k,0) \to \infty, \quad |B(x_k,r_k,g_k(0))|_{g_k(0)} \ge cbr_k^3. \tag{7.4.1}$$

Here $c$ is a positive, generic constant. The notation $B(x_k,l,g_k(0))$ stands for the ball of radius $l$ in the manifold $\mathbf{M}_k$, under the metric $g_k(0)$. Here $l = r_k$ or another positive number. The $\kappa$ solutions are allowed to have different noncollapsing constants.

Without loss of generality, we assume that $B(x_k,5r_k,g_k(0))$ is a proper subset of $\mathbf{M}_k$. By the point picking Lemma 7.1.1, there exist $y_k \in B(x_k,5r_k,g_k(0))$ and $\rho > 0$, depending on $k$, such that

$$\begin{aligned}
&\rho^2 R(y_k,0) \ge r_k^2 R(x_k,0) \to \infty;\\
&R(z,0) \le 2R(y_k,0), \quad \forall z \in B(y_k,\rho,g_k(0)) \subset B(x_k,5r_k,g_k(0)).
\end{aligned} \tag{7.4.2}$$

Since $B(y_k, \rho, g_k(0)) \subset B(x_k, 5r_k, g_k(0)) \subset B(y_k, 8r_k, g_k(0))$ and the sectional curvature is nonnegative, the classical volume comparison theorem shows

$$\begin{aligned}|B(y_k, \rho, g_k(0))|_{g_k(0)} &\geq \left(\frac{\rho}{8r_k}\right)^3 |B(y_k, 8r_k, g_k(0))|_{g_k(0)} \\ &\geq \left(\frac{\rho}{8r_k}\right)^3 |B(x_k, r_k, g_k(0))|_{g_k(0)}.\end{aligned}$$

From this the assumption (7.4.1) implies

$$|B(y_k, \rho, g_k(0))|_{g_k(0)} \geq cb\rho^3. \tag{7.4.3}$$

Now we introduce the scaled metric $\tilde{g}_k = R(y_k, 0)g_k$. Then the ball $B(y_k, \rho, g_k(0))$ under $\tilde{g}_k$ becomes $B(y_k, \rho_k, \tilde{g}_k(0))$ where, by the choice of $\rho$ and $r_k$ in (7.4.2),

$$\rho_k = R(y_k, 0)^{1/2}\rho \to \infty, \quad k \to \infty.$$

Let $\tilde{R}_k$ be the scalar curvature under $\tilde{g}_k$. Then by (7.4.2) and (7.4.3) we have

$$\tilde{R}_k(x, 0) \leq 2, \quad z \in B(y_k, \rho_k, \tilde{g}_k(0));$$

$$\frac{|B(y_k, \rho_k, \tilde{g}_k(0))|_{\tilde{g}_k(0)}}{\rho_k^3} = \frac{|B(y_k, \rho, g_k(0))|_{g_k(0)}}{\rho^3} \geq cb.$$

The same scalar curvature bound holds if we replace time 0 by a previous time $s$ in view of the monotonicity $\partial_t R_k \geq 0$, a result of Hamilton's trace Harnack inequality. Therefore we can apply Hamilton's compactness Theorem 5.3.5 to conclude that a subsequence of $(\mathbf{M}_k, \tilde{g}_k(t), y_k)$ converge in $C^\infty_{loc}$ topology to a nonflat $\kappa$ solution $(\mathbf{M}_\infty, \tilde{g}_\infty(t), y_\infty)$. But the asymptotic volume ratio for the limiting manifold at time 0 is positive, which contradicts Proposition 7.4.1. $\square$

**Remark 7.4.1** *It is important to notice that the constant $C(b, A)$ does not depend on the noncollapsing constant $\kappa$. This fact will be useful later in the section when we prove the universal noncollapsing of nonround $\kappa$ solutions.*

The following result shows that the scalar curvature is bounded at bounded distance. It can also be viewed as a Harnack inequality for the scalar curvature.

**Proposition 7.4.4** *(bounded curvature at bounded distance) Let* $(\mathbf{M}, g(t))$ *be a 3-dimensional* $\kappa$ *solution. Then for every* $A > 0$ *and* $x_0 \in \mathbf{M}$*, there exists a positive constant* $C = C(\kappa, A)$*, depending only on* $\kappa, A$ *such that*

$$R(x,t) \le C(\kappa, A) R(x_0, t), \quad x \in B(x_0, AR(x_0,t)^{-1/2}, t).$$

PROOF. Without loss of generality we take $t = 0$ again. Suppose the proposition were not true. Then there exists $A_0 > 0$ and a sequence of marked $\kappa$ solutions $(\mathbf{M}_k, g_k(t), x_k)$ and $y_k \in B(x_k, \rho_k, g_k(0))$ such that

$$R_k(y_k, 0)/R_k(x_k, 0) \to \infty. \tag{7.4.4}$$

Here and later in the proof, we take

$$\rho_k \equiv A_0 R_k(x_k, 0)^{-1/2}. \tag{7.4.5}$$

Then Proposition 7.4.3 tells us that

$$|B(x_k, \rho_k, g_k(0))|_{g_k(0)}/\rho_k^3 \to 0, \quad k \to \infty.$$

Recall

$$\lim_{\rho \to 0} |B(x_k, \rho, g_k(0))|_{g_k(0)}/\rho^3 = w_3 = 4\pi/3$$

which is the volume of the unit ball in $\mathbf{R^3}$. We can find $r_k = o(\rho_k)$ such that

$$|B(x_k, r_k, g_k(0))|_{g_k(0)}/r_k^3 = w_3/2. \tag{7.4.6}$$

After rescaling, we may take $r_k = 1$. For simplicity we still use $g_k$ to denote the scaled metric and $R_k$ to denote the scaled scalar curvature. We claim that

$$R_k(x_k, 0) \to 0, \quad k \to \infty.$$

Suppose the claim is not true; there is a subsequence, still denoted by $\{R_k(x_k, 0)\}$ and a positive constant $c$ such that, for all large $k$,

$$R_k(x_k, 0) \ge c.$$

Then (7.4.5) shows that $\rho_k$ is bounded. By Proposition 7.4.3 again, for any $A > 0$, we have $R_k(x, 0) \le C(A)$ for $x \in B(x_k, A, g_k(0))$. This implies $R_k(y_k, 0)$ is bounded since $d(y_k, x_k, 0) < \rho_k \le A$ for some large and fixed $A$. Hence $R_k(y_k, 0)/R_k(x_k, 0)$ is also bounded, which contradicts with (7.4.4). Hence the claim is true.

Since $R_k(x_k, 0) \to 0$ when $k \to \infty$, by Hamilton's compactness Theorem 5.3.5 again, we can find a limiting solution $(\mathbf{M}_\infty, g_\infty(t), x_\infty)$

with $R_\infty(x_\infty, 0) = 0$. Hamilton's strong maximum principle implies that $\mathbf{M}_\infty$ is flat. Since $\mathbf{M}_\infty$ is $\kappa$ noncollapsed on all scales, it must be $\mathbf{R^3}$. A proof of this well-known fact can be found on p301 of [Pet] e.g. Therefore $B_{g_\infty}(x_\infty, 1, 0) = w_3$, which contradicts with (7.4.6). □

An immediate consequence of this proposition is

**Theorem 7.4.1** *(compactness of $\kappa$ solutions) For any fixed $\kappa > 0$, the set of nonflat 3 dimensional ancient $\kappa$ solutions is compact modulo scaling in the following sense: for any sequence of such solutions and marking points in space time $(x_k, 0)$ with $R(x_k, 0) = 1$, one can extract a $C^\infty_{loc}$ converging subsequence whose limit is also an ancient $\kappa$ solution.*

**Corollary 7.4.1** *Let $(\mathbf{M}, g(t))$, $-\infty < t \le 0$ be a 3 dimensional ancient $\kappa$ solution for a fixed $\kappa > 0$. Then there exist positive increasing function $w : [0, \infty) \to [0, \infty)$ and a positive constant $\eta$, depending on $\kappa$ such that*

*(i) For every $x, y \in \mathbf{M}$ and $t \in (-\infty, 0]$, there holds*

$$R(x,t) \le R(y,t)\, w(R(y,t)d^2(x,y,t));$$

*(ii) For all $x \in \mathbf{M}$ and $t \in (-\infty, 0]$, there hold*

$$|\nabla R(x,t)| \le \eta R^{3/2}(x,t), \qquad |\partial_t R(x,t)| \le \eta R^2(x,t).$$

*(iii) Suppose for some $(y, t_0)$ in space time and a constant $\zeta > 0$ there holds*

$$\frac{|B(y, R(y,t_0)^{-1/2}, t_0)|_{t_0}}{R(y,t_0)^{-3/2}} \ge \zeta.$$

*Then there exist a positive function $Z$ depending only on $\zeta$ such that, for all $x \in \mathbf{M}$,*

$$R(x,t_0) \le R(y,t_0)\, Z(R(y,t_0)d^2(x,y,t_0)).$$

Proof. Statement (i) is just a reincarnation of Proposition 7.4.4. One just needs to choose $A = d(x_0, x, t)R(x_0, t)^{1/2}$ and shifting the time $t$.

Statement (ii) follows from a combination of (i), Hamilton's trace Harnack inequality Corollary 5.3.1 and Shi's local derivative estimates Theorem 5.3.2.

Part (iii) is just Proposition 7.4.3 with $r = R(y,t_0)^{-1/2}$ and $A = R(y,t_0)d^2(x,y,t_0)$. □

**Remark 7.4.2** *Parts (i) and (ii) of the corollary depend on the noncollapsing constant of the $\kappa$ solution. Interestingly, $\kappa$ solutions have a universal noncollapsing constant unless it is the metric quotients of the round $S^3$. This result will be stated and proven in Propositions 7.4.6 and 7.4.8 below. If $\mathbf{M}$ is round $S^3$ or its metric quotients, then statements (i) and (ii) obviously hold for a universal function $w$ and constant $\eta$. Therefore, the function $w$ and constant $\eta$ are universal for all ancient $\kappa$ solutions.*

Now we would like to describe the global topological structures of $\kappa$ solutions.

**Proposition 7.4.5** *(structure of noncompact $\kappa$ solutions) Let $(\mathbf{M}, g(t))$ be a noncompact 3-dimensional $\kappa$ solution for some $\kappa > 0$.*

*Case 1. If the sectional curvature is zero somewhere, then $\mathbf{M} = S^2 \times \mathbf{R}$ or one of its $Z_2$ quotients: $RP^2 \times \mathbf{R}$ or the twisted product of $S^2 \times \mathbf{R}$ where $Z_2$ flips both $S^2$ and $\mathbf{R}$;*

*Case 2. If the sectional curvature is positive everywhere, then for any $\epsilon > 0$, there exists a constant $C = C(\epsilon)$ depending only on $\epsilon$ and subset $M_\epsilon$ with the following properties:*

*(i). Every point outside of $M_\epsilon$ is a center of an $\epsilon$ neck and no point inside $M_\epsilon$ is a center of an $\epsilon$ neck.*

*(ii). $M_\epsilon$ is compact; there exists a point $x_0 \in \partial M_\epsilon$ such that*

$$diam(M_\epsilon, g(t)) \le C R(x_0, t)^{-1/2}, \quad C^{-1} \le R(x, 0)/R(x_0, t) \le C$$

*for all $x \in M_\epsilon$.*

*(iii). The ball $B(x_0, C R(x_0, 0)^{-1/2})$, which contains $M_\epsilon$, is diffeomorphic to a Euclidean 3 ball.*

Proof. Case 1. If the curvature operator has zero eigenvalue somewhere, then the universal cover of $\mathbf{M}$, called $\tilde{\mathbf{M}}$, splits into the product of a two dimensional $\kappa$ solution and $\mathbf{R}$. This again is a consequence of Hamilton's strong maximum principle Theorem 5.2.1. Theorem 7.1.3 tells us that $\tilde{\mathbf{M}} = S^2 \times \mathbf{R}$. Thus $\mathbf{M} = S^2 \times \mathbf{R}/\Gamma$, a metric quotient of the round cylinder. Note such $\mathbf{M}$ is noncompact automatically with no assumption needed, because if the metric quotient is compact, then it would not be $\kappa$ noncollapsed when $t$ is very negative. The reason is the following. When $t$ is sufficiently negative, the scalar curvature is $O(\frac{1}{|t|})$, the diameter and the volume of $\mathbf{M}$ is $O(\sqrt{|t|})$ and $O(|t|)$ respectively. Therefore $\mathbf{M}$ is noncompact.

Hence $\mathbf{M} = S^2 \times \mathbf{R}$ or one of its $Z_2$ quotients, as proven in Proposition 5.4.2.

Case 2. Now we assume the curvature operator is positive everywhere. We can just prove the result for $t = 0$. Observe that $M_\epsilon$ is nonempty. Otherwise $\mathbf{M}$, which would be covered entirely by $\epsilon$ necks, can not be diffeomorphic to $\mathbf{R^3}$, contradicting the Soul theorem. Next we prove $M_\epsilon$ is compact. Suppose for contradiction there exists a sequence $\{z_k\}$ going to infinity with respect to $g(0)$ such that $z_k$ are not centers of an $\epsilon$ neck. Pick a point $z_0 \in \mathbf{M}$. According to Corollary 7.4.1 (i), there exists a positive increasing function $w : [0, \infty) \to (0, \infty)$ depending on $\kappa$ such that

$$0 < R(z_0, 0) \le R(z_k, 0)\, w(R(z_k, 0) d^2(z_0, z_k, 0)).$$

But this implies

$$\lim_{k \to \infty} R(z_k, 0) d^2(z_0, z_k, 0) = \infty.$$

Because, otherwise a subsequence of $\{R(z_k, 0)\}$ tends to 0. This would force $R(z_0, 0)$ to be zero. Therefore we can apply Proposition 7.1.1 and Theorem 7.1.3 to conclude that $z_k$, with $k$ large, has a neighborhood $U_k$ which, after scaling, is $\epsilon$ close to the round cylinder $S^2 \times [-\epsilon^{-1}, \epsilon^{-1}]$ or one of its $Z_2$ quotients: $RP^2 \times \mathbf{R}$. Since $\mathbf{M}$ has positive sectional curvature, the Soul theorem claims that $\mathbf{M}$ is diffeomorphic to $\mathbf{R^3}$ and hence $\mathbf{M}$ does not contain an embedded $RP^2$ with trivial normal bundle. Consequently $U_k$ is $\epsilon$ close to the round cylinder $S^2 \times [-\epsilon^{-1}, \epsilon^{-1}]$, i.e. $z_k$ is a center of an $\epsilon$ neck. This contradicts the assumption on $z_k$. Thus $M_\epsilon$ is compact. This also implies that $\mathbf{M} - M_\epsilon$ is nonempty.

Pick a point $x_0 \in \partial M_\epsilon$, we need to show that

$$R(x_0, 0)^{1/2} diam M_\epsilon \le C = C(\epsilon) \tag{7.4.7}$$

which depends only on $\epsilon$. Suppose the claim is false. There exists a sequence of $\kappa$ solutions $(\mathbf{M}^k, g_k(t))$ with the following property: let $M_\epsilon^k$ be the set of points in $\mathbf{M}^k$, which is not a center of an $\epsilon$ neck. There exist $x_{k,0} \in \partial M_\epsilon^k$ such that

$$R_k(x_{k,0}, 0)^{1/2} diam M_\epsilon^k \to \infty, \qquad k \to \infty. \tag{7.4.8}$$

Note we are not assuming $\mathbf{M}^k$ has the same noncollapsing constant. So it is not possible to use the above compactness theorem to extract a convergent subsequence. To get around this obstacle, we notice that

$x_{k,0} \in \partial M^k_\epsilon$. Therefore there exists a point, arbitrarily close to $x_{k,0}$, which is a center of an $\epsilon$ neck. Hence $x_{k,0}$ is a center of a $2\epsilon$ neck. This shows that there exists a universal constant $\zeta > 0$, determined by the noncollapsing constant of $S^2 \times \mathbf{R}$ such that

$$|B_{g_k}(x_{k,0}, R_k(x_{k,0}, 0)^{-1/2}, 0)|_{g_k(0)} \geq \zeta R_k(x_{k,0}, 0)^{-3/2}. \qquad (7.4.9)$$

Now we can apply Corollary 7.4.1 (iii) to find a universal positive function $Z$ depending only on $\zeta$ such that, for all $x \in \mathbf{M}^k$,

$$R_k(x, 0) \leq R_k(x_{k,0}, 0)\, Z(R_k(x_{k,0}, 0) d_k^2(x_{k,0}, x, 0)). \qquad (7.4.10)$$

The bounds in (7.4.9) and (7.4.10) permit us to use Hamilton's compactness Theorem 5.3.5. Thus we can extract a subsequence of the marked manifolds $(\mathbf{M}^k, R_k(x_{k,0}, 0)g_k, x_{k,0})$, which converges in $C^\infty_{loc}$ topology to manifold $(\mathbf{M}_\infty, g_\infty, x_\infty)$. By (7.4.8), the limit manifold $\mathbf{M}_\infty$ contains a geodesic line. According to Toponogov splitting theorem (Theorem 7.1.1, part 1), $\mathbf{M}_\infty = N \times \mathbf{R}$. Since $\mathbf{M}_\infty$ is orientable and $N$ is a two dimensional $\kappa$ solution, we conclude as earlier in the proof that $N = S^2$. This contradicts with the assumption that $x_{k,0}$ is not the center of an $\epsilon$ neck. The bound in (7.4.7) is proven. Finally, the bounds for the ratio of scalar curvatures in (ii) follow from Corollary 7.4.1 (iii) and the diameter bound.

Statement (iii) of the proposition follows from the Soul theorem 7.1.2, which says that $\mathbf{M}$ is diffeomorphic to $\mathbf{R}^3$. □

**Remark 7.4.3** *The above proof is modeled on Section 3.2 of the paper of Chen and Zhu [ChZ1], where a certain more general 4-dimensional result is proven. The difference between this proof and the original one by Perelman is that the current proof does not use universal noncollapsing of noncompact $\kappa$ solutions. In fact the latter is a byproduct of the proposition.*

Let us state the result.

**Proposition 7.4.6** *(universal noncollapsing of noncompact $\kappa$ solutions) There exists a positive constant $\kappa_0$ such that every nonflat, 3 dimensional, noncompact ancient $\kappa$ solution for some $\kappa > 0$ is $\kappa_0$ noncollapsed on all scales.*

Proof. Without loss of generality we take $t = 0$. Let $r > 0$ and $x \in \mathbf{M}$. Assume $R \leq 1/r^2$ in $B(x, r, 0)$, we need to prove that

$$|B(x, r, 0)|_{g(0)} \geq \kappa_0 r^3$$

where $\kappa_0$ is some universal positive constant.

The proof uses the previous Proposition 7.4.5 in an essential way. For a fixed small $\epsilon > 0$, let $M_\epsilon$ be as in Proposition 7.4.5 and $\rho$ be the diameter of $M_\epsilon$. Pick $x \in M_\epsilon$. Since the sectional curvature is nonnegative, we can use the classical volume comparison theorem (Theorem 3.5.1) to conclude that

$$|B(x, \rho, 0)| \geq b|B(x, 4\rho, 0)|. \tag{7.4.11}$$

Here $b$ is a universal positive constant and the volume is measured by $g(0)$. By the triangle inequality, we can find a point $y \in B(x, 4\rho, 0)$ such that

$$B(y, \rho, 0) \subset B(x, 4\rho, 0) \cap N \tag{7.4.12}$$

where $N$ is the $\epsilon$ neck adjacent to $M_\epsilon$. According to Proposition 7.4.5, there exists a constant $c$ depending only on $\epsilon$ such that $c^{-1} \leq R(z, 0)\rho^2 \leq c$ for all $z \in M_\epsilon$. Since $N$ lies adjacent to $M_\epsilon$, we can find $c_1$ depending only on $\epsilon$ such that

$$c_1^{-1} \leq R(z, 0)\rho^2 \leq c_1$$

for all $z \in M_\epsilon \cup N$. Hence the ball $B(y, \rho, 0)$, as a subset of the $\epsilon$ neck $N$, satisfies

$$|B(y, \rho, 0)| \geq c_2 \rho^3$$

where $c_2$ is a positive universal constant determined by the noncollapsing constant of $S^2 \times \mathbf{R}$. Combining this inequality with (7.4.11) and (7.4.12), we have

$$|B(x, \rho, 0)| \geq b c_2 \rho^3,$$

which shows, by Theorem 3.5.1 again

$$|B(x, r, 0)| \geq c_3 r^3, \quad 0 < r \leq \rho. \tag{7.4.13}$$

Here $c_3$ is a universal positive constant.

Now let $r > 0$ and $x \in \mathbf{M}$ be such that $R(y, 0) \leq 1/r^2$ when $y \in B(x, r, 0)$. We need to find a suitable lower bound for $|B(x, r, 0)|$.

If $x \in M_\epsilon$, then $c^{-1} \leq R(x, 0)\rho^2 \leq c$ and $R(x, 0) \leq 1/r^2$. Here $c$ depends only on $\epsilon$ as stated in Proposition 7.4.5. Thus $r \leq \sqrt{c}\rho$. If $r \in [\rho, \sqrt{c}\rho]$, then

$$|B(x, r, 0)| \geq |B(x, \rho, 0)| \geq c_3 \rho^3 \geq c_3/c^{3/2}\ r^3.$$

If $r \leq \rho$, then (7.4.13) holds. Anyway, there is a universal positive constant $\kappa_0$ such that

$$|B(x,r,0)| \geq \kappa_0 r^3.$$

If $x \in \mathbf{M} - M_\epsilon$ then $x$ is at the center of one $\epsilon$ neck. Therefore there exists a universal positive constant $c_4$ such that

$$|B(x,R(x,0)^{-1/2},0)| \geq c_4 R(x,0)^{-3/2}.$$

The assumed inequality $R(x,0) \leq 1/r^2$ implies that $r \leq R(x,0)^{-1/2}$. The classical volume comparison Theorem 3.5.1 again shows that

$$\frac{|B(x,r,0)|}{r^3} \geq \frac{|B(x,R(x,0)^{-1/2},0)|}{R(x,0)^{-3/2}} \geq \kappa_0$$

for a universal positive constant $\kappa_0 > 0$. This proves the proposition for noncompact $\kappa$ solutions. □

Mirroring the noncompact case, we can state and prove a structure result for certain compact $\kappa$ solutions.

**Proposition 7.4.7** *(structure of compact $\kappa$ solutions containing an $\epsilon$ neck) Let $(\mathbf{M}, g(t))$ be a compact 3-dimensional $\kappa$ solution for some $\kappa > 0$. For any fixed $\epsilon > 0$ and fixed $t$, suppose $(\mathbf{M}, g(t))$ contains an $\epsilon$ neck. Then there exists a constant $C = C(\epsilon)$ depending only on $\epsilon$ and a subset $M_\epsilon \subset \mathbf{M}$ with the following properties:*

*(i) Every point outside of $M_\epsilon$ is a center of an $\epsilon$ neck and no point inside $M_\epsilon$ is a center of an $\epsilon$ neck.*

*(ii) $M_\epsilon = M_1 \cup M_2$, a disjoint union of two compact regions $M_1$ and $M_2$; there exists a point $x_i \in \partial M_i$ such that*

$$diam(M_i, g(t)) \leq CR(x_i,t)^{-1/2}, \quad C^{-1} \leq R(x,t)/R(x_i,t) \leq C$$

*for all $x \in M_i$, $i = 1, 2$.*

*(iii) $M_i$, $i = 1, 2$, is diffeomorphic to the unit ball $B^3$ in $\mathbf{R}^3$ or the punctured $RP^3$, i.e. $RP^3 - \bar{B}^3$.*

(i)

Proof. As usual we can take $t = 0$. Just like the noncompact case, we denote by $M_\epsilon$ the set of points in $\mathbf{M}$, which are not centers of an $\epsilon$ neck. First we show that $M_\epsilon$ is not empty. Suppose not. Then every point in $\mathbf{M}$ is at the center of an $\epsilon$ neck. Since $\mathbf{M}$ is compact, it is a $S^2$ fibration over $S^1$, which has infinite fundamental group. If

the sectional curvature of $\mathbf{M}$ is 0 somewhere, by Case 2 of the proof of Theorem 7.3.1, $\mathbf{M}$ would be the quotients of the flat $\mathbf{R^3}$ or that of $S^2 \times \mathbf{R}$. Since $\mathbf{M}$ is nonflat, it can not be the quotient of $\mathbf{R^3}$. As explained in the proof of Proposition 7.4.5 Case 1, compact quotients of $S^2 \times \mathbf{R}$ can not be $\kappa$ noncollapsed at scale $\sqrt{|t|}$ when $|t|$ is large. Hence $\mathbf{M}$ must have positive sectional curvature. Meyer's theorem says that the fundamental group is finite. This contradiction shows that $M_\epsilon$ is not empty. By the assumption that $\mathbf{M}$ contains an $\epsilon$ neck, we can write $M_\epsilon = M_1 \cup M_2$, a disjoint union separated by a union of $\epsilon$ necks.
(ii) We just need to prove the proposition for, say, $M_2$. Pick a point $x_0 \in \partial M_2$, we need to show that

$$R(x_0, 0)^{1/2} diam M_2 \le C = C(\epsilon) \tag{7.4.14}$$

which depends only on $\epsilon$. Suppose the claim is false. There exists a sequence of compact $\kappa$ solutions $(\mathbf{M}^k, g_k(t))$ with the following property:

let $M^k_\epsilon$ be the set of points in $\mathbf{M}^k$, which are not centers of an $\epsilon$ neck and $M^k_\epsilon = M^k_1 \cup M^k_2$ which is a disjoint union separated by a union of $\epsilon$ necks. There exist $x_{k,0} \in \partial M^k_2$ such that

$$R_k(x_{k,0}, 0)^{1/2} diam M^k_2 \to \infty, \qquad k \to \infty. \tag{7.4.15}$$

As in the noncompact case, the fact that $x_{k,0} \in \partial M^k_2$ implies the existence of a point, arbitrarily close to $x_{k,0}$, which is a center of an $\epsilon$ neck. Hence $x_{k,0}$ is a center of a $2\epsilon$ neck. This shows that there exists a universal constant $\zeta > 0$, determined by the noncollapsing constant of $S^2 \times \mathbf{R}$ such that

$$|B_{g_k}(x_{k,0}, R_k(x_{k,0}, 0)^{-1/2}, 0)|_{g_k(0)} \ge \zeta R_k(x_{k,0}, 0)^{-3/2}. \tag{7.4.16}$$

Now we can apply Corollary 7.4.1 (iii) to find a universal positive function $Z$ depending only on $\zeta$ such that, for all $x \in \mathbf{M}^k$,

$$R_k(x, 0) \le R_k(x_{k,0}, 0)\, Z(R_k(x_{k,0}, 0) d^2_k(x_{k,0}, x, 0)). \tag{7.4.17}$$

The bounds in (7.4.16) and (7.4.17) permit us to use Hamilton's compactness Theorem 5.3.5. Thus we can find a subsequence of the marked manifolds $(\mathbf{M}^k, R_k(x_{k,0}, 0) g_k, x_{k,0})$, which converges in $C^\infty_{loc}$ topology to a noncompact $\kappa$ solution $(\mathbf{M}_\infty, g_\infty, x_\infty)$.

There are a number of cases we have to handle.

Case 1: The limit manifold $\mathbf{M}_\infty$ contains a geodesic line, i.e. a distance minimizing geodesic without ends, which has infinite length.

According to Toponogov splitting theorem (Theorem 7.1.1, part 1), $\mathbf{M}_\infty = N \times \mathbf{R}$. Since $\mathbf{M}_\infty$ is orientable and $N$ is a two dimensional $\kappa$ solution, we conclude as earlier in the proof that $N = S^2$. This contradicts with the assumption that $x_{k,0}$ is not the center of an $\epsilon$ neck.

Case 2: $\mathbf{M}_\infty$ does not contain a geodesic line and $x_\infty$ is a center of $\epsilon/2$ neck.

Then when $k$ is sufficiently large, the point $x_{k,0}$ is a center of $\epsilon$ neck, which contradicts with the assumption on $x_{k,0} \in \partial M_2^k$.

Case 3: $\mathbf{M}_\infty$ does not contain a geodesic line and $x_\infty$ is not a center of $\epsilon/2$ neck.

Since $\mathbf{M}_\infty$ is noncompact, Proposition 7.4.5 shows that there exists a constant $C$ depending only on $\epsilon/2$ such that every point outside the ball $B(x_\infty, C)$ under $g_\infty(0)$ is a center of $\epsilon/2$ neck. Therefore, for sufficiently large $k$, every point in $B(x_{k,0}, 3C) - B(x_{k,0}, 2C)$ under $R_k(x_{k,0}, 0)g_k(0)$ is a center of a $\epsilon$ neck. Therefore the set $M_2^k$ has a uniformly finite diameter under the metric $R_k(x_{k,0}, 0)g_k(0)$. Since $\epsilon$ is fixed, this contradicts with (7.4.15).

The bound in (7.4.14) is proven since a contradiction is reached in all possible cases. Finally, the bounds for the ratio of scalar curvatures in (ii) follow from Corollary 7.4.1 (iii) and the diameter bound.

Finally the proof of statement (iii) is left as an exercise. □

**Exercise 7.4.1** *Prove statement (iii) of the proposition.*

**Proposition 7.4.8** *(universal noncollapsing of nonround compact $\kappa$ solutions) There exists a positive constant $\kappa_0$ such that every nonflat, 3 dimensional, compact ancient $\kappa$ solution for some $\kappa > 0$ is $\kappa_0$ noncollapsed on all scales unless it is a metric quotient of the round 3 sphere.*

Proof. Since we assume $\mathbf{M}$ is not round, the backward limit in time is a noncompact gradient soliton, which is either a cylinder $S^2 \times \mathbf{R}$ or one of its $Z_2$ quotients: $RP^2 \times \mathbf{R}$ or the twisted product $S^2 \tilde{\times} \mathbf{R}$ (c.f. Proposition 5.4.2). The reason is that Hamilton's rounding Theorem 5.2.7 would imply that $\mathbf{M}$ is round if the backward limit is compact, which must be round by Proposition 5.4.2. If the backward limit is $S^2 \times \mathbf{R}$, then clearly for any small $\epsilon > 0$ and sufficiently negative $t$, $(\mathbf{M}, g(t))$ contains a point $x$ which is a center of an $\epsilon$ neck. If the backward limit is $S^2 \tilde{\times} \mathbf{R}$, then for any small $\epsilon > 0$ and sufficiently negative $t$, $(\mathbf{M}, g(t))$ also contains a point $x$ which is a center of an $\epsilon$

neck. The reason is that $S^2 \tilde{\times} \mathbf{R}$, regarded as a half-line crossed $S^2$ with an $RP^2$ attached at an end, contains the asymptotic soliton $S^2 \times \mathbf{R}$ when rescaling about points whose distances, after rescaling, from the end $RP^2 \times \{0\}$, tend to infinity. We thank Bennett Chow for explanation on this. If the backward limit is $RP^2 \times \mathbf{R}$, $(\mathbf{M}, g(t))$ contains a region, which after scaling, is $\epsilon$ close to $RP^2 \times [-\epsilon^{-1}, \epsilon^{-1}]$. The later case is ruled out by topological considerations as follows (See [Tao] Section 17 e.g.). As shown at the beginning of the proof of the previous proposition, $\mathbf{M}$ has positive curvature. Theorem 5.2.7 shows $\mathbf{M}$ is diffeomorphic to a spherical space form that has finite fundamental group. However, a compact manifold with embedded $RP^2$ with trivial normal bundle has infinite fundamental group. The reason is that an embedded $RP^2$ with trivial normal bundle can not separate $\mathbf{M}$ since the Euler characteristic is 1. So a closed loop in $\mathbf{M}$ can have a nontrivial intersection number with such a projective plane, leading to a nontrivial homomorphism from $\pi_1(\mathbf{M})$ to $Z$. (The detail is left as an exercise.) Anyway, we have shown for sufficiently negative $t$, $(M, g(t))$ contains a point $x$ which is a center of an $\epsilon$ neck.

By the last proposition, under the metric $g(t)$, we can write

$$\mathbf{M} = M_1 \cup M_2 \cup N.$$

Here $M_1 \cup M_2 = M_\epsilon$ is again the set of points which are not centers of an $\epsilon$ neck, and every point in $N$ is a center of an $\epsilon$ neck. First we prove that $(\mathbf{M}, g(t))$ is also universal noncollapsed for sufficiently negative $t$. Then we show that $(\mathbf{M}, g(s))$ is universal noncollapsed for all time $s \in [t, 0]$.

Choose $\epsilon$ sufficiently small. We claim that there are universal positive constant $A$ so that the following Sobolev inequality holds for all smooth functions $v$ on $(\mathbf{M}, g(t))$:

$$\left( \int_{\mathbf{M}} v^6 d\mu(g(t)) \right)^{1/3} \le A \int_{\mathbf{M}} (4|\nabla v|^2 + Rv^2) d\mu(g(t)). \tag{7.4.18}$$

The idea of the proof for this inequality is the same as Proposition 8.2.1 where we show that a capped $\epsilon$ horn satisfies the same inequality. We only present part of the proof which differs from that proposition.

Pick a point $x_i \in M_i$, $i = 1, 2$. According to the last proposition, there is a universal constant $C_0 = C_0(\epsilon)$ such that the diameter of $(M_i, R(x_i, t) g(t))$ is bounded from above by $C_0$ and that the corresponding scalar curvature $R_i$ is bounded between $C_1^{-1}$ and $C_1$, another universal constant depending only on $\epsilon$. Moreover the sectional curvature is positive and the scaled volume $|B(x_i, C_0)| \geq \zeta$ for some universal positive constant $\zeta$. This is a result of the classical volume comparison theorem applied on the adjacent $\epsilon$ neck. This shows that the injectivity radius of $M_i$ is bounded below by a universal constant (Theorem 3.6.2). By [Heb1] (See also [Au]), there exist universal positive constant $A_1$ and $B_1$ such that

$$\left(\int_{M_i} v^6 d\mu(g_i(t))\right)^{1/3} \leq A_1 \int_{M_i} |\nabla v|^2 d\mu(g_i(t)) + B_1 \int_{M_i} v^2 d\mu(g_i(t))$$

for all $v \in C_0^\infty(M_i)$, $i = 1, 2$. Here $g_i(t) = R(x_i, t)g(t)$. Therefore

$$\begin{aligned}\left(\int_{M_i} v^6 d\mu(g_i(t))\right)^{1/3} \leq A_1 \int_{M_i} & 4|\nabla v|^2 d\mu(g_i(t)) \\ & + B_1 C_1 \int_{M_i} R_i v^2 d\mu(g_i(t))\end{aligned}$$

for all $v \in C_0^\infty(M_i)$, $i = 1, 2$. Here $R_i$ is the scalar curvature of $(M_i, g_i)$, which is bounded between $C_1^{-1}$ and $C_1$. Recovering the original metric $g(t)$, we see that (7.4.18) holds for all $v \in C_0^\infty(M_i)$ with $A = max\{A_1, B_1C_1\}$.

In the same manner as in the proof of Proposition 8.2.1, we know that (7.4.18) holds for all $v \in C^\infty(N)$. Using a simple cut-off function, we know that (7.4.18) holds for all $v \in C^\infty(\mathbf{M})$, with perhaps a larger $A$. Finally, by the last statement of Theorem 6.2.1 part (a), we know that (7.4.18) holds for all $(\mathbf{M}, g(s))$, $s \geq t$, as long as the Ricci flow is smooth. Applying (7.4.18) together with Theorem 4.1.2 on the balls in the Definition of $\kappa$ noncollapsing, we know that the $\kappa$ solution is universally noncollapsed at all scales. □

**Exercise 7.4.2** *Carry out the details of the proof using cut-off functions.*

Combining the last propositions, we arrive at

**Theorem 7.4.2** *(canonical neighborhoods of $\kappa$ solutions) Let $(\mathbf{M}, g(t))$ be a 3-dimensional $\kappa$ solution other than the round $RP^2 \times \mathbf{R}$. Then, for any $\epsilon > 0$, there exists a positive constant $\beta = \beta(\epsilon)$ with the following properties. Given any point $(x, t)$, there exists $r \in (0, \beta R(x,t)^{-1/2})$ and an open set $B$, satisfying $B(x, r, t) \subset B \subset B(x, 2r, t)$, and falling into one of the three types.*

*(i). $B$ is an evolving $\epsilon$ neck, i.e. after scaling by factor $R(x,t)$ and shifting time $t$ to zero, the region in space time*

$$\{(y,s)\,|\,y\in B, s\in[t-\epsilon^{-2}R(x,t)^{-1},t]\}$$

*is $\epsilon$ close to the subset the evolving standard round cylinder $S^2\times\mathbf{R}$, which at time zero, is $S^2\times[-\epsilon^{-1},\epsilon^{-1}]$ the with scalar curvature 1.*

*(ii) $B$ is an evolving $\epsilon$ cap, i.e. $B$ is an evolving $\epsilon$ neck outside some suitable compact set which is diffeomorphic to the standard three ball in $\mathbf{R}^3$ or the punctured $RP^3$.*

*(iii) $B$ and therefore $\mathbf{M}$, is a compact manifold of positive sectional curvature.*

*Moreover, there exists a positive constant $C=C(\epsilon)$ such that for any $y\in B$, it holds*

$$C^{-1}R(x,t)\le R(y,t)\le CR(x,t);$$

*and the volume of $B$ in cases (i) and (ii) satisfies $(CR(x,t))^{-3/2}\le |B|_{g(t)}\le\epsilon r^3$.*

*(iv) The gradient bounds are valid for $x\in B$ and a universal constant $\eta>0$:*

$$|\nabla R(x,t)|\le\eta R^{3/2}(x,t),\ \ |\partial_t R(x,t)|\le\eta R^2(x,t).$$

## 7.5 Singularity analysis of 3-dimensional Ricci flow

The main result of this section is the next theorem which says that a space time cube in a 3 dimensional Ricci flow resembles a $\kappa$ solution provided that the scalar curvature at one point of the cube is sufficiently large. Intuitively, using Perelman's local noncollapsing Theorem 6.1.2, this result would follow easily by scaling the high curvature cube by the scalar curvature at the vertex of the cube and using Hamilton's compactness theorem. Indeed this is the case if the scalar curvature at the vertex of the cube is comparable with the maximum scalar curvature on the whole manifold at and before the moment. However, such simple scaling does not work if this extra assumption fails, i.e. the scalar curvature at the vertex of the cube could be very high but still much smaller than the scalar curvature at some other points in the cube. Thus the scaled curvature could still be unbounded, preventing the use of the compactness theorem. To overcome this difficulty, Perelman introduced an ingenious induction method which goes from regions of highest scalar curvature downward to all regions with sufficiently high scalar curvature.

Before stating the theorem, we introduce some notations to be used through out the proof. Let $(x,t)$ be a space time point in a Ricci flow $(\mathbf{M}, g(t))$ and $r>0$, we write

$$P(x,t,r) = \{(y,s) \,|\, d(x,y,g(t)) < r,\ t-r^2 \le s \le t\}.$$

During the proof, we will frequently switch metrics. Let $h$ be a metric, we will use the notation $B(x,r,h)$, etc., to denote the ball centered at $x$, with radius $r$ under the metric $h$, etc. Similarly $d(x,y,h)$ denotes the distance between $x$ and $y$ under the metric $h$.

**Theorem 7.5.1** *[P1] (singularity structure theorem or canonical neighborhood property) Let $(\mathbf{M}, g(t))$, $t \in [0,T_0)$, $T_0 > 1$ be a Ricci flow on a three dimensional compact, orientable manifold $\mathbf{M}$ with normalized initial metric. For any $\epsilon > 0$, there exists some $r_0 = r_0(\epsilon, T_0) > 0$ with the following property. Suppose $Q \equiv R(x_0,t_0) \ge r_0^{-2}$ at some arbitrary point $(x_0,t_0)$ where $x_0 \in \mathbf{M}$ and $t_0 \in (1,T_0)$. Then the flow in the region*

$$\{(x,t) \mid d^2(x_0,x,g(t_0)) < \epsilon^{-2}Q^{-1},\ t_0 - \epsilon^{-2}Q^{-1} \le t \le t_0\}$$

*is, after scaling by the factor $Q$, $\epsilon$ close to the corresponding region of some orientable $\kappa$ solution in the $C^{[\epsilon^{-1}]}$ topology.*

Proof. The proof is divided into several steps.

*Step 1.* Induction setup.

Suppose the theorem is false. Then for some $\epsilon > 0$, there exists a sequence of Ricci flows $(M_k, g_k(t))$ with normalized initial condition, defined on the time interval $[0,T_k)$ for some $T_k \in (1,T_0)$, which satisfies the following conditions.

(1) There exist sequences of positive numbers $r_k \to 0$, points $x_k \in M_k$, and times $t_k \in [1,T_k)$ such that

$$Q_k \equiv R_k(x_k,t_k) \ge r_k^{-2};$$

where $R_k$ is the scalar curvature with respect to $g_k$;

(2) For each $(M_k, g_k(t))$ and a constant $a(\epsilon) \in (0,\epsilon^2]$, the parabolic region

$$\begin{aligned} P(x_k,t_k,[a(\epsilon)Q_k]^{-1/2}) \equiv &\{(x,t) \in [0,T_k) \mid d^2(x_k,x,g_k(t_k)) \\ &< (a(\epsilon)Q_k)^{-1},\ t_k - (a(\epsilon)Q_k)^{-1} \le t \le t_k\} \end{aligned}$$

is not, after scaling by the factor $Q_k$, $\epsilon$ close to the corresponding subset of any orientable $\kappa$ solution.

(3) The conclusion of the theorem holds for any $(x,t) \in M_k \times [t_k - h_k Q_k^{-1}, t_k]$ if $R_k(x,t) \geq 2Q_k$. Here

$$h_k = 1/(2r_k)^2.$$

Only condition (3) needs a proof, which is by induction again. Fixing $k$, suppose a point $(x_{0,k}, t_{0,k})$, together with the scalar curvature $Q_{0,k} = R_k(x_{0,k}, t_{0,k})$, has been found to satisfy conditions (1), (2) but violate condition (3). Here by conditions (1)–(3), we mean to replace $(x_k, t_k)$ and $Q_k$ by $(x_{0,k}, t_{0,k})$ and $Q_{0,k}$ respectively. Then there exists a point

$$(x_{1,k}, t_{1,k}) \in M_k \times [t_{0,k} - h_k Q_{0,k}^{-1}, t_{0,k}],$$

which satisfies the inequality

$$Q_{1,k} = R_k(x_{1,k}, t_{1,k}) \geq 2Q_{0,k}.$$

However the flow $(M_k, g_k)$ in the cube $P(x_{1,k}, t_{1,k}, [a(\epsilon)Q_{1,k}]^{-1/2})$ is not, after scaling by the factor $Q_{1,k}$, $\epsilon$ close to the corresponding subset of any orientable $\kappa$ solution.

Notice that conditions (1)–(2) corresponding to the point $(x_{1,k}, t_{1,k})$ are also satisfied. If the point $(x_{1,k}, t_{1,k})$ satisfies condition (3) with $t_k$ replaced by $t_{1,k}$ and $Q_k$ replaced by $Q_{1,k}$, then it is our choice for $(x_k, t_k)$. Otherwise we rerun the induction. After each cycle of the induction we have picked a point in space time whose scalar curvature at least doubled the scalar curvature of the previous point. Since $M_k$ is compact, this induction must terminate in finite cycles when the scalar curvature of the chosen point, say $Q_{m,k} = R_k(x_{m,k}, t_{m,k})$, is at least a quarter of the maximum of the scalar curvature in $M_k \times [0, t_{1,k}]$. Note $t_{1,k} \geq t_{2,k} \geq \ldots$. Hence $Q_{m,k}$ is at least a quarter of the maximum of the scalar curvature of $(M_k, g_k(t))$, $t \leq t_{m,k}$. Now, an easy scaling argument described at the beginning of the section tells us that condition (3) is satisfied by $(x_{m,k}, t_{m,k})$, provided that $Q_{m,k} \geq r_k^{-2}$ with $r_k$ sufficiently small. We take this point as the desired $(x_k, t_k)$.

Having selected $(x_k, t_k)$ satisfying the induction conditions (1)–(3), let $(M_k, \tilde{g}_k(t), x_k, 0)$ be the scaled marked Ricci flow where

$$\tilde{g}_k(t) = Q_k g_k(tQ_k^{-1} + t_k).$$

For simplicity we still use $t$ to denote the scaled time. The goal is to prove that a subsequence of the rescaled Ricci flows converges in $C^\infty_{loc}$

topology to an orientable $\kappa$ solution, which will lead to a contradiction with the induction condition (2) for large $k$.

Before moving to the next step, we verify two claims on $\tilde{g}_k$.

Claim 1. $\tilde{g}_k(t), t \le 0$ has a uniform $\kappa$ noncollapsing constant and its noncollapsing scale tends to $\infty$ as $k \to \infty$.

The reason for this is that $g_k$ has normalized initial metric $g_k(0)$. Perelman's local noncollapsing theorem then shows that $g_k$ has a uniform $\kappa$ noncollapsing constant and scales depending only on $T_0$. But the noncollapsing constant $\kappa$ is scaling invariant. The noncollapsing scale for $\tilde{g}_k$ is $\sqrt{Q_k}$ times that of $g_k$. This proves Claim 1.

Claim 2. The negative sectional curvatures associated with $\tilde{g}_k$ at $(x_k, 0)$ tends to 0 when $k \to \infty$.

By Hamilton-Ivey pinching Theorem 5.2.4, since $Q_k = R_k(x_k, t_k) \to \infty$ when $k \to \infty$, the size of negative sectional curvatures associated with $g_k$ at $(x_k, t_k)$ is $o(Q_k)$. After scaling by $Q_k$, Claim 2 follows.

*Step 2.* Extension of curvature bound from one point to a neighborhood.

In this step, it is shown that if the scalar curvature $R_k$ is bounded at one point in space time, then this bound can be extended to a parabolic cube with vertex as the given point.

**Proposition 7.5.1** *For each $(y, s) \in M_k \times [t_k - (4r_k^2 Q_k)^{-1}, t_k]$ with $Q_k = R_k(x_k, t_k)$, there exists a universal constant $c_0 > 0$ such that*

$$R_k(x,t) \le 4\bar{Q}_k, \quad (x,t) \in P(y, s, (c_0 \bar{Q}_k^{-1})^{1/2})$$

*where $\bar{Q}_k = Q_k + R_k(y, s)$.*

Proof. (detail according to [KL] Section 70) Let $(x,t) \in P(y, s, (c_0 \bar{Q}_k^{-1})^{1/2})$. If $R_k(x,t) \le 2Q_k$, then the proof is done. So we assume $R_k(x,t) > 2Q_k$. We connect $(x,t)$ with $(y,s)$ by a curve $L$ in space time: $L$ connects $(x,t)$ with $(x,s)$ by a straight line and connects $(x,s)$ with $(y,s)$ be a minimizing geodesic with respect to $g_k(s)$. Let $(y_0, s_0)$ be the point on $L$ which is the closest one to $(x,t)$, at which the scalar curvature is $2Q_k$. If such a point does not exist, we take $(y_0, s_0) = (y, s)$. Denote by $L_1$ the segment of the curve $L$, which joins $(y_0, s_0)$ with $(x,t)$. The scalar curvature along $L_1$ is at least $2Q_k$. By the induction hypotheses (3), when $c_o$ is sufficiently small, the segment $L_1$ is contained in the union of parabolic cubes which, after scaling are $\epsilon$ close to portions of $\kappa$ solution. By the canonical neighborhood Theorem 7.4.2, we have the gradient bounds along $L_1$:

$$|\nabla R_k^{-1/2}| \leq \eta, \quad |\partial_t R_k^{-1}| \leq \eta.$$

Here $\eta$ is a universal constant and the inequalities are scaling invariant. The proposition is proven by integrating the gradient bounds along $L_1$. □

Note the proposition shows that $R_k(x,t) \leq 8R_k(x_k,t_k)$ if $(x,t) \in P(x_k, t_k, (c_o Q_k^{-1})^{1/2})$. Hence the scalar curvature $\tilde{R}_k$ of the rescaled solution $(M_k, \tilde{g}_k, x_k)$ are uniformly bounded in a neighborhood around the center, which has a fixed diameter. In the next step we will show that this diameter can be arbitrarily large.

*Step 3.* Bounded curvature at bounded distance from the center.

From now on we just work on the rescaled metric $\tilde{g}_k$. So the time $t = 0$ for $\tilde{g}_k$ corresponds to the time $t_k$ for $g_k$. All geometric quantities are relative to $\tilde{g}_k$ unless stated otherwise. For example $\tilde{R}_k$ will denote the scalar curvature with respect to $\tilde{g}_k$.

For all $\rho > 0$, define

$$\begin{aligned} J(\rho) &= \sup\{\tilde{R}_k(x,0) \mid k \geq 1, x \in B(x_k, \rho, \tilde{g}_k(0))\}, \\ \rho_0 &= \sup\{\rho \mid J(\rho) < \infty\}. \end{aligned}$$

By Step 2, we know that $\rho_0 > 0$. The aim is to show that $\rho_0 = \infty$.

Suppose $\rho_0 < \infty$. Then we know, after passing to a subsequence when necessary, there exists $y_k \in M_k$ such that $d(x_k, y_k, \tilde{g}_k(0)) \to \rho_0$ and $\tilde{R}_k(y_k, 0) \to \infty$. Let $\alpha_k$ be a minimizing geodesic from $x_k$ to $y_k$. Since $\tilde{R}_k(x_k, 0) = 1$, there exists a point $z_k \in \alpha_k$ such that $\tilde{R}_k(z_k, 0) = 2$ and that $z_k$ is the closest such point to $y_k$. We use $\beta_k$ to denote the segment of $\alpha_k$ that connects $z_k$ and $y_k$. Then the scalar curvature $\tilde{R}_k \geq 2$ along $\beta_k$. By Step 2, the length of $\beta_k$ is uniformly bounded away from zero. Hamilton-Ivey pinching Theorem 5.2.4 tells us that the curvature tensor is bounded from below. Therefore, for each fixed $\rho < \rho_0$, the curvature tensor of $M_k$ is uniformly bounded on the balls $B(x_k, \rho, \tilde{g}_k(0))$. The injectivity radii are also uniformly bounded away from zero due to Perelman's local noncollapsing theorem. Hence Hamilton's compactness theorem shows that the marked (sub)sequence $(B(x_k, \rho_0, \tilde{g}_k(0)), \tilde{g}_k(0), x_k)$ converges in the $C_0^\infty$ topology to a marked incomplete manifold $(B_\infty, \tilde{g}_\infty, x_\infty)$. The geodesic segment $\alpha_k$ converges to a geodesic segment $\alpha_\infty \subset B_\infty$ and $\beta_k$ to $\beta_\infty$. The shared end point of $\alpha_\infty$ and $\beta_\infty$ is denoted by $y_\infty$.

Observe that the scalar curvature $\tilde{R}_\infty$ along $\beta_\infty$ is at least 2. By the induction hypotheses (3), for any $q_0 \in \beta_\infty$, $\tilde{g}_\infty$ in i.e. the ball

$$D_\infty(q_0) \equiv \{q \in B_\infty \mid d^2(q_0, q, \tilde{g}_\infty) < \epsilon^{-2}[\tilde{R}_\infty(q_0)]^{-1}\}$$

is $2\epsilon$ close to the corresponding subset of some $\kappa$ solution. By Theorem 7.4.2, these corresponding subsets are either $2\epsilon$ necks, $2\epsilon$ caps or compact manifolds without boundary. As $\tilde{R}_\infty$ becomes unbounded when the curve $\alpha_\infty$ approaches the end point $y_\infty$, the region $D_\infty(q_0)$ is not close to a compact manifold. Since $\alpha_\infty$ is distance minimizing, we know that $D_\infty(q_0)$ can not be close to a $2\epsilon$ cap either. The reason is that any long geodesic going through the center of a cap can not be distance minimizing. So the only possibility is $D_\infty(q_0)$ is close to a $2\epsilon$ neck.

Therefore the limit manifold $(B_\infty, \tilde{g}_\infty, x_\infty)$, as a union of $2\epsilon$ necks, is diffeomorphic to $S^2 \times (0,1)$. The sectional curvature is nonnegative due to Hamilton-Ivey pinching Theorem 5.2.4 and the scalar curvature tends to infinity when a point in $B_\infty$ tends to the end point $y_\infty$.

Now we can cite a basic result in metric geometry, which claims that there exists a three dimensional tangent cone at $y_\infty$ which is a metric cone. Intuitively, if one magnifies the metric $\tilde{g}_\infty$ near $y_\infty$ suitably, then it looks increasingly like the metric of a cone.

**Proposition 7.5.2** *Let $\{\lambda_j\}$ be any sequence of positive numbers tending to infinity. There exists a sequence $\{y_j\} \subset B_\infty$ such that $d(y_j, y_\infty, \tilde{g}_\infty) = \lambda_j^{-1/2}$ and that the marked manifold $(B_\infty, \lambda_j \tilde{g}_\infty, y_j)$ converges in $C^\infty_{loc}$ topology to an open cone.*

One can see Chapter 10 of [BBI] and Chapter 10 of [MT] for a detailed proof.

Let's use $J(x_\infty)$ to denote the cone in the proposition and $g_J$ the metric. Let $V$ be the tangent vector field of a minimizing (radial) geodesic $\gamma$ ending at the vertex $y_\infty$. Then $Ric(V,V)|_{g_J} = 0$ as a property of the cone. On the other hand, for a fixed point $p \in \gamma$ other than $y_\infty$, there exists a constant $c > 0$ such that $(B(p, c, g_J), g_J)$ can be regarded as a time slice of an incomplete Ricci flow, which is the scaling limit of certain balls $B(z_k, c_k, \tilde{g}_k(t)) \subset M_k$ over some time interval. We observe that there are actually two limits involved. One is $\tilde{g}_\infty$ as a limit of $\tilde{g}_k$ and the other is $g_J$ as a scaled limit of $\tilde{g}_\infty$ near the vertex. But it is not hard to show that the limit of limit is a limit of a subsequence of the original sequence for $C^\infty_{loc}$ convergence.

Applying Hamilton's strong maximum principle, we know that $(B(p, c, g_J), g_J)$ splits locally as a product metric along the radial direction. But this is impossible for a nonflat cone. The contradiction shows that $\rho_0 = \infty$, i.e. $(M_k, \tilde{g}_k(0), x_k)$ has uniformly bounded scalar curvatures $\tilde{R}_k$ in the ball $B(x_k, A, \tilde{g}_k(0))$ for any $A > 0$. Let $B(A)$ denote the above scalar curvature bound. By Proposition 7.5.1 in Step 2, there exists some fixed number $\delta > 0$, depending only on $B(A)$, such that the scalar curvature $\tilde{R}_k$ are uniformly bounded in the space time region $B(x_k, A, \tilde{g}_k(0)) \times [-\delta, 0]$. By Shi's local derivative bounds (Theorem 5.3.2), the derivatives of the curvatures of $(M_k, \tilde{g}_k)$ in smaller regions are also bounded.

Recall that $\tilde{g}_k(t)$, $t \geq 0$, are uniformly $\kappa$ noncollapsed (Claim 1 at the end of step 1). Therefore we can use Hamilton's compactness Theorem 5.3.5 to obtain a $C^\infty_{loc}$ limit $(M_\infty, \tilde{g}_\infty, x_\infty)$. This is a Ricci flow existing on an open subset of $M_\infty \times (-\infty, 0]$, which contains the time slice $M_\infty \times \{0\}$. At this stage, the life span for the flow at a point in $M_\infty$ may tend to zero as the point tends to infinity. Nevertheless, we know that the following two nice properties hold for $(M_\infty, \tilde{g}_\infty(0), x_\infty)$:

Property (i) It is $\kappa$ noncollapsed at all scales;

Property (ii) It has nonnegative sectional curvature.

(i) Holds because $\tilde{g}_k$ is a blow up of $g_k$, i.e. a big constant multiple of $g_k$. This process also blows up the noncollapsing scales of $g_k$ (Claim 1 at the end of Step 1).

(ii) Is a result of Hamilton-Ivey pinching Theorem 5.2.4. Let $\nu_k(x, 0)$, $\tilde{\nu}_k(x, 0)$ be a negative sectional curvature under $g_k(0)$, $\tilde{g}_k(0)$ respectively, at the space time point $(x, 0)$. Fixing $x$, we just proved that $\tilde{R}_k(x, 0) \leq C$, a constant depending on $x$. Hence for the unscaled metric $g_k$, we have $R_k(x, t_k) \leq C R_k(x_k, t_k) \to \infty$ when $k \to \infty$. There are two cases to deal with. If $R_k(x, t_k) = O(R_k(x_k, t_k))$, then Hamilton-Ivey pinching tells us $|\nu_k(x, t_k)| = o(R_k(x, t_k))$ and hence $|\tilde{\nu}_k(x, 0)| = o(1)$. If $R_k(x, t_k) = o(R_k(x_k, t_k))$, then Hamilton-Ivey pinching also tells us $|\nu_k(x, t_k)| \leq c + R_k(x, t_k)$. So after scaling by $Q_k = R_k(x_k, t_k)$, we have $\tilde{\nu}_k(x, 0) \to 0$, proving Property (ii).

Our next task is to show that the life span of the limit flow at each point is bounded away from zero and the above properties (i) and (ii) hold during the life span. This is achieved by showing that the scalar curvature of $M_\infty$ under $\tilde{g}^{(0)}_\infty$ is bounded at all points and then applying Proposition 7.5.1.

*Step 4*. Bounded curvature at all points for $(M_\infty, \tilde{g}_\infty(0), x_\infty)$.

Now that we know the sectional curvature is nonnegative, we just need to prove $\tilde{R}_\infty$, the scalar curvature is bounded. The idea is to show that $M_\infty$ would contain an $\epsilon$ neck of arbitrarily small radius if the scalar curvature becomes unbounded. This would contradict Proposition 7.1.2.

Suppose $\tilde{R}_\infty$ is unbounded, then by a point picking result similar to Lemma 7.1.1, there exists a sequence of points $\{p_j\} \subset M_\infty$, diverging to $\infty$ such that

$$\begin{aligned} \tilde{R}_\infty(p_j, 0) \to \infty, \qquad \tilde{R}_\infty(x, 0) \le 4\tilde{R}_\infty(p_j, 0), \\ d(x, p_j, \tilde{g}_\infty(0)) \le j[\tilde{R}_\infty(p_j, 0)]^{-1/2}. \end{aligned} \tag{7.5.1}$$

By $\kappa$ noncollapsing and compactness theorem, a subsequence of the scaled marked manifolds $(M_k, \tilde{R}_\infty(p_j, 0)\tilde{g}_\infty(0), p_j)$ converges in the $C^\infty_{loc}$ topology to a smooth nonflat manifold. According to Proposition 7.1.1, this limit manifold is isometric to a metric product $N \times \mathbf{R}$ where $N$ is a 2 dimensional manifold.

On the other hand, for a sufficiently large $j$, $\tilde{R}_\infty(p_j, 0) \ge 4$. This means that for a sufficiently large $k$, there is a point $y_k \in M_k$ such that the unscaled scalar curvature $R_k(y_k, t_k) \ge 2Q_k$. So the induction condition (3) in Step 1 shows that the point $y_k$ has a canonical neighborhood which is either an $\epsilon$ neck, an $\epsilon$ cap or a compact manifold without boundary. By the definition of $(M_\infty, \tilde{g}_\infty)$, for any $r > 0$, the balls $B(y_k, r/Q_k, g_k(t_k))$, after scaling by $Q_k$, converges to $B(p_j, r, \tilde{g}_\infty(0))$ in $C^\infty_{loc}$ sense. Hence point $p_j$ has a canonical neighborhood which is either a $2\epsilon$ neck, a $2\epsilon$ cap or a compact manifold without boundary. By the last step, the canonical neighborhood must be a $2\epsilon$ neck of radius $[R_\infty(p_j, 0)]^{-1/2}$. The later converges to 0 when $j \to \infty$. This contradiction with Proposition 7.1.2 shows that $\tilde{R}_\infty(\cdot, 0)$ is bounded.

*Step 5.* Bounded curvature at all points $(M_\infty, \tilde{g}_\infty(t), x_\infty)$, $t \le 0$.

Using Proposition 7.5.1 again, we know that $(M_\infty, \tilde{g}_\infty(t), x_\infty)$ is defined and that $\tilde{R}_\infty(\cdot, t)$ is bounded in the time interval $(-a, 0]$ for some $a > 0$. We need to extend $a$ to $\infty$.

Suppose for contradiction that $(-a, 0]$ is the maximum time interval when the limiting solution is defined. Then Proposition 7.5.1 shows that the maximum of the scalar curvature $\tilde{R}_\infty(\cdot, t)$ must go to infinity as $t \to -a$. However the infimum of the scalar curvature, being nondecreasing in time, must remain bounded when $t \to -a$. We will use the method in the previous steps to show that this bound on the infimum can be extended to every point, leading to a contradiction.

Hamilton's trace Harnack inequality (Corollary 5.3.1): $\partial_t \tilde{R}_\infty + (\tilde{R}_\infty/(t + a)) \ge 0$ shows

$$\tilde{R}_\infty(x, t) \le Qa/(t + a)$$

where $Q$ is the maximum of $\tilde{R}_\infty(\cdot, 0)$.

By Proposition 5.1.5 (ii) with $r_0 = [Qa/(t+a)]^{-1/2}$, and the non-negativity of sectional curvature, we have, for a constant $c > 0$,

$$-c\sqrt{Qa/(t+a)} \le \partial_t d(x, y, \tilde{g}_\infty(t)) \le 0.$$

Here and until the end of the proof, the distance $d(x, y, \tilde{g}_\infty(t))$ is with respect to $\tilde{g}_\infty(t)$. After integration, we arrive at

$$d(x, y, \tilde{g}_\infty(0)) \le d(x, y, \tilde{g}_\infty(t)) \le d(x, y, \tilde{g}_\infty(0)) + ca\sqrt{Q} \tag{7.5.2}$$

for all $x, y \in M_\infty$ and $t \in (-a, 0]$.

Since $\tilde{R}_\infty(x_\infty, 0) = 1$ by construction, we know, for any fixed $h > 0$, there exists a point $y_\infty$ (depending on $h$) such that

$$\tilde{R}_\infty(y_\infty, -a+h) \le 1. \tag{7.5.3}$$

Otherwise the infimum of $\tilde{R}_\infty$ at time $-a+h$ would be no smaller than 1, contradicting with the fact that the infimum of scalar curvature is increasing unless the manifold is Ricci flat. The later claim is a consequence of the maximum principle which can be applied since the curvature at each time slice is bounded. In order to extend $\tilde{g}_\infty$ before the time $-a$, we go back to $(M_k, \tilde{g}_k)$, the sequence which defines $(M_\infty, \tilde{g}_\infty)$ as the limit and which are defined before $-a$.

From (7.5.3), when $k$ is sufficiently large, there exists $y_k \in M_k$ such that

$$\tilde{R}_k(y_k, -a+h) \le 2.$$

We fix $h \le c_0/2$, the constant in Proposition 7.5.1. When $k$ is sufficiently large, the time $-a+h$ falls into the range where Proposition 7.5.1 is in force. Hence

$$\tilde{R}_k(y_k, t) \le 12, \qquad t \in [-a-h, -a+h].$$

By repeating Step 3, we know that for any $A > 0$ and one small $\delta_0 \in (0, h]$, there exists a positive constant $C(A, \delta_0)$ such that

$$\tilde{R}_k(x, -a+\delta_0) \le C(A, \delta_0), \qquad d(x, y_k, \tilde{g}_k(-a+\delta_0)) \le A.$$

By Proposition 7.5.1, it holds

$$\tilde{R}_k(x, -a+\delta) \le 2C(A, \delta_0), \qquad d(x, y_k, \tilde{g}_k(-a+\delta_0)) \le A$$

for all $\delta \in [0, \delta_0]$. Taking the limit $k \to \infty$, we obtain, for all $\delta \in (0, \delta_0]$,

$$\tilde{R}_\infty(x, -a+\delta) \le 2C(A, \delta_0), \qquad d(x, y_\infty, \tilde{g}_\infty(-a+\delta_0)) \le A.$$

Since $d(y_\infty, x_\infty, \tilde{g}_\infty(-a+h))$ is fixed, by (7.5.2) we can find another positive constant $c = c(a, Q)$ such that

$$\tilde{R}_\infty(x, -a+\delta) \le cC(A, \delta_0), \qquad d(x, x_\infty, \tilde{g}_\infty(0)) \le A.$$

Hence, there holds, for a fixed $\delta \le \delta_0$, and sufficiently large $k$,

$$\tilde{R}_k(x, -a+\delta) \le 2cC(A, \delta_0), \qquad d(x, x_k, \tilde{g}_k(0)) \le A.$$

Choosing $\delta << cC(A, \delta_0)$, by Proposition 7.5.1 again, there exists $c_1 > 0$ such that

$$\tilde{R}_k(x,t) \le 8cC(A,\delta_0), \quad d(x, x_k, \tilde{g}_k(0)) \le A, \quad t \in [-a - c_1 C(A, \delta_0)^{-1}, 0].$$

Here we have used the fact that $\tilde{R}_k$ are uniformly bounded in a fixed compact set after the time $-a+\delta_0$, because the limiting metric would have to blow up at this moment otherwise.

Consequently, we know that a subsequence of $\{(M_k, \tilde{g}_k(t), x_k)\}$ converges, in $C^\infty_{loc}$ sense, to $(M_\infty, \tilde{g}_\infty(t), x_\infty)$ in an open subset of $M_\infty \times (-\infty, 0]$, which contains $M_\infty \times [-a, 0]$. Repeating Step 4 we know that $\tilde{R}_\infty$ is bounded on $M_\infty \times [-a, 0]$, which contradicts the assumption on $a$. Hence $(M_\infty, \tilde{g}_\infty(t), x_\infty)$ is a $\kappa$ solution. Therefore, when $k$ is large, the flow $(M_k, g_k)$ in the region $P(x_k, t_k, [a(\epsilon)Q_k]^{-1/2})$ is, after scaling by the factor $Q_k$, $\epsilon$ close to the corresponding region of some orientable $\kappa$ solution in the $C^{[\epsilon^{-1}]}$ topology. This statement contradicts the induction hypotheses (2) in Step 1. This proves the theorem. □

**Exercise 7.5.1** *Prove (7.5.1).*

**Definition 7.5.1** *The number $r$ in the above Theorem 7.5.1 is called the parameter of the canonical neighborhood property and $\epsilon$ is called the accuracy of the canonical neighborhood property.*

With the understanding of the singularity structure in a 3-dimensional Ricci flow, we can move to the next phase: Ricci flow with surgeries.

# Chapter 8

# Sobolev inequality and the Ricci flow, the case with surgeries

## 8.1 A brief description of the surgery process

At certain times, a Ricci flow may develop singularity at some parts of the manifold but stay smooth in other parts. In order to prolong the Ricci flow beyond this time, one needs to cut off certain regions with high curvature, which are replaced by cap shaped manifold. The resulting manifold serves as the initial data for a new Ricci flow. This process is called a surgery. Due to the canonical neighborhood Theorem 7.5.1, for 3 dimensional Ricci flows, regions with higher curvature have simple topological and geometrical structures. The most essential ones are $\epsilon$ horns where surgeries take place. During a surgery, the singular part of an $\epsilon$ horn is cut off along a central 2 sphere, which is then pasted to a surgery cap that is diffeomorphic to the Euclidean three ball.

Here we collect some basic facts concerning Ricci flow with surgery. Materials in the section are taken from Perelman's paper [P2], Hamilton's paper [Ha7], [CZ], [KL] and [MT], where more details can be found. All Ricci flows in this chapter have dimension 3, unless otherwise stated.

In the next definition we recall or introduce a few geometric objects which we have to deal with in doing surgeries.

**Definition 8.1.1** *($\epsilon$ neck, $\epsilon$ horn, double $\epsilon$ horn, $\epsilon$ tube, $\epsilon$ cap, and capped $\epsilon$ horn) An $\epsilon$ neck (of radius $r$) is an open set with a metric* *which is, after scaling the metric with factor* $r^{-2}$*,* $\epsilon$ *close, in the* $C^{[\epsilon^{-1}]}$ *topology, to the standard neck* $S^2 \times (-\epsilon^{-1}, \epsilon^{-1})$.

*Let* $I$ *be an open interval in* $\mathbf{R}$*. An* $\epsilon$ *horn (of radius* $r$*) is* $S^2 \times I$ *with a metric with the following properties: each point is contained in some* $\epsilon$ *neck; one end is contained in an* $\epsilon$ *neck of radius* $r$*; the scalar curvature tends to infinity at the other end.*

*An* $\epsilon$ *tube is* $S^2 \times I$ *with a metric such that each point is contained in some* $\epsilon$ *neck and the scalar curvature stays bounded on both ends.*

*A double* $\epsilon$ *horn is* $S^2 \times I$ *with a metric such that each point is contained in some* $\epsilon$ *neck and the scalar curvature tends to infinity at both ends.*

*An* $\epsilon$ *cap is the three ball* $B^3$ *or* $RP^3 - \bar{B}^3$ *equipped with a smooth (incomplete) metric, which is an* $\epsilon$ *neck outside a compact set.*

*A capped* $\epsilon$ *horn is the union of an* $\epsilon$ *cap and a* $\epsilon$ *horn, joined smoothly at the common boundary, which is a 2 sphere in an* $\epsilon$ *neck.*

**Definition 8.1.2** *(surgery) ( $(r,\delta)$ surgery with cut off radius* $h$*) Let* $(M, g(t))$*,* $t \in [T_0, T)$*, be a smooth Ricci flow which becomes singular at time* $T$*. Suppose the flow satisfies the canonical neighborhood property with parameter* $r$ *and accuracy* $\epsilon$ *(c.f. Definition 7.5.1). Given* $\delta \in (0, \epsilon)$*, a* $(r, \delta)$ *surgery with radish at time* $T$ *is a process of modifying the manifold and the metric at a* $\delta$ *neck described as follows.*

*Let* $\mathbf{N}$ *be a* $\delta$ *neck of radius* $h$*, which is a part of an* $\epsilon$ *horn whose one end has bounded curvature at time* $T$*. Recall that* $(\mathbf{N}, h^{-2}g)$ *is* $\delta$ *close in the* $C^{[\delta^{-1}]}$ *topology to the standard round neck* $S^2 \times (-\delta^{-1}, \delta^{-1})$ *of scalar curvature* 1*. Let* $\Pi$ *be the diffeomorphism from the standard round neck to* $\mathbf{N}$ *in the definition of* $\delta$ *closeness. Denote by* $z$ *for a number in* $(-\delta^{-1}, \delta^{-1})$*. For* $\theta \in S^2$*,* $(\theta, z)$ *is a parametrization of* $\mathbf{N}$ *via the diffeomorphism* $\Pi$*. We can identify the metric on* $\mathbf{N}$ *with its pull back on the round neck by* $\Pi$ *in this manner.*

*The metric* $\tilde{g} = \tilde{g}(T)$ *right after the surgery is given by*

$$\tilde{g} = \begin{cases} \bar{g}, & z \le 0, \\ e^{-2f}\bar{g}, & z \in [0,2], \\ \phi e^{-2f}\bar{g} + (1-\phi)e^{-2f}h^2 g_0, & z \in [2,3], \\ e^{-2f}h^2 g_0, & z \in [3,4]. \end{cases} \tag{8.1.1}$$

*Here* $\bar{g}$ *is the nonsingular part of the* $\lim_{t \to T^-} g(t)$*;* $g_0$ *is the standard product metric on the round neck* $S^2 \times \mathbf{R}$ *with scalar curvature* 1*; and*

*$f = f(z)$ is a smooth function given by*

$$\begin{aligned} & f(z) = 0,\ z \le 0; \\ & f(z) = q_0 e^{-p_0/z},\ z \in (0,3]; \\ & f''(z) > 0,\ z \in [3, 3.9]; \\ & f(z) = -\frac{1}{2}\ln(16 - z^2),\ z \in [3.9, 4]. \end{aligned} \tag{8.1.2}$$

*Here $q_0 > 0$ and $p_0 > 0$ are constants. $\phi$ is a smooth cut-off function with $\phi = 1$ for $z \le 2$ and $\phi = 0$ for $z \ge 3$.*

**Remark 8.1.1** *There are three parameters associated with a $(r, \delta)$ surgery. One is the accuracy parameter $\delta$. Another one, $r$, is the parameter in the canonical neighborhood property, i.e. if the scalar curvature at a point is greater than $1/r^2$, then the point has a neighborhood which is $\epsilon$ close to a corresponding region in a $\kappa$ solution. It can be regarded as the largest cross section radius of the $\epsilon$ horn inside which a cut is made. The third, called cut off radius $h$, is the radius of the $\delta$ neck (imbedded in a $\delta$ horn) where the cut takes place. In Proposition 8.1.1 below, it is shown that any $\epsilon$ horn with sufficiently small $\epsilon$ contains a $\delta$ neck which is "strong" when $\delta$ is sufficiently small.*

*The definition is modeled on that on p424 of [CZ], which is based on [Ha7].*

**Definition 8.1.3** *(standard capped infinite cylinders) A standard capped infinite cylinder is $\mathbf{R}^3$ equipped with a rotationally symmetric metric with nonnegative sectional curvature and positive scalar curvature such that outside a compact set it is a semi-infinite standard round cylinder $S^2 \times (-\infty, 0)$ whose scalar curvature is 1.*

*Let $g_0$ be the standard product metric of scalar curvature 1 on the semi-infinite cylinder $N_0 \equiv S^2 \times (-\infty, 4)$. Then $N_0$ equipped with the following metric $g_s$ is called the standard capped infinite cylinder.*

$$g_s = \begin{cases} g_0, & z \le 0, \\ e^{-2f} g_0, & z \in (0, 4]. \end{cases}$$

*Here $z$ and $f$ are the same as those in Definition 8.1.2.*

**Definition 8.1.4** *(standard solutions) The noncompact Ricci flow with a standard capped cylinder as the initial value, and with bounded curvature at each time slice, is called a standard solution.*

*The noncompact Ricci flow with the standard capped cylinder as the initial value, and with bounded curvature at each time slice, is called the standard solution.*

**Remark 8.1.2** *By Shi's Theorem 5.1.2, a standard solution exists in a finite time interval and is complete at each time level. See Lemma 8.1.3 below.*

Next we show by computation that the Hamilton-Ivey pinching Theorem 5.2.4 is preserved by the surgery provided that the parameters $q_0$ and $p_0$ are chosen correctly. This is important in understanding the singularity structure for the restarted Ricci flow. Recall that much of the understanding of singularities of 3-d Ricci flow relies on this fact: ancient solutions obtained by blowing up singularities have nonnegative sectional curvature, a consequence of Hamilton-Ivey pinching.

To simplify the computation, we need to use the following property for 3 manifolds:

**Lemma 8.1.1** *At each point of a 3 manifold, there is an orthonormal frame $\{e_1, e_2, e_3\}$ such that the curvature operator Rm is diagonal in the frame $\{e_1 \wedge e_2, e_3 \wedge e_1, e_2 \wedge e_3\}$. The corresponding eigenvalues are given, under the notations of Definition 3.2.2, by*

$$\lambda = 2 < Rm(e_1, e_2)e_2, e_1 >= 2R_{1221},$$
$$\mu = 2 < Rm(e_3, e_1)e_1, e_3 >= 2R_{3113},$$
$$\nu = 2 < Rm(e_2, e_3)e_3, e_2 >= 2R_{2332}.$$

The proof of this lemma is left as

**Exercise 8.1.1** *Give a proof of Lemma 8.1.1.*

**Lemma 8.1.2** *(Hamilton-Ivey pinching under surgeries) There are universal positive constants $\delta_0, q_0, p_0$ with the following properties. Suppose, at time $T$, a $(r, \delta)$ surgery with cut off radius $h$ has taken place (at a $\delta$ neck of radius $h$). Here $\delta \le \delta_0$, $h^2 \le 1/(2e^2 \ln(1+T))$ and $q_0$, $p_0$ are the numbers chosen in (8.1.2) to define the surgery. Then the following conclusions hold.*

*(i) Let $\tilde{R}$ be the scalar curvature of the metric $\tilde{g}$ and $\tilde{\nu}$ be the least eigenvalues of the curvature operator. Then*

$$\tilde{R} \ge (-\tilde{\nu})[\ln(-\tilde{\nu}) + \ln(1+T) - 3]$$

*when $\tilde{\nu} < 0$.*

*(ii) After scaling with $h^{-2}$, any metric ball $B(p, \delta^{-1/2}h, \tilde{g})$, centered in the surgery cap is $\delta$ close, in the $C^{[\delta^{-1/2}]}$ topology, to the corresponding ball of the standard capped infinite cylinder.*

PROOF. (i) The proof is by direct computation along the surgery cap.

*Step 1.* We compute $\tilde{R}_{ijkl}$ (the curvature tensor of the metric $\tilde{g} = e^{-2f}\bar{g}$) in the case $0 \le z \le 2$.

Let $\{e_1, e_2, e_3\}$ be an orthonormal basis for $\bar{g}$, centered at a point $x$. By direct computation using (3.2.4) and (3.1.3), we have

$$\begin{aligned}\tilde{R}_{abcd} = e^{-2f}\big[\bar{R}_{abcd} - |\bar{\nabla} f|^2(\bar{g}_{ad}\bar{g}_{bc} - \bar{g}_{ac}\bar{g}_{bd}) - (f_{ac} + f_a f_c)\bar{g}_{bd} \\ - (f_{bd} + f_b f_d)\bar{g}_{ac} + (f_{ad} + f_a f_d)\bar{g}_{bc} + (f_{bc} + f_b f_c)\bar{g}_{ad}\big].\end{aligned} \tag{8.1.3}$$

Here $\bar{\nabla} f$ is the gradient of $f$ with respect to $\bar{g}$; $f_a, \ldots$ are the components of $df$, and $f_{ac}, \ldots$ are the components of Hessian of $f$ under the local basis.

Note that

$$\{\hat{e}_a \equiv e^f e_a, \quad a = 1, 2, 3\}$$

is an orthonormal frame for $\tilde{g}$. Then

$$\hat{R}_{abcd} = \tilde{R}_{klmn} e^{4f} \delta_a^k \delta_b^l \delta_c^m \delta_d^n,$$

are the components of the curvature tensor under $\{\hat{e}_a, \quad a = 1, 2, 3\}$. By (8.1.3), the following formulas hold at $x$:

$$\begin{aligned}\hat{R}_{abcd} = e^{2f}\big[\bar{R}_{abcd} - |\bar{\nabla} f|^2(\delta_{ad}\delta_{bc} - \delta_{ac}\delta_{bd}) - (f_{ac} + f_a f_c)\delta_{bd} \\ - (f_{bd} + f_b f_d)\delta_{ac} + (f_{ad} + f_a f_d)\delta_{bc} + (f_{bc} + f_b f_c)\delta_{ad}\big],\end{aligned} \tag{8.1.4}$$

$$\tilde{R} = e^{2f}(\bar{R} + 4\bar{\Delta} f - 2|\bar{\nabla} f|^2). \tag{8.1.5}$$

Here $\bar{\Delta}$ is the Laplacian under $\bar{g}$.

By definition of $\delta$ necks, we know that the scaled metric $h^{-2}\bar{g}$ can be regarded as a metric defined on a segment of the cylinder $S^2 \times \mathbf{R}$. At a point $x$ in such a segment, let $\{e_1, e_2, e_3\}$ be an orthonormal basis with respect to $\bar{g}$, such that $\{e_1 \wedge e_2, e_3 \wedge e_1, e_2 \wedge e_3\}$ diagonalizes the curvature operator $\bar{R}m$, as specified by Lemma 8.1.1. We use $\bar{\lambda}$, $\bar{\mu}$ and $\bar{\nu}$ to denote the eigenvalues of $\bar{R}m$ under this basis, which are arranged in decreasing order.

Under the standard metric $g_0$ on $S^2 \times \mathbf{R}$, the curvature operator has eigenvalues $1/2, 0, 0$. Since $h^{-2}\bar{g}$ is $\delta$ close to $g_0$ in $C^{[\delta^{-1}]}$ topology, we have

$$|\bar{R}_{3113}| + |\bar{R}_{2332}| = O(\delta)h^{-2}, \qquad |\bar{R}_{1221} - \frac{1}{2h^2}| = O(\delta)h^{-2}. \tag{8.1.6}$$

We recall that the $z$ direction is the axis of the cylinder $S^2 \times \mathbf{R}$. So the $\delta$ closeness also allows us to arrange $e_i$, $i = 1, 2, 3$ in such a way that

$$|e_3 - h^{-1}\frac{\partial}{\partial z}|_{g_0} = O(\delta)h^{-1}, \quad |\bar\nabla_3 z - h^{-1}| = O(\delta)h^{-1},$$
$$|\bar\nabla_1 z| + |\bar\nabla_2 z| = O(\delta)h^{-1}, \quad |\bar\nabla^2_{a,b} z| = O(\delta)h^{-2}, \quad a, b = 1, 2, 3.$$

Here $\bar\nabla_a z$ means the covariant derivative $\bar\nabla_{e_a} z$, and $\bar\nabla^2_{a,b} z$ means the Hessian $\bar\nabla^2_{e_a,e_b}$, as specified in Definition 3.2.1. Note that

$$\bar\nabla_a f(z) = f'(z)\bar\nabla_a z, \quad \bar\nabla^2_{a,b} f(z) = f'(z)\bar\nabla^2_{a,b} z + f''(z)\bar\nabla_a z \bar\nabla_b z.$$

Observe also

$$f'(z) = q_0 e^{-p_0/z} p_0 z^{-2}, \quad f''(z) = q_0 e^{-p_0/z}(p_0^2 z^{-4} - 2p_0 z^{-3}).$$

Hence for any small $\theta > 0$, we can choose $q_0 > 0$ sufficiently small and $p_0 > 0$ sufficiently large such that

$$|e^{2f(z)} - 1| + |f'(z)| + |f'(z)|^2 < \theta f''(z), \qquad f''(z) < \theta, \quad z \in [0, 3]. \tag{8.1.7}$$

This shows

$$\begin{aligned} &|\bar\nabla_a f(z)| \le 2\theta h^{-1} f''(z), \qquad a = 1, 2, 3, \\ &|\bar\nabla^2_{a,b} f(z)| = O(\delta)h^{-2} f''(z), \quad \text{unless} \quad a = b = 3 \\ &|\bar\nabla^2_{3,3} f(z) - h^{-2} f''(z)| = O(\delta)h^{-2} f''(z). \end{aligned} \tag{8.1.8}$$

Combining (8.1.4), (8.1.5), (8.1.6) with (8.1.8), we arrive at the estimate: for sufficiently small $\theta, \delta > 0$,

$$\begin{aligned} &\hat R_{1221} = \bar R_{1221} - (O(\theta) + O(\delta))h^{-2} f''(z), \\ &\hat R_{3113} = \bar R_{3113} - (O(\theta) + O(\delta))h^{-2} f''(z) + h^{-2} f''(z), \\ &\hat R_{2332} = \bar R_{2332} - (O(\theta) + O(\delta))h^{-2} f''(z) + h^{-2} f''(z), \\ &\hat R_{abcd} = (O(\theta) + O(\delta))h^{-2} f''(z), \qquad \text{the rest of indices} \quad abcd \end{aligned} \tag{8.1.9}$$

Let $\tilde\lambda$, $\tilde\mu$ and $\tilde\nu$ be the eigenvalues of $\tilde Rm$, listed in decreasing order. Since $\{\hat e_a, a = 1, 2, 3\}$ is an orthonormal frame for $\tilde g$, we know from Remark 3.2.5 that

$$\tilde R = 2[\hat R_{1221} + \hat R_{3113} + \hat R_{2332}].$$

Hence

$$\tilde R \ge \bar R + [4 - (O(\theta) + O(\delta))]h^{-2} f''(z) \ge \bar R,$$

$$\tilde{\nu} \geq \bar{\nu} + [2 - (O(\theta) + O(\delta))]h^{-2} f''(z) \geq \bar{\nu}.$$

If $\tilde{\nu} \geq -e^2$, then the assumption that $h^{-2} \geq 2e^2 \ln(1+t)$ implies

$$\begin{aligned} \tilde{R} \geq \bar{R} &\geq \frac{1}{2} h^{-2} \geq e^2 \ln(1+t) \\ &\geq (-\tilde{\nu})[\ln(-\tilde{\nu}) + \ln(1+t) - 3]. \end{aligned}$$

Here we have used the property that the scalar curvature of $S^2 \times \mathbf{R}$ is 1 and $\bar{R}h^2$ is a perturbation of 1.

If $\tilde{\nu} < -e^2$, then the pinching property on $\bar{R}$ shows

$$\begin{aligned} \tilde{R} \geq \bar{R} &\geq (-\bar{\nu})[\ln(-\bar{\nu}) + \ln(1+t) - 3] \\ &\geq (-\tilde{\nu})[\ln(-\tilde{\nu}) + \ln(1+t) - 3]. \end{aligned}$$

Here we have used the fact that the function $x \ln x$ is increasing when $x > e$ and that $-\tilde{\nu} \leq -\bar{\nu}$. This shows the desired pinching holds in the zone $0 \leq z \leq 2$.

*Step 2.* The case $2 \leq z \leq 4$.

Since $h^{-2}\bar{g}$ is $\delta$ close, in $C^{[\delta^{-1}]}$ topology, to $g_0$, we can write

$$\tilde{g} = e^{-2f} h^2 g_0 + \phi h^2 e^{-2f} O(\delta).$$

In this case $f''(z)$ is bounded from below by a positive constant. When $q_0$ is sufficiently small and $p_0$ is sufficiently large, we know the curvature operator of the metric $e^{-2f} g_0$ is positive definite. This can be confirmed by direct calculation in the similar manner as in the previous step, which we leave as Exercise 8.1.2 below. Hence, when $\delta$ is sufficiently small, the curvature operator $\tilde{Rm}$ is also positive definite, and the pinching property holds automatically.

The last statement of the lemma is obvious from the definition of surgeries. □

**Exercise 8.1.2** *Carry out the calculation in Step 2 above, i.e. show that the curvature of $e^{-2f} g_0$ is positive definite when $z \in [2, 4]$, $q_0$ is sufficiently small and $p_0$ is sufficiently large.*

The following lemma tells us that the life span of any standard solution is $[0, 1)$ and the scalar curvature at every point tends to infinity when time approaches 1. Note that the life span for all standard solutions are the same even though their initial values are different in a compact set.

**Lemma 8.1.3** *A standard solution $g = g(t)$ satisfies the following properties.*

*(i) The curvature operator is nonnegative and the scalar curvature is positive everywhere during the time of existence.*

*(ii) The life span of the solution is the time interval $[0, 1)$.*

*(iii) The scalar curvature is bounded at each time slice and satisfies the lower bound:*

$$R(x, t) \geq C/(1 - t)$$

*for some constant $C$ depending only on the initial value.*

PROOF. (i) This is an immediate consequence of the strong maximum principle for Ricci flow Theorem 5.2.2, which is applicable since the curvature is uniformly bounded in each time slice.

(ii) Let $[0, T_s)$ be the life span of a standard solution. First we show that $T_s \leq 1$.

Suppose $T_s > 1$. Pick a sequence of points $x_i \in M$, $i = 1, 2, \ldots$, which tends to infinity in the time zero slice. Then by Hamilton's compactness Theorem5.3.5, there exists a subsequence of the marked manifold $(M, x_i, g(t))$, $t \in [0, 1)$, which converges, in $C^\infty_{loc}$, topology, to some marked Ricci flow $(M_\infty, x_\infty, g_\infty(t))$. The required injectivity radius lower bound in the cited theorem is the consequence of finite time $\kappa$ noncollapsing property of standard solutions. This property holds as in the case of compact Ricci flows since the curvature is bounded at each time slice.

Observe that $g_\infty(0)$ is the standard metric on $S^2 \times \mathbf{R}$. So the uniqueness of standard solutions ( [ChZ2], [LT]) implies that $(M_\infty, x_\infty, g_\infty(t))$ is the standard Ricci flow on $S^2 \times \mathbf{R}$, i.e. the shrinking cylinder. The later has life span $[0, 1)$ and the curvature blows up as $t \to 1$ uniformly. Hence there exists $x_i$ and $t_i \to 1$ such that the curvature of $(M, g)$ at $(x_i, t_i)$ is arbitrarily large. This shows $T_s \leq 1$.

Suppose $T_s < 1$. Then there exists a sequence $\{x_i\} \subset M$ and $t_i \to T_s$ such that $\lim_{i\to\infty} R(x_i, t_i) = \infty$. We observe that $d(x_i, x_1, 0)$ is uniformly bounded. Otherwise, argue as in the previous paragraph, there exists a subsequence, still denoted by $\{x_i\}$, such that the marked manifold $(M_\infty, x_\infty, g_\infty(t))$ converges to the shrinking cylinder in $C^\infty_{loc}$ topology. But the later does not blow up until $t = 1$ and the corresponding curvature is bounded by $C/(1 - T_s)$. Hence the curvature of $(M, g)$ at $(x_i, t_i)$ is bounded by $2C/(1 - T_s)$ when $i$ is large. This is not possible by the choice of $x_i$. Therefore $\{x_i\}$ is contained in a compact region measured by the initial metric. Whence we can apply the singularity structure Theorem 7.5.1. Note this theorem was stated only for compact Ricci flows. However as a local result, it obviously covers compact regions in a noncompact Ricci flow. Thus there exist $\epsilon_i \to 0$ and $i \to \infty$ such that the region

$$\{(x,t) \mid d^2(x_i, x, t_i) < \epsilon_i^{-2} Q_i^{-1},\ t_0 - \epsilon_i^{-2} Q_i^{-1} \le t \le t_0\}, \quad Q_i = R(x_i, t_i),$$

is, after scaling by the factor $Q_i$, $\epsilon_i$ close to the corresponding region of some orientable $\kappa$ solution in the $C^{[\epsilon_i^{-1}]}$ topology. By Proposition 7.4.1, the asymptotic volume ratio of the $\kappa$ solution is 0. Hence, for any small $\delta > 0$, there exists $A > 1$ such that

$$|B(x_i, (\sqrt{Q_i})^{-1} A, t_i)|_{g(t_i)} \, [(\sqrt{Q_i})^{-1} A]^{-3} < \delta.$$

Since the curvature is nonnegative, the classical volume comparison Theorem 3.5.1 shows, for any $r > (\sqrt{Q_i})^{-1} A$, it holds

$$|B(x_i, r, t_i)|_{g(t_i)} \, r^{-3} < \delta.$$

Since $Q_i \to \infty$, this implies, for any fixed $r > 0$,

$$\lim_{i \to \infty} |B(x_i, r, t_i)|_{g(t_i)} \, r^{-3} = 0.$$

Pick a fixed large $r$ and a point $y$ such that $d(x_1, y, 0) = r$ and

$$x_i \in B(y, r, 0).$$

Then we would have, by the classical volume comparison theorem,

$$\lim_{i \to \infty} |B(y, r/2, t_i)|_{g(t_i)} \, r^{-3} = 0.$$

This contradicts the finite time $\kappa$ noncollapsing property of standard solutions. This property is proven in exactly the same way as Theorem 6.1.2 in the compact case, since the curvature is bounded in each time slice. This contradiction shows $T_s = 1$. The proof of (ii) is done.

(iii) We claim that $\lim_{t \to 1^-} R(x,t) = \infty$ for any $x \in M$. Suppose not. Then there exist $\{x_i\} \subset M$ and $t_i \to 1^-$ such that $R(x_i, t_i) \le C_0 < \infty$, $i = 1, 2, 3, \ldots$. First we point out that $\{x_i\}$ must stay in $D$, a compact set under $g(0)$. For otherwise there is a subsequence, still called $\{x_i\}$, going to $\infty$, such that $R(x_i, t_i) \le C_0$. From the proof of (ii), for any $\delta > 0$ and $A > 0$, when $i$ is sufficiently large, the standard solution in $B(x_i, A, 0) \times [0, 1-\delta]$ is close to the corresponding region of the standard shrinking cylinder. The scalar curvature of the shrinking cylinder at time $t$ is $a/(1-t)$ for some $a > 0$. Hence $R(x_i, t_i) > 2a/(1-t_i)$, which contradicts with the assumption $R(x_i, t_i) \le C_0$. Therefore $\{x_i\}$ must stay in some compact set $D$. Thus there exists a point $z$ and a subsequence of $\{t_i\}$, still called $\{t_i\}$ such that $R(z, t_i) \le C_0$, $i = 1, 2, \ldots$. But the principle of bounded curvature at bounded distance (established in Step 3 of the proof of Theorem 7.5.1) would mean that $R(y, t_i)$ is bounded if $d(z, y, t_i)$ is bounded. Following Step 4 of the same theorem, we know that $R(x, 1)$ is actually bounded for all $x \in M$. So the standard solution would live beyond $t = 1$. This is a contradiction to part (ii). Therefore the claim is true.

As explained in the proof of (ii), the singularity structure Theorem 7.5.1 holds for standard solutions. So high curvature regions of the standard solution are sufficiently close to a portion of the $\kappa$ solution. By this proximity and Theorem 7.4.2 (iv), the gradient bounds hold for the standard solution. Let $x \in M$, $t \in [0, 1)$. Suppose $R(x, t)$ is sufficiently large, then there exists constant $\eta > 0$ such that

$$|\partial_t R(x, t)| \le \eta R^2.$$

Since $\lim_{t \to 1^-} R(x, t) = \infty$, we can integrate the above to deduce $R(x, t) > c/(1-t)$. □

The above proof is adopted from Section 61 of [KL]. One can also consult Section 7.4 in [CZ], and Chapter 12 of [MT] for very detailed proof.

**Lemma 8.1.4** *The standard solution* $(\mathbf{R}^3, g(t))$ *enjoys a canonical neighborhood property in the following sense.*

*For any sufficiently small* $\epsilon > 0$, *there is a positive constant* $C(\epsilon)$ *such that each point* $(x, t) \in \mathbf{R}^3 \times [0, 1)$ *has an open neighborhood* $B$, *with* $B(x, r, t) \subset B \subset B(x, 2r, t)$, $r \in (0, \quad C(\epsilon) R(x, t)^{-1/2})$, *which is one of the following two types:*

*(i).* $B$ *is an* $\epsilon$ *cap, or*

*(ii).* $B$ *is an* $\epsilon$ *neck.*

*Moreover,* $B$ *in case (ii) is the time* $t$ *slice of the parabolic neighborhood*

$$B(x, \epsilon^{-1} R(x, t)^{-1/2}, t) \times [t - min\{R(x, t)^{-1}, t\}, \quad t].$$

*The later is, after scaling by* $R(x, t)$ *and shifting time* $t$ *to zero,* $\epsilon$ *close in the* $C^{[\epsilon^{-1}]}$ *topology, to the corresponding subset of the evolving standard cylinder* $S^2 \times \mathbf{R}$ *over the time interval* $[-min\{tR(x, t), 1\}, \quad 0]$.

Proof. The proof is similar to that of Theorem 7.5.1. See [CZ] p427 e.g. □

**Exercise 8.1.3** *Prove Lemma 8.1.4.*

Now we define precisely the surgery process for a 3 dimensional Ricci flow.

Let $(M, g(t))$ be a Ricci flow on a compact manifold $M$, which is smooth on the time interval $[S, T)$, but which becomes singular at time $T$. We write

$$\Omega = \{x \in \mathbf{M} \mid \limsup_{t\to T^-} R(x,t) < \infty\}.$$

By Shi's local derivative estimate (Theorem 5.3.2), we know that $\lim_{t\to T^-} R(x,t)$ exists when $x \in \Omega$. Let $r$ be the parameter in the canonical neighborhood property for $g(t)$, $t \in [S,T)$. For some $\rho < r$, we write

$$\Omega_\rho = \{x \in \Omega \mid \lim_{t\to T^-} R(x,t) \le 1/\rho^2\}.$$

**Definition 8.1.5** *(surgery procedures) The surgery at the singular time $T$ is consisted of the following procedures.*

*(i) Perform $(r, \delta)$ surgery for all $\epsilon$ horns connected with $\Omega_\rho$.*

*(ii) Discard every compact component (without boundary) having positive sectional curvature.*

*(iii) Discard all capped horns and double horns lying in $\Omega - \Omega_\rho$.*

*(iv) Discard all compact components (without boundary) lying in $\Omega - \Omega_\rho$.*

**Remark 8.1.3** *The topology of the discarded parts are all well understood. The discarded components in procedure (ii) are diffeomorphic to $S^3$ or its metric quotient by the pioneering work of Hamilton [Ha1]. Since $\rho < r$, by the canonical neighborhood property, shortly before the singular time, those discarded items in (iv) are covered by canonical neighborhoods. If they are compact with positive sectional curvature, then they are thrown away by procedure (ii). Otherwise, they are covered entirely by $\epsilon$ necks or caps.*

*Let $N$ be one of these components. If $N$ contains a cap, then either there is a cap or a neck adjacent to it. In the former case, $N$ is diffeomorphic to $S^3$, $RP^3$ or the connected sum of two copies of $RP^3$. In the latter case, there is either a cap or a neck adjacent to the neck. This process must stop with the appearance of a cap eventually. Therefore $N$ again is diffeomorphic to $S^3$, $RP^3$ or the connected sum of two copies of $RP^3$. Recall we are looking at a time shortly before the singular time $T$.*

*If $N$ does not contain a cap, then it is covered entirely by necks. Since $N$ is smooth, compact and orientable, the necks must repeat. Thus $N$ is diffeomorphic to $S^2 \times S^1$.*

Next we introduce certain a priori conditions (also called a priori assumptions) on a Ricci flow $(M, g)$ with surgeries, which is hoped persist for a long time.

**Definition 8.1.6** *(a priori assumption of accuracy $\epsilon$ for Ricci flow with surgeries)*

*1. Pinching assumption. The eigenvalues $\lambda \geq \mu \geq \nu$ of the curvature operator and the scalar curvature $R$ at each space time point satisfy, if $\nu < 0$,*

$$R \geq -\nu[\ln(-\nu) + \ln(1+t) - 3].$$

*2. Strong canonical neighborhood property (assumptions) with parameter:(scale) $r$ and accuracy $\epsilon$.*

*For any fixed and sufficiently small $\epsilon > 0$, there exists a positive nonincreasing function of time $r = r(t)$ such that at each space time point $(x, t)$ with $R(x, t) \geq r^{-2}(t)$ has a neighborhood $B$, satisfying $B(x, \sigma, t) \subset B \subset B(x, 2\sigma, t)$ for some $\sigma \in (0, cR^{-1/2}(x, t))$, $c = c(\epsilon) > 0$, and which falls into one of the three types.*

*(i) $B$ is a strong (evolving) $\epsilon$ neck, i.e. after scaling by factor $R(x, t)$ and shifting time $t$ to zero, the region in space time*

$$\{(y, s) \,|\, y \in B, s \in [t - R(x, t)^{-1}, t]\}$$

*is $\epsilon$ close, in $C^{[\epsilon^{-1}]}_{loc}$ topology, to the subset of the evolving standard round cylinder $S^2 \times \mathbf{R}$ over the time interval $[-1, 0]$, which at time zero, is $S^2 \times [-\epsilon^{-1}, \epsilon^{-1}]$ with scalar curvature 1.*

*(ii) $B$ is an evolving $\epsilon$ cap, i.e. $B$ is an evolving $\epsilon$ neck outside some suitable compact set that is diffeomorphic to the standard three ball or the punctured $RP^3$.*

*(iii) $B$ and therefore $M$, is a compact manifold of positive sectional curvature.*

*(iv) Moreover, there exists a positive constant $C = C(\epsilon)$ such that for any $y \in B$, it holds*

$$C^{-1} R(x, t) \leq R(y, t) \leq C R(x, t);$$

*and the volume of $B$ in cases (i) and (ii) satisfies $(CR(x, t))^{-3/2} \leq |B|_{g(t)} \leq \epsilon\sigma^3$.*

*(v) Finally, the gradient bounds are valid for $x \in B$ and a universal constant $\eta > 0$:*

$$|\nabla R(x,t)| \le \eta R^{3/2}, \ |\partial_t R(x,t)| \le \eta R^2.$$

**Remark 8.1.4** *Note that an evolving $\epsilon$ cap is not required to have 1 as the scaled life span. The life span can be smaller. In contrast, a strong $\epsilon$ neck has a scaled life span 1. So there can not be a surgery that is "too close" to the vertex $(x,t)$ of the neck in the time direction. This property is desirable when one tries to rule out infinitely many surgeries in finite time.*

Perelman proved that any small $\epsilon$ horn contains a strong $\delta$ neck when $\delta$ is also sufficiently small.

**Proposition 8.1.1** *(Lemma 4.3 [P2]) Let $(M,g)$ be a Ricci flow with surgery, starting with a normalized initial metric, satisfying the a priori assumption with accuracy $\epsilon$, defined on the time interval $[0,T)$, going singular at time $T$. Let $r(T)$ be the scale in the strong canonical neighborhood property. Pick $\delta \in (0,1)$ and write $\rho = \delta r(T)$. Suppose $(x,T)$ lies in an $\epsilon$ horn whose boundary is contained in $\Omega_\rho$. Here $\epsilon < \epsilon_0$ which is sufficiently small. Then there exists $h \in (0, \delta\rho)$ depending only on $\delta$, $\epsilon_0$, $r(T)$ such that the following holds:*

*Suppose $R(x,T) \ge h^{-2}$. Then the parabolic region*

$$\begin{aligned}&P(x,T,\delta^{-1}R(x,T)^{-1/2}, T - R(x,t)^{-1}, T)\\ &\equiv \{(y,s)\,|\,d(x,y,s) < \delta^{-1}R(x,T)^{-1/2},\ s \in (T - R(x,t)^{-1}, T)\}\end{aligned}$$

*is a strong $\delta$ neck.*

Proof. The proof is by scaling and contradiction, which is similar to Steps 3–5 of Theorem 7.5.1. One can see Lemma 7.3.2 [CZ], Lemma 71.1 of [KL] and Theorem 11.29 [MT] for the same proof.

Suppose the proposition is not true. Then, for a fixed $\delta \in (0,\epsilon)$, there exists a sequence of Ricci flow with surgery $\{(M^k, g^k)\}$, $k = 1, 2, \ldots$, satisfying the conditions of the proposition, and points $x^k \in M^k$, lying inside an $\epsilon$ horn on $M^k$, and numbers $h(x^k) \to 0^+$, such that the parabolic regions

$$P^k(x^k, T, \delta^{-1}h(x^k), T - h^2(x^k), T), \quad k = 1, 2, \ldots$$

are not strong $\delta$ necks. Here $h^2(x^k)$ is the reciprocal of scalar curvature at $(x^k, T)$.

Consider the rescaled flow

$$\tilde{g}^k(\cdot, s) = h^{-2}(x^k) g^k(\cdot, h^2(x^k)s + T).$$

Since $h(x^k) \to 0$, each $x^k$ lying deep in the $\epsilon$ horn has a strong $\epsilon$ neck as its canonical neighborhood. This statement is a result of the a priori assumption. Therefore, for each $A > 0$ and large $k$, the ball $B(x^k, A, \tilde{g}^k(\cdot, 0))$ is unscathed by surgery in a uniform time interval $[-s_0(A), 0]$. Here the uniformity means $s_0(A)$ is independent of $k$. This observation tells us there is large enough region under the scaled metrics, where surgeries do not interfere.

Now we are in a situation where we can apply the argument in Step 3 of the proof of Theorem 7.5.1 (bounded curvature at bounded distance). Thus there exists $J(A) > 0$ such that

$$|Rm_{\tilde{g}^k}(y, 0)| \le J(A), \quad y \in B(x^k, A, \tilde{g}^k(\cdot, 0)).$$

Hence, a subsequence of $\tilde{g}^k$ converges, in $C^\infty_0$ topology, to a limit Ricci flow $(M^\infty, \tilde{g}^\infty(\cdot, s))$. This limit flow exists in a relatively open domain of $M^\infty \times (-\infty, 0]$, which contains the time slice $M^\infty \times \{0\}$.

The pinching assumption, as part of the a priori assumptions, shows that the limit flow has nonnegative curvature. Since $x^k$ is contained in a strong $\epsilon$ neck, we know the limit manifold has two ends. The Toponogov splitting theorem (Theorem 7.1.1, part 1) implies that the limit manifold is $M^\infty = N \times \mathbf{R}$ where $N$ is compact 2 manifold with positive curvature. Observe that $N$ is diffeomorphic to $S^2$ because a segment of $M^\infty$ is the limit of $\epsilon$ necks. Using the a priori assumptions, the limit flow $(M^\infty, \tilde{g}^\infty)$, as limit of scaled strong $\epsilon$ necks, exists in the time interval $[-1, 0]$. Now we can just follow step 5 in the proof of Theorem 7.5.1 to conclude that $(M^\infty, \tilde{g}^\infty)$ exists in the time interval $(-\infty, 0]$ and is a $\kappa$ solution. Thus $(N, g^\infty|_N)$ is a 2 dimensional $\kappa$ solution. Here $g^\infty|_N$ is the restriction of $g^\infty$ to $N$. Since $N$ is diffeomorphic to $S^2$, Theorem 7.1.3 shows $M^\infty = S^2 \times \mathbf{R}$.

Hence the region $P^k(x^k, T, \delta^{-1}h(x^k), T - h^2(x^k), T)$ is a strong $\delta$ neck when $k$ is large. This contradiction proves the proposition. □

## 8.2 $W$ entropy, Sobolev inequalities, little loop conjecture with surgeries

A crucial step in Perelman's work on Poincaré and Geometrization conjectures is the $\kappa$ noncollapsing result for Ricci flow with or without surgeries [P1] and [P2]. The proof of this result in the surgery case requires truly complicated calculation using such new concepts as reduced distance, admissible curve, barely admissible curve, gradient estimate of scalar curvature etc. This is elucidated in great length by Cao and Zhu [CZ], Kleiner and Lott [KL], Morgan and Tian [MT] and Tao [Tao].

In this section we present a result in [Z4]. There we prove a uniform Sobolev inequality for Ricci flow, which is independent of the number of surgeries. It is well known that uniform Sobolev inequalities are essential in that they encode rich analytical and geometrical information on the manifold. These include noncollapsing, isoperimetric inequalities, etc. As a consequence, a strong noncollapsing result is obtained. It includes Perelman's $\kappa$ noncollapsing with surgery as a special case. The result also requires less assumptions. For instance we do not need the canonical neighborhood property for the whole manifold (see Remark 8.2.1 below). In the proof, we use only Perelman's $W$ entropy and some analysis of the minimizer equation of the $W$ entropy on horn like manifolds. Hence it is shorter and seems more accessible. It also gives a proof of Hamilton's little loop conjecture in the surgery case, which was still open. The nonsurgery case was proved by Perelman as a result of his $\kappa$ noncollapsing theorem. See Remark 8.2.3 after the statement of the theorem in this section.

Let $\mathbf{M}$ be a compact Riemann manifold of dimension $n \geq 3$ and $g$ be the metric. Then a Sobolev inequality of the following form holds: there exist positive constants $A, B$ such that, for all $v \in W^{1,2}(\mathbf{M}, g)$,

$$\left( \int v^{2n/(n-2)} d\mu(g) \right)^{(n-2)/n} \leq A \int |\nabla v|^2 d\mu(g) + B \int v^2 d\mu(g). \quad (8.2.1)$$

This inequality was proven by Aubin [Au] for $A = K^2(n) + \epsilon$ with $\epsilon > 0$ and $B$ depending on bounds on the injectivity radius, sectional curvatures and $\epsilon$. Here $K(n)$ is the best constant in the Sobolev imbedding for $\mathbf{R}^n$. Hebey [Heb1] showed that $B$ can be chosen to depend only on $\epsilon$, the injectivity radius and the lower bound of the Ricci curvature. Hebey and Vaugon [HV] proved that one can even take $\epsilon = 0$. However the constant $B$ will also depend on the derivatives of the curvature tensor. Hence, the controlling geometric quantities for $B$ as stated above are not invariant under the Ricci flow in general. Theorem 8.2.1 below states that a Sobolev inequality of the above type holds uniformly under 3 dimensional Ricci flow in finite time, even in the presence of indefinite number of surgeries.

In order to state the theorem, we first introduce and recall some notations. We use $(\mathbf{M}, g(t))$ to denote Hamilton's Ricci flow, $\frac{dg}{dt} = -2Ric$. If a surgery occurs at time $t$, then $(\mathbf{M}, g(t^-))$ denotes the pre surgery manifold (the one right before the surgery); and $(\mathbf{M}, g(t^+))$ denotes the post surgery manifold (the one right after the surgery). As usual, the ball of radius $r$ with respect to the metric $g(t)$, centered at $x$, is denoted by $B(x, r, t)$. The scalar curvature is denoted by $R = R(x, t)$ and $R_0^- = \sup R^-(x, 0)$. $Rm$ denotes the full curvature tensor. $d\mu(g(t))$ denotes the volume element. $vol(\mathbf{M}(g(t))$ or $|\mathbf{M}(g(t))|$ is the total volume of $\mathbf{M}$ under $g(t)$. For a point $(x_0, t_0)$ in space time and $r > 0$, we define the parabolic ball

$$P(x_0, t_0, r, -r^2) = \{(x, t) \,|\, d(x_0, x, t) < r,\ t \in (t_0 - r^2, t_0)\}.$$

**Definition 8.2.1** *(scathed region) We say that a region in space time is scathed if a surgery cuts off some points of the region. Otherwise we say a region is unscathed.*

We recall the following definition of $\kappa$ noncollapsing by Perelman [P2], as elucidated in Definition 77.9 of [KL].

**Definition 8.2.2** *((weak) $\kappa$ noncollapsing with surgeries).*

*Let $(\mathbf{M}, g(t))$ be a 3 dimensional Ricci flow with surgery defined on the time interval $[a, b]$. Suppose that $x_0 \in \mathbf{M}$, $t_0 \in [a, b]$ and $r > 0$ are such that $t_0 - r^2 \ge a$, $B(x_0, r, t_0) \subset \mathbf{M}$ is a proper ball and the parabolic ball $P(x_0, t_0, r, -r^2)$ is unscathed. Then $\mathbf{M}$ is (weak) $\kappa$ noncollapsed or (weak) $\kappa$ noncollapsing at $(x_0, t_0)$ at scale $r$ if $|Rm| \le r^{-2}$ on $P(x_0, t_0, r, -r^2)$ and $|B(x_0, r, t_0)| \ge \kappa r^3$.*

Here we introduce

**Definition 8.2.3** *(strong $\kappa$ noncollapsing) Let $\mathbf{M}$ be a 3 dimensional Ricci flow with surgery defined on the time interval $[a, b]$. Suppose that $x_0 \in \mathbf{M}$, $t_0 \in [a, b]$ and $r > 0$ are such that $B(x_0, r, t_0) \subset \mathbf{M}$ is a proper ball. Then $\mathbf{M}$ is strong $\kappa$ noncollapsed or strong $\kappa$ noncollapsing at $(x_0, t_0)$ at scale $r$ if the scalar curvature satisfies $R \le r^{-2}$ on $B(x_0, r, t_0)$ and $|B(x_0, r, t_0)| \ge \kappa r^3$.*

This strong $\kappa$ noncollapsing improves the $\kappa$ noncollapsing on two aspects. One is that only information on the metric balls on one time level is needed. Thus it bypasses the complicated issue that a parabolic ball may be cut by a surgery. The other is that it only requires scalar curvature upper bound.

The main result of the section is

**Theorem 8.2.1** *Given real numbers $T_1 < T_2$, let $(\mathbf{M}, g(t))$ be a $n = 3$ dimensional Ricci flow with normalized initial metric defined on the time interval containing $[T_1, T_2]$. Suppose the following conditions are met.*

*(a) There are finitely many $(r, \delta)$ surgeries in $[T_1, T_2]$, occurring in $\epsilon$ horns of radii $r$. Here $r \le r_0$ and $\epsilon \le \epsilon_0$, with $r_0$ and $\epsilon_0$ being fixed sufficiently small positive numbers less than 1. The surgery radii are $h \le \delta^2 r$, i.e. the surgeries occur in $\delta$ necks of radius $h \le \delta^2 r$. Here $0 < \delta \le \delta_0$ where $\delta_0 = \delta_0(r_0, \epsilon_0) > 0$ is sufficiently small. Outside of the $\epsilon$ horns, the Ricci flow is smooth.*

*(b) For a constant $c > 0$ and any point $x$ in all the above $\epsilon$ horns, the following holds: There is a region $U$, satisfying, $B(x, c\epsilon^{-1} R^{-1/2}(x)) \subset U \subset B(x, 2c\epsilon^{-1} R^{-1/2}(x))$, such that, after scaling by a factor $R(x)$, it is $\epsilon$ close in the $C^{[\epsilon^{-1}]}$ topology to $S^2 \times (-\epsilon^{-1}, \epsilon^{-1})$.*

*Also for any $x$ in the modified part of the $\epsilon$ horn immediately after a surgery, the following holds: the ball $B(x, \epsilon^{-1} R^{-1/2}(x))$, is, after scaling by a factor $R(x)$, $\epsilon$ close in the $C^{[\epsilon^{-1}]}$ topology to the corresponding ball of the standard capped infinite cylinder.*

*(c) For $A_1 > 0$, the Sobolev imbedding for $n = 3$*

$$\left( \int v^{2n/(n-2)} d\mu(g(T_1)) \right)^{(n-2)/n} \le A_1 \int (4|\nabla v|^2 + Rv^2) d\mu(g(T_1)) + A_1 \int v^2 d\mu(g(T_1))$$

*holds for all $v \in W^{1,2}(\mathbf{M}, g(T_1))$.*

*Then for all $t \in (T_1, T_2]$, the Sobolev imbedding below holds for all $v \in W^{1,2}(\mathbf{M}, g(t))$.*

$$\left( \int v^{2n/(n-2)} d\mu(g(t)) \right)^{(n-2)/n} \le A_2 \int (4|\nabla v|^2 + Rv^2) d\mu(g(t)) + A_2 \int v^2 d\mu(g(t)).$$

*Here*

$$A_2 = C(A_1, \sup R^-(x, 0), T_2, T_1, \sup_{t \in [T_1, T_2]} Vol(\mathbf{M}(g(t))) \ )$$

*is independent of the number of surgeries or $r$.*

*Moreover, the Ricci flow is strong $\kappa$ noncollapsed in the whole interval $[T_1, T_2]$ under scale 1 where $\kappa$ depends only on $A_2$.*

*Finally, the surgery radius $h$ can be chosen as any $h \le z_0 r^p$ where $p$ is any number greater than one and $z_0 = z_0(\epsilon, p) > 0$ is sufficiently small.*

By the work of Hebey [Heb1], at any given time, a Sobolev imbedding always holds with constants depending on lower bound of Ricci curvature and injectivity radius. So one can replace assumption (c) by the assumption that $(\mathbf{M}, g(T_1))$ is $\kappa$ noncollapsed and that the canonical neighborhood assumption (with a fixed radius $r_0 > 0$ and $\epsilon_0 > 0$) at $T_1$ holds. It is easy to see that these together imply the Sobolev imbedding at time $T_1$.

We assume as usual that, at a surgery, we throw away all compact components with positive sectional curvature, and also capped horns, double horns and all compact components lying in the region where $R > r^{-2}$. In the extra assumption that the Ricci flow is smooth outside of the $\epsilon$ horns, we have excluded these deleted items. It is known that $vol(M(g(t)) \le C(1+t^{3/2})$. If one can choose the initial scalar curvature to be nonnegative everywhere, then $A_2$ can be chosen as a constant independent of the life span of Ricci flow. This is due to the fact that the volume does not increase with time and that the Sobolev constant is uniformly bounded in this case (c.f. Theorem 6.2.1).

**Remark 8.2.1** *With the exception of using the monotonicity of Perelman's $W$ entropy, the proof of Theorem 8.2.1 uses only long established results. Under $(r, \delta)$ surgery, assumption (b) is clearly implied by, but much weaker than the canonical neighborhood property on the whole manifold $\mathbf{M}$, which was used in all the papers so far. In particular there is no need for the gradient estimate on the scalar curvature, which is difficult to prove by itself. Also the surgery does not have to fall on a strong $\delta$ neck for the theorem to hold.*

*However, in proving long time existence of Ricci flow with surgery, one has to show that the canonical neighborhood property holds, by a delicate contradiction argument. (See [P2], [CZ], [KL], [MT], [Tao] and the next chapter of the book). In this argument one supposes the canonical neighborhood property with a fixed accuracy first breaks down at certain time. Then it can be shown that the canonical neighborhood property with a worse accuracy holds at this time. Using this one can prove the weak $\kappa$ noncollapsing property which in term will lead to a* *contradiction through a blow-up argument. During this process the gradient estimate for the scalar curvature is still required. Also this proof of noncollapsing with surgeries via Theorem 8.2.1 seems to work only in the case of Poincaré conjecture. For the full geometrization conjecture, so far one has to use Perelman's argument via reduced distance and volume to derive a localized $\kappa$ noncollapsing result.*

**Remark 8.2.2** *In [Z2], it was shown that under a Ricci flow with finite number of surgeries in finite time, a uniform Sobolev imbedding holds. Recently, in the preprint, The logarithmic Sobolev inequality along the Ricci flow, by Ye, Rugang, arXiv: 0707.2424v4, 2007, a similar result depending on the number of surgeries was stated without proof.*

**Remark 8.2.3** *The strong noncollapsing result clearly implies Hamilton's little loop conjecture with surgeries [Ha7] Section 15, i.e. if the scalar curvature in a geodesic ball of radius $W$ is bounded from above by const./$W^2$, then the injectivity radius at the center of the ball is bounded from below by const.$W$.*

*The conjecture was proved by Perelman [P1] only in the case without surgery. In the case with surgery, using the method of reduced distance etc., Perelman [P2] proved the lower bound of the injectivity radius under the more restrictive assumption that the curvature tensor is bounded in a parabolic cube that is also unscathed by surgeries. However, a priori, there is no knowledge on whether a surgery takes place.*

Let us outline the proof of the theorem. Recall Perelman's $W$ entropy and its monotonicity. They are in fact the monotonicity of the best constants of the Log Sobolev inequality with certain parameters. If a Ricci flow is smooth over a finite time interval, then the best constants of the Log Sobolev inequality with a changing parameter does not decrease. If a Ricci flow undergoes a $(r, \delta)$ surgery with $\delta$ sufficiently small, then the best constant only decreases by at most a constant times the change in volume. This proves the essential monotonicity of the infimum of the $W$ entropy under surgeries (see (8.2.36) and (8.2.37) below). Due to its potential independent interest, we single out this result as a separate theorem here.

At a given time $t$ in a Ricci flow $(\mathbf{M}, g(t))$ and for $\sigma > 0$, let us define

$$\lambda_{\sigma^2}(g(t)) = \inf\{\int [\sigma^2(4|\nabla v|^2 + Rv^2) - v^2 \ln v^2] d\mu(g(t)) - n \ln \sigma \mid v \in C^\infty(\mathbf{M}),\ \|v\|_2 = 1\}. \tag{8.2.2}$$

The number

$$\lambda_{\sigma^2} - \frac{n}{2}\ln(4\pi)$$

can be regarded as the infimum of the $W$ entropy with parameter $\sigma^2$ by defining $u = v^2$. Note that $\lambda_{\sigma^2}(g(t))$ can also be thought as the best Log Sobolev constant with parameter $\sigma^2$. If $t$ is a surgery time, then $\lambda_{\sigma^2}(g(t^+))$ stands for the best Log Sobolev constant with parameter $\sigma^2$ for the manifold right after surgery; and

$$\lambda_{\sigma^2}(g(t^-)) \equiv \lim_{s\to t^-} \lambda_{\sigma^2}(g(s)).$$

The main work in proving Theorem 8.2.1 is to prove the following

**Theorem 8.2.2** *Suppose a Ricci flow* $(\mathbf{M}, g)$ *satisfies conditions (a) and (b) of the previous theorem. Suppose a* $(r,\delta)$ *surgery occurs at time* $T$ *and* $\sigma \in (0,1]$. *There exist positive constants* $\Lambda_0$ *and* $h_0$, *independent of* $T$ *or* $\sigma$ *such that the following holds: if* $\lambda_{\sigma^2}(g(T^+)) \le -\Lambda_0$ *and the surgery radius is smaller than* $h_0$, *then*

$$\lambda_{\sigma^2}(g(T^-)) \le \lambda_{\sigma^2}(g(T^+)) + c|vol(\mathbf{M}(T^-)) - vol(\mathbf{M}(T^+))|.$$

*Here* $vol(\mathbf{M}(T^-))$ *is the volume of the pre-surgery manifold at* $T$ *and* $vol(\mathbf{M}(T^+))$ *is the volume of the post-surgery manifold at* $T$, *and* $c$ *is a positive constant.*

The proof is achieved by a weighted estimate of Agmon type for the minimizing equation of the $W$ entropy. The method is motivated by those at the end of [P2] and [KL] where the change of eigenvalues of the linear operator $4\Delta - R$ was studied. Since our case is nonlinear and contains an extra parameter, the more analysis and estimates are needed. In the end we prove, in finite time, the best constant of the log Sobolev inequality (c.f. L. Gross [Gro]) with certain parameters is uniformly bounded from below by a negative constant, regardless of the number of surgeries. This uniform log Sobolev inequality is then converted by known method to the desired uniform Sobolev inequality which in turn yields strong noncollapsing. The estimate about the change of the best constant of the log Sobolev inequality under one surgery seems to have independent interest in addition to Ricci flow.

We will need three lemmas before carrying out the proof of the theorem. Much of the analysis is focused on the $\epsilon$ horn where a surgery takes place. So we will fix some notations and basic facts concerning the $\epsilon$ horn and the surgery cap. Also we will use $c$ with or without index to denote generic positive constants.

Recall that a $(r, \delta)$ surgery occurs deep inside an $\epsilon$ horn of radius $r$. The horn is cut open at the place where the radius is $h \le \delta^2 r$. Then a cap is attached and a smooth metric is constructed by interpolating between the metric on the horn and the metric on the cap. The resulting manifold right after the surgery is denoted by $\mathbf{M}^+$ and the $\epsilon$ horn thus modified by the surgery is called a capped $\epsilon$ horn with radius $r$.

Let $D$ be a capped $\epsilon$ horn. By assumption, a region $\mathbf{N}$ around the boundary $\partial D$ equipped with the scaled metric $cr^{-2}g$ is $\epsilon$ close, in the $C^{[\epsilon^{-1}]}$ topology, to the standard round neck $S^2 \times (-\epsilon^{-1}, \epsilon^{-1})$. Here $c$ is a generic positive constant such that $cr^{-2}$ equals the scalar curvature at a point on $\partial D$. For this reason we will often take $c = 1$.

Let $\Pi$ be the diffeomorphism from the standard round neck to $\mathbf{N}$ in the definition of $\epsilon$ closeness. Denote by $z$ for a number in $(-\epsilon^{-1}, \epsilon^{-1})$. For $\theta \in S^2$, $(\theta, z)$ is a parametrization of $\mathbf{N}$ via the diffeomorphism $\Pi$. We can identify the metric on $\mathbf{N}$ with its pull back on the round neck by $\Pi$ in this manner. We normalize the parameters so that the capped $\epsilon$ horn lies in the region where $z \ge 0$.

Next we define

$$Y(D) = \inf\{\frac{\int (4|\nabla v|^2 + Rv^2) d\mu(g)}{\left(\int v^{2n/(n-2)} d\mu(g)\right)^{(n-2)/n}} \mid v \in C_0^\infty(D \cup \mathbf{N}), v > 0\}. \tag{8.2.3}$$

**Proposition 8.2.1** *For sufficiently small $\epsilon > 0$, there exists positive constants $C_1, C_2$ such that*

$$C_1 \le Y(D) \le C_2. \tag{8.2.4}$$

Proof. Since $R$ is positive in $D \cup \mathbf{N}$, Y(D) is bounded from above and below by constant multiples of the Yamabe constant

$$Y_0(D) = \inf\{\int \frac{(4\frac{n-1}{n-2}|\nabla v|^2 + Rv^2) d\mu(g)}{\left(\int v^{2n/(n-2)} d\mu(g)\right)^{(n-2)/n}} \mid v \in C_0^\infty(D \cup \mathbf{N}), v > 0\}.$$

So it suffices to prove that the Yamabe constant is bounded between two positive constants.

Let $g = g(x)$ be the metric on $D \cup \mathbf{N}$ then $Y_0(D)$ stays the same under the metric $g_1(x) = R(x)g(x)$.

Consider the manifold $(D \cup \mathbf{N}, g_1)$. By assumption and the $(r, \delta)$ surgery procedure, there is a fixed $r_0 > 0$ such that for any $x \in D \cup \mathbf{N}$, the ball $B(x, r_0)$ under $g_2(y) = R(x)g(y)$, $y \in B(x, r_0)$ is $\epsilon$ close (in

$C^{[\epsilon^{-1}]}$ topology) to a part of the standard capped infinite cylinder. Therefore, for $y$ in the same geodesic ball, the scaled scalar curvature

$$R^{-1}(x)R(y)$$

is $\epsilon$ close under $C^{\epsilon^{-1}-2}$ norm, to a positive function. This positive function is the scalar curvature in the standard capped cylinder, which is both uniformly bounded away from 0 and bounded from above. Actually

$$R(y) = R(x)(h(y) + \xi(y, \epsilon))$$

where $h(y) = 1$ when $y$ is away from the surgery cap and $h(y)$ is the scalar curvature of the surgery cap otherwise. The $C^{[\epsilon^{-1}-2]}$ norm of $\xi$ is less than $const.\epsilon$.

Hence, for $y$ away from the surgery cap and under the metric $g_1(y) = R(y)g(y)$, the same geodesic ball, is $\epsilon$ close (in $C^{[\epsilon^{-1}-2]}$ topology) to a part of the standard capped infinite cylinder. For $y$ in the surgery cap, since $h = h(y)$ has bounded $C^2$ norm, the curvatures under $g_1(y) = R(y)g(y)$ are uniformly bounded.

Since $\epsilon$ is sufficiently small, we know that the injectivity radius of $(D \cup \mathbf{N}, g_1)$ is bounded from below by a positive constant; and its Ricci curvature is bounded from below. Actually it is easy to see that these hold for a much larger domain containing $(D \cup \mathbf{N}, g_1)$. By Proposition 6 in [Heb1], we can find a positive constant $C$ such that

$$\left( \int v^{2n/(n-2)} d\mu(g_1) \right)^{(n-2)/n} \le C \int |\nabla_1 v|^2 d\mu(g_1) + C \int v^2 d\mu(g_1)$$

for all $v \in C_0^\infty(D \cup \mathbf{N})$. Recall the scalar curvature of $(D \cup \mathbf{N}, g_1)$ is bounded between two positive constants outside of the surgery cap. Inside the surgery cap the scalar curvature is bounded from below by an absolute negative constant. Therefore for a constant still named $C$,

$$\left( \int v^{2n/(n-2)} d\mu(g_1) \right)^{(n-2)/n} \le C \int (4\frac{n-1}{n-2}|\nabla_1 v|^2 + R_1 v^2) d\mu(g_1) + C \int v^2 \alpha^2 d\mu(g_1)$$

for all $v \in C_0^\infty(D \cup \mathbf{N})$. Here $\alpha$ is a nonnegative, smooth function supported in a neighborhood of the surgery cap, which is bounded from above by an absolute constant. Also $\nabla_1$ and $R_1$ are the gradient and scalar curvature under $g_1$ respectively. Note $R_1$ may not be positive inside the surgery cap.

Now we scale back to the metric $g = R^{-1}(y)g_1(y)$. By conformal invariance of all but the last term, it is easy to check that, after renaming $R^{(n-2)/4}v$ by $v$,

$$\left(\int v^{2n/(n-2)} d\mu(g)\right)^{(n-2)/n} \le C\int (4\frac{n-1}{n-2}|\nabla v|^2 + Rv^2) d\mu(g) + C\int v^2(x)R(x)\alpha^2(x)d\mu(g)$$

for all $v \in C_0^\infty(D \cup \mathbf{N})$. Now the scalar curvature is positive everywhere.

Hence we see that $Y_0(D)$ is bounded from below by a positive constant when $\epsilon$ is sufficiently small. It is also bounded from above by the Yamabe constant of $S^n$. Since $Y_0(D)$ and $Y(D)$ are comparable, we have shown that

$$0 < Const_1 \le Y(D) \le Const_2 \tag{8.2.5}$$

when $\epsilon$ is sufficiently small.

This proves the proposition. □

Next we present

**Lemma 8.2.1** *Let $(\mathbf{M}^+, g)$ be a manifold right after a $(r, \delta)$ surgery. Let $D \subset \mathbf{M}^+$ be a capped $\epsilon$ horn of radius $r$. Here $\epsilon$ is a sufficiently small positive number.*

*Suppose $u$ with $\|u\|_{L^2(\mathbf{M}^+)} = 1$ is a positive solution to the equation*

$$\sigma^2(4\Delta u - Ru) + 2u\ln u + \Lambda u + n(\ln\sigma)u = 0. \tag{8.2.6}$$

*Here $\sigma > 0$ and $\Lambda \le 0$.*

*Then there exists a positive constant $C$ depending only on $Y(D)$, $n$ but not on the smallness of $\epsilon$ such that*

$$\sup_D u^2 \le C\max(r^{-n}, \sigma^{-n}).$$

Proof. After taking the scaling

$$g_1 = \sigma^{-2}g, \ R_1 = \sigma^2 R, \ u_1 = \sigma^{n/2}u$$

we see that $u_1$ satisfies

$$4\Delta_1 u_1 - R_1 u_1 + 2u_1 \ln u_1 + \Lambda u_1 = 0.$$

Since the result in the lemma is independent of the above scaling, we can just prove it for $\sigma = 1$.

So let $u$ be a positive solution to the equation

$$4\Delta u - Ru + 2u \ln u + \Lambda u = 0$$

in $\mathbf{M}^+$ such that its $L^2$ norm is 1. Given any $p \geq 1$, it is easy to see that

$$-4\Delta u^p + pRu^p \leq 2pu^p \ln u. \tag{8.2.7}$$

We select a smooth cut off function $\phi$ which is one in $D$ and 0 outside of $D \cup \mathbf{N}$. Writing $w = u^p$ and using $w\phi^2$ as a test function in (8.2.7), we deduce

$$4 \int \nabla(w\phi^2)\nabla w + p \int R(w\phi)^2 \leq 2p \int (w\phi)^2 \ln u.$$

Here in this lemma and next, we will ignore the volume element in integrations because the metric is fixed. Since the scalar curvature $R$ is positive in the support of $\phi$ and $p \geq 1$, this shows

$$4 \int \nabla(w\phi^2)\nabla w + \int R(w\phi)^2 \leq p \int (w\phi)^2 \ln u^2.$$

Using integration by parts, we have

$$4 \int |\nabla(w\phi)|^2 + \int R(w\phi)^2 \leq 4 \int |\nabla\phi|^2 w^2 + p \int (w\phi)^2 \ln u^2. \tag{8.2.8}$$

We need to dominate the last term in (8.2.8) by the left-hand side of (8.2.8). For one positive number $a$ to be chosen later, it is clear that

$$\ln u^2 \leq u^{2a} + c(a).$$

Hence for any fixed $q > n/2$, the Hölder inequality implies

$$\begin{aligned} p \int (w\phi)^2 \ln u^2 &\leq p \int (w\phi)^2 u^{2a} + pc(a) \int (w\phi)^2 \\ &\leq p \left( \int u^{2aq} \right)^{1/q} \left( \int (w\phi)^{2q/(q-1)} \right)^{(q-1)/q} \\ &\quad + pc(a) \int (w\phi)^2. \end{aligned}$$

We take $a = 1/q$ so that $2aq = 2$. Since the $L^2$ norm of $u$ is 1 by assumption, the above implies

$$p \int (w\phi)^2 \ln u^2 \leq p \left( \int (w\phi)^{2q/(q-1)} \right)^{(q-1)/q} + pc(a) \int (w\phi)^2.$$

By interpolation inequality (see p84 [HL] e.g.), it holds, for any $b > 0$,

$$\left(\int (w\phi)^{2q/(q-1)}\right)^{(q-1)/q} \le b\left(\int (w\phi)^{2n/(n-2)}\right)^{(n-2)/n} + c(n,q)b^{-n/(2q-n)}\int (w\phi)^2.$$

Therefore

$$p\int (w\phi)^2 \ln u^2 \le pb\left(\int (w\phi)^{2n/(n-2)}\right)^{(n-2)/n} + c(n,q)pb^{-n/(2q-n)}\int (w\phi)^2 + pc(a)\int (w\phi)^2. \tag{8.2.9}$$

By the definition of $Y(D)$ in (8.2.3), we see that (8.2.8) gives

$$Y(D)\left(\int (w\phi)^{2n/(n-2)}\right)^{(n-2)/n} \le 4\int |\nabla\phi|^2 w^2 + p\int (w\phi)^2 \ln u^2. \tag{8.2.10}$$

Substituting (8.2.9) to the right-hand side of (8.2.10), we arrive at

$$\begin{aligned} Y(D)\left(\int w^{2n/(n-2)}\right)^{(n-2)/n} &\le 4\int |\nabla\phi|^2 w^2 \\ &+ pb\left(\int (w\phi)^{2n/(n-2)}\right)^{(n-2)/n} + c(n,q)pb^{-n/(2q-n)}\int (w\phi)^2 \\ &+ pc(a)\int (w\phi)^2. \end{aligned}$$

Take $b$ so that $pb = Y(D)/2$. It is clear there exist positive constant $c = c(Y(D), n, q)$ and $\alpha = \alpha(n, q)$ such that

$$\left(\int (w\phi)^{2n/(n-2)}\right)^{(n-2)/n} \le c(p+1)^{\alpha}\int (|\nabla\phi|^2 + 1)w^2. \tag{8.2.11}$$

From here one can use Moser's iteration to prove the desired bound. Let $z$ be the longitudinal parameter for $D$ described before the lemma. For $z_2$ and $z_1$ such that $-1 \le z_2 < z_1 < 0$, we construct a smooth function of $z$, called $\xi$ such that $\xi(z) = 1$ when $z \ge z_1$; $\xi(z) = 0$ when $z < z_2$ and $\xi(z) \in (0,1)$ for the rest of $z$. Set the test function $\phi = \xi(z) = \xi(z(x))$. Then it is clear that

$$|\nabla\phi| \le \frac{c}{r(z_1 - z_2)}. \tag{8.2.12}$$

Write

$$D_i = \{x \in \mathbf{M}^+ \mid z(x) > z_i\}, \qquad i = 1, 2.$$

By (8.2.11) and (8.2.12)

$$\left(\int_{D_1} w^{2n/(n-2)}\right)^{(n-2)/n} \leq c \max\{\frac{1}{[(z_1 - z_2)r]^2}, 1\}(p+1)^\alpha \int_{D_2} w^2. \tag{8.2.13}$$

Recall that $w = u^p$. We iterate (8.2.13) with $p = (n/(n-2))^i$, $i = 0, 1, 2, \ldots$ in conjunction with choosing

$$z_1 = -(1/2 + 1/2^{i+2}), \qquad z_2 = -(1/2 + 1/2^{i+1}).$$

Following Moser, we will get

$$\sup_D u^2 \leq C \max(r^{-n}, 1) \int u^2.$$

□

**Remark 8.2.4** *One can avoid using (8.2.5) but work on each $\epsilon$ neck and the surgery cap directly as above. Then one can show that $u^2(x) \leq C \max\{R^{n/2}(x), \sigma^{-n}\}$. This weaker bound is sufficient for proving the main result. This will be clear in the proof below.*

The next lemma is a nonlinear version of the result in [P2] and Lemma 92.10 in [KL]. This estimate has its origin in the weighted Agmon type estimate of eigenfunctions of the Laplacian.

**Lemma 8.2.2** *Let $(\mathbf{M}, g)$ be any compact manifold without boundary. Suppose $u$ is a positive solution to the inequality*

$$4\Delta u - Ru + 2u \ln u + \Lambda u \geq 0 \tag{8.2.14}$$

*with $\Lambda \leq 0$.*

*Given a nonnegative function $\phi \in C^\infty(\mathbf{M})$, $\phi \leq 1$, suppose there is a smooth function $f$ such that $R \geq 0$ in the support of $\phi$ and that*

$$4|\nabla f|^2 \leq R - 2\ln^+ u + \frac{|\Lambda|}{2}$$

*also in the support of $\phi$. Then*

$$\frac{|\Lambda|}{2}\|e^f \phi u\|_2 \leq 8\left[\sup_{x \in supp \nabla\phi} e^f \sqrt{R - 2\ln^+ u + \frac{|\Lambda|}{2}} + \|e^f \nabla\phi\|_\infty\right] \|u\|_2.$$

Proof. The main point of the lemma is that the right-hand side depends only on information in the support of $\nabla\phi$.

Using integration by parts,

$$\int e^f\phi u\left(-4\Delta + R - 2\ln u - \Lambda - 4|\nabla f|^2\right)(e^f\phi u)$$
$$= 4\int |\nabla(e^f\phi u)|^2 + \int (e^f\phi u)^2(R - 2\ln u - \Lambda - 4|\nabla f|^2).$$

By assumption

$$R - 2\ln u - \Lambda - 4|\nabla f|^2 \geq |\Lambda|/2.$$

Hence

$$\int e^f\phi u\left(-4\Delta + R - 2\ln u - \Lambda - 4|\nabla f|^2\right)(e^f\phi u) \geq \frac{|\Lambda|}{2}\int (e^f\phi u)^2. \tag{8.2.15}$$

By straightforward calculation,

$$\begin{aligned}\text{left side of (8.2.15)} = &\int (e^f\phi)^2 u\,(-4\Delta u + Ru - 2u\ln u - \Lambda u)\\ &- \int e^f\phi u\left[8\nabla(e^f\phi)\nabla u + 4\Delta(e^f\phi)u\right]\\ &- 4\int (e^f\phi u)^2|\nabla f|^2\\ \leq &- \int e^f\phi u\left[8\nabla(e^f\phi)\nabla u + 4\Delta(e^f\phi)u\right]\\ &- 4\int (e^f\phi u)^2|\nabla f|^2.\end{aligned}$$

The last step is due to (8.2.14). This together with (8.2.15) yield

$$\frac{|\Lambda|}{2}\int (e^f\phi u)^2 \leq -\int e^f\phi u\left[8\nabla(e^f\phi)\nabla u + 4\Delta(e^f\phi)u\right] - 4\int (e^f\phi u)^2|\nabla f|^2.$$

Performing integration by parts on the term containing $\Delta$, we deduce

$$\begin{aligned}\frac{|\Lambda|}{2}\int (e^f\phi u)^2 \leq &-8\int e^f\phi u\nabla(e^f\phi)\nabla u + \int 4\nabla(e^f\phi)\nabla(e^f\phi u^2)\\ &- 4\int (e^f\phi u)^2|\nabla f|^2.\end{aligned}$$

This shows

$$\frac{|\Lambda|}{2}\int (e^f\phi u)^2 \leq 4\int |\nabla(e^f\phi)|^2u^2 - 4\int (e^f\phi u)^2|\nabla f|^2.$$

Hence

$$\frac{|\Lambda|}{2}\int (e^f \phi u)^2 \le 4\int \left[(e^f\phi)|^2|\nabla f|^2 + 2e^{2f}(\nabla f \nabla \phi)\phi + e^{2f}|\nabla\phi|^2\right] u^2$$
$$- 4\int (e^f\phi u)^2 |\nabla f|^2.$$

The first and the last term on the right-hand side cancel to give

$$\frac{|\Lambda|}{2}\int (e^f \phi u)^2 \le 8\int e^{2f}(\nabla f \nabla \phi)\phi u^2 + 4\int e^{2f}|\nabla\phi|^2 u^2.$$

Note that the integrations on the right side only take place in the support of $\nabla\phi$. Thus it shows, by assumption on $|\nabla f|^2$,

$$\frac{|\Lambda|}{2}\int (e^f \phi u)^2 \le 4\int_{supp\nabla\phi} e^{2f}|\nabla f|^2\phi^2 u^2 + 8\int e^{2f}|\nabla\phi|^2 u^2$$
$$\le \int_{supp\nabla\phi} e^{2f}(R - 2\ln^+ u + \frac{|\Lambda|}{2})\phi^2 u^2 + 8\int e^{2f}|\nabla\phi|^2 u^2.$$

So finally

$$\frac{|\Lambda|}{2}\int (e^f \phi u)^2 \le \sup_{x\in supp\nabla\phi} e^{2f}(R - 2\ln^+ u + \frac{|\Lambda|}{2}) \int u^2$$
$$+ 8\sup e^{2f}|\nabla\phi|^2 \int u^2. \qquad \square$$

**Lemma 8.2.3** *Let* $(\mathbf{M}, g)$ *be any compact manifold without boundary and* $\mathbf{X}$ *be a domain in* $\mathbf{M}$*. Define*

$$\lambda_X = \inf\{\int (4|\nabla v|^2 + Rv^2 - v^2\ln v^2) \mid v \in C_0^\infty(\mathbf{X}),\ \|v\|_2 = 1\}, \tag{8.2.16}$$

$$\lambda_M = \inf\{\int (4|\nabla v|^2 + Rv^2 - v^2\ln v^2) \mid v \in C^\infty(\mathbf{M}),\ \|v\|_2 = 1\}, \tag{8.2.17}$$

*Let* $u(>0)$ *be a minimizer for* $\lambda_M$*. For any smooth cut-off function* $\eta \in C_0^\infty(\mathbf{X}),\ \ 0 \le \eta \le 1$*, it holds*

$$\lambda_X \le \lambda_M + 4\frac{\int u^2|\nabla\eta|^2}{\int (u\eta)^2} - \frac{\int (u\eta)^2\ln\eta^2}{\int (u\eta)^2}.$$

PROOF. Since $\eta u/\|\eta u\|_2 \in C_0^\infty(\mathbf{X})$ and its $L^2$ norm is 1, we have, by definition,

$$\lambda_X \le \int \left[ 4\frac{|\nabla(\eta u)|^2}{\|\eta u\|_2^2} + R\frac{(\eta u)^2}{\|\eta u\|_2^2} - \frac{(\eta u)^2}{\|\eta u\|_2^2} \ln \frac{(\eta u)^2}{\|\eta u\|_2^2} \right].$$

This implies

$$\lambda_X \|\eta u\|_2^2 \le \int [4|\nabla(\eta u)|^2 + R(\eta u)^2 - (\eta u)^2 \ln(\eta u)^2] + \|\eta u\|_2^2 \ln \|\eta u\|_2^2. \tag{8.2.18}$$

On the other hand, $u$ is a smooth positive solution (cf [Ro]) of the equation

$$4\Delta u - Ru + 2u \ln u + \lambda_M u = 0.$$

Using $\eta^2 u$ as a test function for the equation, we deduce

$$\lambda_M \int (\eta u)^2 = -4 \int (\Delta u)\eta^2 u + \int R(\eta u)^2 - 2 \int (\eta u)^2 \ln u.$$

By direct calculation

$$-4 \int (\Delta u)\eta^2 u = 4 \int |\nabla(\eta u)|^2 - 4 \int u^2 |\nabla \eta|^2.$$

Hence

$$\lambda_M \int (\eta u)^2 = 4 \int |\nabla(\eta u)|^2 - 4 \int u^2 |\nabla \eta|^2 + \int R(\eta u)^2 - 2 \int (\eta u)^2 \ln u. \tag{8.2.19}$$

Comparing (8.2.19) with (8.2.18) and noting that $\|\eta u\|_2 < 1$, we obtain

$$\lambda_X \|\eta u\|_2^2 \le \lambda_M \|\eta u\|_2^2 + 4 \int |\nabla \eta|^2 u^2 - \int (\eta u)^2 \ln \eta^2.$$

□

Now we are ready to give a

**Proof of Theorem 8.2.1.**

First we note that the proof of Theorem 8.2.2 is included here.

At a given time $t$ in a Ricci flow $(\mathbf{M}, g(t))$ and for $\sigma > 0$, let us recall, from (8.2.2),

$$\lambda_{\sigma^2}(g(t)) = \inf\{ \int [\sigma^2(4|\nabla v|^2 + Rv^2) - v^2 \ln v^2] d\mu(g(t)) - n \ln \sigma \mid v \in C^\infty(\mathbf{M}),\ \|v\|_2 = 1 \}.$$

The main aim is to find a uniform lower bound for $\lambda_{\sigma^2}(g(t))$, $t \in [T_1, T_2]$, $\sigma \in (0, 1]$. So without loss of generality, we assume it is negative.

The rest of the proof is divided into 5 steps.

*Step 1.* We estimate the change of $\lambda_{\sigma^2}(t)$, the best constant of the log Sobolev inequality, after one $(r, \delta)$ surgery.

It will be clear that the proof below is independent of the number of cut offs occurring in one surgery time $T$. Therefore we just assume there is one $\epsilon$ horn and one cut off at $T$.

Let $(\mathbf{M}, g(T^+))$ be the manifold right after the surgery and

$$\Lambda \equiv \lambda_{\sigma^2}(g(T^+))$$

be the best constant for this post surgery manifold, defined in (8.2.2).

By [Ro], there is a smooth positive function $u$ that reaches the infimum in (8.2.2) and $u$ solves

$$\sigma^2(4\Delta u - Ru) + 2u \ln u + \Lambda u + n(\ln \sigma)u = 0. \tag{8.2.20}$$

After taking the scaling

$$g_1 = \sigma^{-2} g(T^+), \ R_1 = \sigma^2 R, d_1 = \sigma^{-1} d, \ u_1 = \sigma^{n/2} u$$

we see that $u_1$ satisfies

$$4\Delta_1 u_1 - R_1 u_1 + 2u_1 \ln u_1 + \Lambda u_1 = 0 \tag{8.2.21}$$

and

$$\Lambda = \inf\{\int ((4|\nabla_{g_1} v|^2 + R_1 v^2 - v^2 \ln v^2) d\mu(g_1) \mid v \in C^\infty(\mathbf{M}^+), \ \|v\|_2 = 1\}. \tag{8.2.22}$$

Denote by $U$ the $\sigma^{-1}h$ neighborhood of the surgery cap $\mathbf{C}$ under $g_1$, i.e.

$$U = \{x \in (\mathbf{M}, g_1(T^+)) \mid d_1(x, \mathbf{C}) < \sigma^{-1}h\} = \{x \in M^+ \mid d(x, \mathbf{C}) < h\}.$$

Note that $U - \mathbf{C}$ is part of the $\epsilon$ tube which is unaffected by the surgery. Therefore, $U - \mathbf{C}$ is $\epsilon$ close to a portion of the standard round neck under the scaled metric $\sigma^2 h^{-2} g_1$. Actually it is even $\delta(< \epsilon)$ close if it is part of the strong $\delta$ neck. But we do not need this fact. Following the description at the beginning of the section, there is a longitudinal parametrization of $U - \mathbf{C}$, called $z$ which maps $U - \mathbf{C}$ to $(-1, 0) \subset$

$(-\epsilon^{-1}, \epsilon^{-1})$. Let $\zeta : [-1, 0] \to [0, 1]$ be a smooth decreasing function such that $\zeta(-1) = 1$ and $\zeta(0) = 0$. Then $\eta \equiv \zeta(z(x))$ maps $U - \mathbf{C}$ to $(0, 1)$. We then extend $\eta$ to be a cut off function on the whole manifold by setting $\eta = 1$ in $\mathbf{M}^+ - U$ and $\eta = 0$ in $\mathbf{C}$.

Define

$$\Lambda_X = \inf\{ \int ((4|\nabla_{g_1} v|^2 + R_1 v^2 - v^2 \ln v^2) d\mu(g_1) \; |v \in C_0^\infty(\mathbf{M}^+ - \mathbf{C}), \\ \|v\|_2 = 1\}. \tag{8.2.23}$$

Then it is clear that

$$\lambda_{\sigma^2}(g(T^-)) \le \Lambda_X.$$

By Lemma 8.2.3,

$$\Lambda_X \le \Lambda + 4 \frac{\int u_1^2 |\nabla_{g_1} \eta|^2 d\mu(g_1)}{\int (u_1 \eta)^2 d\mu(g_1)} - \frac{\int (u_1 \eta)^2 \ln \eta^2 d\mu(g_1)}{\int (u_1 \eta)^2 d\mu(g_1)}.$$

Observe that the supports of $\nabla_{g_1} \eta$ and $\eta \ln \eta$ are in $U - \mathbf{C}$. Moreover

$$|\nabla_{g_1} \eta| \le \frac{c\sigma}{h}, \qquad -\eta^2 \ln \eta^2 \le c.$$

Therefore the above shows

$$\lambda_{\sigma^2}(g(T^-)) \le \Lambda_X \le \Lambda + \frac{4c\sigma^2}{h^2} \frac{\int_U u_1^2 d\mu(g_1)}{1 - \int_U u_1^2 d\mu(g_1)} + c \frac{\int_U u_1^2 d\mu(g_1)}{1 - \int_U u_1^2 d\mu(g_1)}. \tag{8.2.24}$$

Recall that $\Lambda = \lambda_{\sigma^2}(g(T^+))$. So, in order to bound it below, we need to show that $\int_U u_1^2 d\mu(g_1)$ is small. This is where we will use Lemma 8.2.1 and 8.2.2.

Under the metric $g_1 = \sigma^{-2} g$, the capped $\epsilon$ horn $D$ of radius $r$ under $g(T^+)$ is just a capped $\epsilon$ horn of radius $r_1 = \sigma^{-1} r$. Using the longitudinal parametrization $z$ of $D$ as described at the beginning the section, we can construct a cut-off function $\phi = \phi(z(x))$ for $x \in M^+$, which satisfies the following property.

i) $\{x \in \mathbf{M} \mid z(x) = 0\}$ is the boundary of $D$.

ii) If $z \le 0$, then $\phi(z) = 0$; and if $z \ge 1$, then $\phi(z) = 1$.

iii) $0 \le \phi \le 1$; $\qquad |\nabla_{g_1} \phi| \le \frac{c}{r_1}$.

iv) $\phi$ is set to be zero outside of $D$ and is set to be 1 to the right of the set

$$\{x \in \mathbf{M}^+ \mid z(x) = 1\}.$$

Notice that the support of $\nabla\phi$ is in the set where $z$ is between 0 and 1. Applying Lemma 8.2.1 on $u_1$, which satisfies (8.2.21), we know that

$$u_1(x) \le c \max\{\frac{1}{r_1^{n/2}}, 1\}, \qquad x \in D.$$

Hence, for a negative number $\Lambda_0$ with $|\Lambda_0|$ being sufficiently large,

$$\begin{cases} R_1(x) - 2\ln^+ u_1(x) + \frac{|\Lambda_0|}{2} \le c r_1^{-2} + \frac{|\Lambda_0|}{2}, \qquad x \in supp \nabla_{g_1}\phi; \\ R_1(x) - 2\ln^+ u_1(x) + \frac{|\Lambda_0|}{2} \\ \quad \ge \frac{R_1(x)}{2} + c r_1^{-2} - c_1 \ln^+ \max\{\frac{1}{r_1}, 1\} + \frac{|\Lambda_0|}{2} \ge \frac{R_1(x)}{2} + \frac{|\Lambda_0|}{4}, \qquad x \in D. \end{cases} \tag{8.2.25}$$

We stress that $\Lambda_0$ is independent of the size of $r_1 = \sigma^{-1} r$, which could be large or small due to the scaling factor $\sigma$.

Recall that we aim to find a uniform lower bound for $\Lambda$. If $\Lambda = \lambda_{\sigma^2}(g(T^+)) \ge \Lambda_0$, then we are in good shape. So we assume throughout that $\Lambda \le \Lambda_0$. Then, by (8.2.21), it holds

$$4\Delta_1 u_1 - R_1 u_1 + 2u_1 \ln u_1 + \Lambda_0 u_1 \ge 0 \tag{8.2.26}$$

Motivated by the last section of [P2] and Lemma 92.10 in [KL], we choose a function $f = f(x)$ as the distance between $x$ and the set $z^{-1}(0)$ under the metric

$$\frac{1}{4}(R_1(x) - 2\ln^+ u_1(x) + \frac{|\Lambda_0|}{2}) g_1(x), \qquad x \in D.$$

By the first inequality in (8.2.25), in the support of $\nabla_{g_1}\phi$,

$$4|\nabla_{g_1} f|^2 \le c r_1^{-2} + \frac{|\Lambda_0|}{2} \tag{8.2.27}$$

and in $D$,

$$4|\nabla_{g_1} f|^2 \le R_1(x) - 2\ln^+ u_1(x) + \frac{|\Lambda_0|}{2}. \tag{8.2.28}$$

Note that the right-hand side of (8.2.28) is positive by the second inequality in (8.2.25).

Inequalities (8.2.28) and (8.2.26) allow us to use Lemma 8.2.2 (with $\Lambda$ there replaced by $\Lambda_0$ here) to conclude

$$\frac{|\Lambda_0|}{2}\|e^f \phi u_1\|_2 \le 8 \left[ \sup_{x \in supp \nabla_{g_1}\phi} e^f \sqrt{R_1 - 2\ln^+ u_1 + \frac{|\Lambda_0|}{2}} + \|e^f \nabla_{g_1}\phi\|_\infty \right] \times \|u_1\|_2.$$

Here the underlying metric is $g_1$. By (8.2.25) (first item) this shows

$$\frac{|\Lambda_0|}{2}\|e^f\phi u_1\|_2 \le c \sup_{x\in supp\nabla_{g_1}\phi} e^f\sqrt{(\frac{1}{r_1^2}+|\Lambda_0|)} \quad \|u_1\|_2. \tag{8.2.29}$$

From (8.2.29), we will derive a bound for $\|u_1\|_{L^2(U)}$ which holds for all finite $\sigma$. Here and later $\|u_1\|_{L^2(U)}$ stands for the norm under the metric $g_1$.

First, we note from (8.2.29)

$$\frac{|\Lambda_0|}{2}\inf_U e^f\|u_1\|_{L^2(U)} \le c \sup_{x\in supp\nabla_{g_1}\phi} e^f \sqrt{(\frac{\sigma^2}{r^2}+|\Lambda_0|)} \quad \|u_1\|_2. \tag{8.2.30}$$

Let us remember that $U$ lies deep inside the capped $\epsilon$ horn $D$. Going from $\partial D$ (i.e. $z^{-1}(0)$) to $U$, one must traverse a number of disjoint $\epsilon$ necks. The ratio of scalar curvatures between the two ends of an $\epsilon$ neck is bounded by $e^{c_2\epsilon}$ for some fixed $c_2>0$. The ratio of the scalar curvatures between $\partial U$ and $\partial D$ is $c_3r^2h^{-2}$, which is independent of the scaling factor $\sigma$. Therefore one must traverse at least

$$K \equiv \frac{1}{c_2\epsilon}\ln(c_3r^2h^{-2}) \tag{8.2.31}$$

number of $\epsilon$ necks to reach $U$. Note $K$ is independent of $\sigma$.

Let $G_i$ be one of the $\epsilon$ necks. The distance between its two ends under the metric $g$ is comparable to $2\epsilon^{-1}R^{-1/2}(x_i)$ where $x_i$ is a point in $G_i$. So, under the metric,

$$\frac{1}{4}(R_1(x)-2\ln^+ u_1(x)+\frac{|\Lambda_0|}{2})g_1(x)$$

the distance between the two ends is bounded from below by

$$c_4\inf_{x\in G_i}\sqrt{\frac{1}{4}(R_1(x)-2\ln^+ u_1(x)+\frac{|\Lambda_0|}{2})}R_1^{-1/2}(x_i)\epsilon^{-1} \ge c_5\epsilon^{-1}.$$

Here the last inequality comes from the second item in (8.2.25). This means that the function $f$ increases by at least $c_5\epsilon^{-1}$ when traversing one $\epsilon$ neck.

Next we observe that

$$\inf_{G_2} f \ge \sup_{supp\nabla_{g_1}\phi} f$$

since the support of $\nabla_{g_1}\phi$ is contained in the first $\epsilon$ neck $G_1$. Therefore

$$\inf_U f \ge c_5 \epsilon^{-1}(K-2) + \inf_{G_2} f \ge c_5\epsilon^{-1}(K-2) + \sup_{supp\nabla_{g_1}\phi} f.$$

Substituting this to (8.2.30), we deduce

$$\|u_1\|_{L^2(U)} \le 2c|\Lambda_0^{-1}| e^{-c_5\epsilon^{-1}(K-2)} \sqrt{(\frac{\sigma^2}{r^2}+|\Lambda_0|)} \quad \|u_1\|_2.$$

Therefore, by the formula for $K$ in (8.2.31),

$$\|u_1\|_{L^2(U)} \le c_6|\Lambda_0^{-1}|(r^{-2}h^2)^{c_7\epsilon^{-2}} \sqrt{(\frac{\sigma^2}{r^2}+|\Lambda_0|)} \quad \|u_1\|_2.$$

Since $r \le 1$ by assumption, we know that

$$\begin{aligned}\|u_1\|_{L^2(U)} &\le c_8 C(\Lambda_0)(\sigma+1)r^{-1}(r^{-2}h^2)^{c_7\epsilon^{-2}} \ \|u\|_2 \\ &= c_8 C(\Lambda_0) h^5 \|u\|_2 \, \frac{h^{2c_7\epsilon^{-2}-5}}{r^{2c_7\epsilon^{-2}+1}}.\end{aligned}$$

Since $h \le \delta^2 r \le 1$, it is easy to see that we can choose $\delta$ as a suitable power of $r$ so that

$$\|u\|_{L^2(U,d\mu(g))} = \|u_1\|_{L^2(U)} \le c_9(\sigma+1)h^5\|u\|_2 \tag{8.2.32}$$

if $\epsilon$ is made sufficiently small, once and for all. For instance, one can choose $\delta \le r^{1/2}$.

Substituting (8.2.32) to (8.2.24), we see that

$$\lambda_{\sigma^2}(g(T^-)) \le \Lambda + c_{10}(\sigma+1)^3 h^3 \frac{1}{1-c_9(\sigma+1)h^5}.$$

Hence, given any $\sigma_0 > 0$, we have, for all $\sigma \in (0, \sigma_0)$, either

$$\lambda_{\sigma^2}(g(T^+)) \ge \Lambda_0$$

or

$$\lambda_{\sigma^2}(g(T^-)) \le \Lambda + c_{11}(\sigma+1)^3h^3 = \lambda_{\sigma^2}(g(T^+)) + c_{11}(\sigma+1)^3h^3$$

provided that $h \le (2(\sigma_0+1)c_9)^{-1/5}$. This shows, for all $\sigma \in (0, \sigma_0]$, either $\lambda_{\sigma^2}(g(T^+)) \ge \Lambda_0$, or

$$\lambda_{\sigma^2}(g(T^-)) \le \lambda_{\sigma^2}(g(T^+)) + c_{12}|vol(\mathbf{M}(T^-)) - vol(\mathbf{M}(T^+))|. \tag{8.2.33}$$

Here $vol(\mathbf{M}(T^-))$ is the volume of the pre-surgery manifold at $T$ and $vol(\mathbf{M}(T^+))$ is the volume of the post-surgery manifold at $T$. Note this proves Theorem 8.2.2.

From the calculation around (8.2.32), when $\epsilon > 0$ is sufficiently small, the surgery radius $h$ can be chosen as

$$h = z_0 r^2 \tag{8.2.34}$$

where $z_0$ is some constant smaller than 1. Actually the power on $r$ can also be any number greater than 1.

*Step 2.* We estimate the change of the best constant in the log Sobolev inequality in a given time interval without surgery.

Suppose the Ricci flow is smooth from time $t_1$ to $t_2$. Let $t \in (t_1, t_2)$ and $\sigma > 0$. Recall that, for $(\mathbf{M}, g(t))$, Perelman's $W$ entropy with parameter $\tau$ is

$$W(g, f, \tau) = \int_{\mathbf{M}} \big(\tau(R + |\nabla f|^2) + f - n\big)\, \tilde{u}\, d\mu(g(t))$$

where $\tilde{u} = \frac{e^{-f}}{(4\pi\tau)^{n/2}}$. We are using $\tilde{u}$ in this step to distinguish from $u$ in the last step.

We define

$$\tau = \tau(t) = \sigma^2 + t_2 - t$$

so that $\tau_1 = \epsilon^2 + t_2 - t_1$ and $\tau_2 = \sigma^2$ (by taking $t = t_1$ and $t = t_2$ respectively).

Let $\tilde{u}_2$ be a minimizer of the entropy $W(g(t), f, \tau_2)$ for all $\tilde{u}$ such that $\int \tilde{u} d\mu(g(t_2)) = 1$.

We solve the conjugate heat equation with the final value chosen as $\tilde{u}_2$ at $t = t_2$. Let $\tilde{u}_1$ be the value of the solution of the conjugate heat equation at $t = t_1$. As usual, we define functions $f_i$ with $i = 1, 2$ by the relation $\tilde{u}_i = e^{-f_i}/(4\pi\tau_i)^{n/2}$, $i = 1, 2$. Then, by the monotonicity of the $W$ entropy ([P1])

$$\begin{aligned} inf_{\int \tilde{u}_0 d\mu(g(t_1))=1} W(g(t_1), f_0, \tau_1) &\le W(g(t_1), f_1, \tau_1) \le W(g(t_2), f_2, \tau_2) \\ &= inf_{\int \tilde{u} d\mu(g(t_2))=1} W(g(t_2), f, \tau_2). \end{aligned}$$

Here $f_0$ and $f$ are given by the formulas

$$\tilde{u}_0 = e^{-f_0}/(4\pi\tau_1)^{n/2}, \qquad \tilde{u} = e^{-f}/(4\pi\tau_2)^{n/2}.$$

Using these notations we can rewrite the above as

$$\begin{aligned}
&\inf_{\|\tilde{u}\|_1=1} \int_{\mathbf{M}} \left( \sigma^2 (R + |\nabla \ln \tilde{u}|^2) - \ln \tilde{u} - \ln(4\pi\sigma^2)^{n/2} \right) \tilde{u}\, d\mu(g(t_2)) \\
&\geq \inf_{\|\tilde{u}_0\|_1=1} \int_{\mathbf{M}} \Big( (\sigma^2 + t_2 - t_1)(R + |\nabla \ln \tilde{u}_0|^2) \\
&\qquad\qquad - \ln \tilde{u}_0 - \ln(4\pi(\sigma^2 + t_2 - t_1))^{n/2} \Big) \tilde{u}_0\, d\mu(g(t_1)).
\end{aligned}$$

Denote $v = \sqrt{\tilde{u}}$ and $v_0 = \sqrt{\tilde{u}_0}$. This inequality is converted to

$$\begin{aligned}
&\inf_{\|v\|_2=1} \int_{\mathbf{M}} \left( \sigma^2 (Rv^2 + 4|\nabla v|^2) - v^2 \ln v^2 \right)\, d\mu(g(t_2)) - \ln(4\pi\sigma^2)^{n/2} \\
&\geq \inf_{\|v_0\|_2=1} \int_{\mathbf{M}} \left( 4(\sigma^2 + t_2 - t_1)(\frac{1}{4} R v_0^2 + |\nabla v_0|^2) - v_0^2 \ln v_0^2 \right)\, d\mu(g(t_1)) \\
&\quad - \ln(4\pi(\sigma^2 + t_2 - t_1))^{n/2}.
\end{aligned}$$

That is

$$\lambda_{\sigma^2}(g(t_2)) \geq \lambda_{\sigma^2 + t_2 - t_1}(g(t_1)). \tag{8.2.35}$$

*Step 3.* We estimate the change of the best constant in the log Sobolev inequality in the time interval $[T_1, T_2]$, with surgeries.

Now, let

$$T_1 \leq t_1 < t_2 < \ldots < t_k \leq T_2$$

and $t_i$, $i = 1, 2, \ldots, k$ be all the surgery times from $T_1$ to $T_2$. Here, without loss of generality, we assume that $T_1$ and $T_2$ are not surgery times. Otherwise we can just directly apply step 1 two more times at $T_1$ and $T_2$. We also fix a

$$\sigma_0 = T_2 - T_1 + 1,$$

where $\sigma_0$ is the upper bound for the parameter $\sigma$ in step 1, (8.2.33).

For any $\sigma \in (0, 1]$, by (8.2.35), we have

$$\lambda_{\sigma^2}(g(T_2)) \geq \lambda_{\sigma^2 + T_2 - t_k}(g(t_k^+)).$$

By Step 1 (8.2.33), either

$$\lambda_{\sigma^2 + T_2 - t_k}(g(t_k^+)) \geq \Lambda_0$$

or

$$\lambda_{\sigma^2+T_2-t_k}(g(t_k^+)) \ge \lambda_{\sigma^2+T_2-t_k}(g(t_k^-)) - c_{12}|vol(\mathbf{M}(t_k^-) - vol(\mathbf{M}(t_k^+))|.$$

In the first case, we have

$$\lambda_{\sigma^2}(g(T_2)) \ge \Lambda_0.$$

So a uniform lower bound is already found.

In the second case,

$$\lambda_{\sigma^2}(g(T_2)) \ge \lambda_{\sigma^2+T_2-t_k}(g(t_k^-)) - c_{12}|vol(\mathbf{M}(t_k^-) - vol(\mathbf{M}(t_k^+))|.$$

From here we start with $\lambda_{\sigma^2+T_2-t_k}(g(t_k^-))$ and repeat the above process. We have, from (8.2.35), with $\sigma^2$ in (8.2.35) replaced by $\sigma^2 + T_2 - t_k$,

$$\lambda_{\sigma^2+T_2-t_k}(g(t_k^-)) \ge \lambda_{\sigma^2+T_2-t_{k-1}}(g(t_{k-1}^+)).$$

Continue like this, until $T_1$, we have either

$$\lambda_{\sigma^2}(g(T_2)) \ge \lambda_{\sigma^2+T_2-T_1}(g(T_1)) - c_{12}\Sigma_{i=1}^k|vol(\mathbf{M}(t_i^-) - vol(\mathbf{M}(t_i^+))| \tag{8.2.36}$$

or

$$\lambda_{\sigma^2}(g(T_2)) \ge \Lambda_0 - c_{12}\Sigma_{i=1}^k|vol(\mathbf{M}(t_i^-) - vol(\mathbf{M}(t_i^+))|. \tag{8.2.37}$$

Note that the above process can be carried out since all the parameters under $\lambda$ is bounded from above by $\sigma_0$.

It is known that

$$\Sigma_{i=1}^k|vol(\mathbf{M}(t_i^-) - vol(\mathbf{M}(t_i^+))| \le \sup_{t\in[T_1,T_2]} vol(\mathbf{M}(t)).$$

Hence, either

$$\lambda_{\sigma^2}(g(T_2)) \ge \lambda_{\sigma^2+T_2-T_1}(g(T_1)) - c_{12} \sup_{t\in[T_1,T_2]} vol(\mathbf{M}(t)), \tag{8.2.38}$$

or

$$\lambda_{\sigma^2}(g(T_2)) \ge \Lambda_0 - c_{12} \sup_{t\in[T_1,T_2]} vol(\mathbf{M}(t)). \tag{8.2.39}$$

In either case, the lower bound is independent of the number of surgeries.

If (8.2.38) holds, then we have to find a lower bound for $\lambda_{\sigma^2+T_2-T_1}(g(T_1))$, which is independent of $\sigma$. Remember that it is assumed that $(\mathbf{M}, g(T_1))$ satisfies a Sobolev inequality with constant $A_1$. It is well known that this implies a log Sobolev inequality. Indeed, from

$$\left(\int v^{2n/(n-2)} d\mu(g(T_1))\right)^{(n-2)/n} \leq A_1 \int (4|\nabla v|^2 + Rv^2) d\mu(g(T_1)) + A_1 \int v^2 d\mu(g(T_1)),$$

using Hölder inequality and Jensen inequality for ln, we have:

for those $v \in W^{1,2}(\mathbf{M}, g(T_1))$ such that $\|v\|_2 = 1$, it holds

$$\int v^2 \ln v^2 d\mu(g(T_1)) \leq \frac{n}{2} \ln \left( A_1 \int (4|\nabla v|^2 + Rv^2) d\mu(g(T_1)) + A_1 \right). \tag{8.2.40}$$

Recall the elementary inequality: for all $z, q > 0$,

$$\ln z \leq qz - \ln q - 1.$$

By (8.2.40), this shows

$$\int v^2 \ln v^2 d\mu(g(T_1)) \leq \frac{n}{2} q \left( A_1 \int (4|\nabla v|^2 + Rv^2) d\mu(g(T_1)) + A_1 \right) - \frac{n}{2} \ln q - \frac{n}{2}.$$

Take $q$ such that $\frac{n}{2} q A_1 = \sigma^2 + T_2 - T_1$. Since $\sigma \leq 1$, this shows, for some

$$B = B(A_1, T_1, T_2, n) = c\ (T_2 - T_1) + c > 0,$$

$$\begin{aligned} \lambda_{\sigma^2+T_2-T_1}(g(T_1)) \equiv \inf_{\|v\|_2=1} \int [\ (\sigma^2 + T_2 - T_1)(4|\nabla v|^2 + Rv^2) \\ - v^2 \ln v^2] d\mu(g(T_1)) - \frac{n}{2} \ln(\sigma^2 + T_2 - T_1) \\ \geq -B. \end{aligned}$$

Therefore we can conclude from (8.2.38) and (8.2.39) that

$$\lambda_{\sigma^2}(g(T_2)) \geq \min\{-B, \Lambda_0\} - c_{12} \sup_{t \in [T_1, T_2]} vol(\mathbf{M}(t)) \equiv A_2$$

for all $\sigma \in (0, 1]$. By Definition 8.2.2, this is nothing but a (restricted) log Sobolev inequality for $(\mathbf{M}, g(T_2))$, i.e.

$$\int v^2 \ln v^2 d\mu(g(T_2)) \leq \sigma^2 \int (4|\nabla v|^2 + Rv^2) d\mu(g(T_2)) - \frac{n}{2} \ln \sigma^2 - A_2 \tag{8.2.41}$$

where $\sigma \in (0, 1]$. Obviously this implies the full log Sobolev inequality: for all $\epsilon > 0$,

$$\int v^2 \ln v^2 d\mu(g(T_2)) \le \epsilon^2 \int (4|\nabla v|^2 + Rv^2) d\mu(g(T_2)) - \frac{n}{2} \ln \epsilon^2 + \epsilon^2 - A_2 + C \tag{8.2.42}$$

where $C$ is a numerical constant.

*Step 4.* The log Sobolev inequality (8.2.41) or (8.2.42) implies certain heat kernel estimate.

Let $p(x, t, y)$ be the heat kernel of $\Delta - \frac{1}{4}R$ in $(\mathbf{M}, g(T_2))$ (with the fixed metric $g(T_2)$). Then (8.2.41) or (8.2.42) implies, for $t \in (0, 1]$,

$$p(x, t, y) \le \exp(4(T_2 + 1) + \frac{n}{2}|\ln|A_2|| + c + R_0^-) \frac{1}{(4\pi t)^{n/2}} \equiv \frac{\Lambda}{t^{n/2}}. \tag{8.2.43}$$

Here $R_0 = \sup R^-(x, 0)$ again. This follows from a generalization of Davies' argument [Da], as done in Theorem 6.2.1.

*Step 5.* The heat kernel estimate (8.2.43) implies Sobolev inequality perturbed with scalar curvature $R$ and strong noncollapsing.

This is more or less routine. By adapting the standard method in heat kernel estimate in [Da], as demonstrated in the proof of Theorem 6.2.1, it is known that (8.2.43) implies the desired Sobolev imbedding for $g(T_2)$, i.e. for all $v \in W^{1,2}(\mathbf{M}, g(T_2))$, there is $B_2 > 0$,

$$\begin{aligned}\left( \int v^{2n/(n-2)} d\mu(g(T_2)) \right)^{(n-2)/n} \le & B_2 \int (4|\nabla v|^2 + Rv^2) d\mu(g(T_2)) \\ & + B_2 \int v^2 d\mu(g(T_2)).\end{aligned}$$

The strong noncollapsing result follows from the work of Carron [Ca] and Akutagawa [Ak], as described in Chapter 4.

The last statement in the theorem is a result of (8.2.34). □

# Chapter 9

# Applications to the proof of Poincaré conjecture

In this chapter we will use the strong noncollapsing result (Theorem 8.2.1) to clarify and simplify another key part of the proof of the Poincaré conjecture, namely that there are only finitely many surgeries in finite time.

The main work is to prove a strong canonical neighborhood property for Ricci flow with surgeries (Theorem 9.2.1, part (1) below) that mirrors Theorem 7.5.1, which deals with the case without surgeries. In this chapter, unless stated otherwise, when we use the term canonical neighborhood property with or without the prefix "strong", we mean the one defined as part of the a priori assumption in Definition 8.1.6.

Our proof seems to simplify and clarify Perelman's original proof both logically and technically. Recall Perelman's proof in [P2] of Theorem 9.2.1, part (1). One first assumes that such a strong canonical neighborhood property for Ricci flow with surgeries holds with a worse accuracy $2\epsilon$. Then one proves Lemma 9.1.1 below ( [P2], Lemma 4.5), which describes the evolution of the surgery cap. Using this lemma, one proves the (weak) $\kappa$ noncollapsing property. Then one finally proves the strong canonical neighborhood property with surgeries with accuracy $\epsilon$, using Lemma 9.1.1, the (weak) $\kappa$ noncollapsing property and the method in Theorem 7.5.1. In contrast, we already know that the strong noncollapsing result holds as long as one performs $(r, \delta)$ surgery with sufficiently small $\delta$. This allows us to prove Lemma 9.1.1 directly with a different method. The proof of Lemma 9.1.1 outlined in [P2] seems in need of further justification, as explained in Claim 9.1.1 in Step 2 of the proof of Lemma 9.1.1 below. A combination of this lemma with the method in Theorem 7.5.1 leads to the strong canonical neighborhood property with accuracy $\epsilon$. This method also allows one to bypass the sophisticated uniqueness theorem of [ChZ2].See Remark 9.1.2 below.

## 9.1 Evolution of regions near surgery caps

Let us begin with two preliminary results on Ricci flow with surgeries. One is called bounded curvature at bounded distance. The other is a surgery version of Hamilton's compactness theorem.

**Proposition 9.1.1** *(bounded curvature at bounded distance for Ricci flow with surgery) (Claim II [P2] 4.2. See also [MT] Theorem 10.2 and [KL] section 70.)*

*Let $(M, g)$ be a Ricci flow with surgery that satisfies the a priori assumption (Definition 8.1.6) with the canonical neighborhood parameter $r = r(t)$ and accuracy $\epsilon$.*

*Assume $\epsilon$ is sufficiently small. Then for any $A > 0$ and $t_0 > 0$, there exist positive constants $K = K(A, \epsilon, r(t_0))$ and $E = E(A, \epsilon, r(t_0))$ such that the following statement holds.*

*If $Q \equiv R(x_0, t_0) \geq E$, and for any $x \in M$, the space time segment*

$$\{x\} \times [t_0 - (\max(Q, R(x, t_0)))^{-1}, t_0]$$

*is unscathed by surgery, then*

$$R(x, t_0) \leq K(A, \epsilon, r(t_0))R(x_0, t_0), \quad x \in B(x_0, AQ^{-1/2}, t_0).$$

PROOF. This result is essentially contained in the proof of the singularity structure Theorem 7.5.1, Step 3. One difference is that there are surgeries here, which may interfere with the limiting process in the proof. However the existence of a suitable time interval and spatial domain where the Ricci flow is smooth is guaranteed by the a priori assumption which entails the strong canonical neighborhood property. So the previous method still works.

Here are the details of the proof.

Suppose the lemma is not true. Then there exists a sequence of Ricci flows $(M_k, g_k)$ and space time points $(x_k, t_0)$ with the following properties:

(i) The scalar curvature $R_k(x_k, t_0)$ tends to $\infty$ when $k \to \infty$;

(ii) There exists $A > 0$ and $z_k \in B(x_k, A, g_k(t_0))$ such that

$$R_k(z_k, t_0)R_k^{-1}(x_k, t_0) \to \infty, \quad k \to \infty.$$

By scaling and time shifting, we will take $t_0 = 0$ and assume that

$$Q_k \equiv R_k(x_k, t_0) = 1, \quad k = 1, 2, \ldots.$$

By (i), we may also assume that any point with (scaled) scalar curvature higher than 1 has a strong canonical neighborhood property. For the rest of the proof, we only deal with the scaled flows without changing the notations.

For all $\rho > 0$, define

$$\begin{aligned} J(\rho) &= \sup\{R_k(x, 0) \mid x \in B(x_k, \rho, g_k(0)), k = 1, 2, \ldots\}, \\ \rho_0 &= \sup\{\rho \mid J(\rho) < \infty\}. \end{aligned}$$

Integrating the spatial gradient bound in the a priori assumption, as in step 2 of the proof of Theorem 7.5.1, we deduce that $\rho_0 > 0$.

Since the lemma is assumed false, we know $\rho_0 < \infty$. We will show that this leads to a contradiction, which proves the lemma. By $\rho_0 < \infty$, after passing to a subsequence when necessary, there exists $y_k \in M_k$ such that $d(x_k, y_k, g_k(0)) \to \rho_0$ and $R_k(y_k, 0) \to \infty$. Let $\alpha_k$ be a minimizing geodesic from $x_k$ to $y_k$, lying in $M_k$. Since $R_k(x_k, 0) = 1$, there exists a point $z_k \in \alpha_k$ such that $R_k(z_k, 0) = 2$ and that $z_k$ is the closest such point to $y_k$. We use $\beta_k$ to denote the segment of $\alpha_k$ that connects $z_k$ and $y_k$. Then the scalar curvature $R_k \geq 2$ along $\beta_k$. The pinching property, as part of the a priori assumption, tells us that the curvature tensor is bounded from below. Therefore, for each fixed $\rho < \rho_0$, the curvature tensor of $(M_k, g_k)$ is uniformly bounded on the balls $B(x_k, \rho, g_k(0))$. The injectivity radii are also uniformly bounded away from zero due to the strong noncollapsing theorem. Hence Hamilton's compactness Theorem 5.3.5 implies the following. The marked (sub)sequence $(B(x_k, \rho_0, g_k(0)), g_k(0), x_k)$ converges in the $C_0^\infty$ topology to a marked incomplete manifold $(B_\infty,\ g_\infty(0), x_\infty)$. Also the incomplete Ricci flows $(M_k, g_k)$ in the region

$$P_k \equiv \{(x, t) \mid d(x_k, x, g_k(0)) < \rho_0,\ -(\max(1, R_k(x, 0)))^{-1} < t \leq 0\}$$

with marked space time point $(x_k, 0)$, converges, in subsequence, to a marked incomplete Ricci flow with metric $g_\infty(t)$ in the region:

$$P_\infty \equiv \{(x, t) \mid x \in B_\infty,\ -(\max(1, R_\infty(x, 0)))^{-1} < t \leq 0.\}$$

Here $R_\infty$ is the scalar curvature of $g_\infty$. Note that $P_k$ is unscathed by surgery by assumption. Hence the limit process is meaningful. Notice also the geodesic segment $\alpha_k$ converges to a geodesic segment $\alpha_\infty \subset B_\infty$ and $\beta_k$ to $\beta_\infty$. The common end point of $\alpha_\infty$ and $\beta_\infty$ is denoted by $y_\infty$.

Observe that the scalar curvature $R_\infty$ along $\beta_\infty$ is at least 2. This corresponds to high scalar curvature for the flows $(M_k, g_k)$, where the strong canonical neighborhood property with accuracy $\epsilon$ holds as part of the a priori assumption. Hence, for any $q_0 \in \beta_\infty$, the ball

$$D_\infty(q_0) \equiv \{q \in B_\infty \mid d^2(q_0, q, g_\infty(0)) < \epsilon^{-2}[R_\infty(q_0)]^{-1}\}$$

is $2\epsilon$ close to either a time slice of strong $\epsilon$ necks, $\epsilon$ caps or compact manifolds without boundary. Also $R_\infty$ becomes unbounded when the curve $\alpha_\infty$ approaches the end point $y_\infty$. This shows $D_\infty(q_0)$ is not close to a compact manifold. Since $\alpha_\infty$ is distance minimizing, we know that $D_\infty(q_0)$ can not be close to an $\epsilon$ cap either. The reason is that any long geodesic going through the center of a cap can not be distance minimizing. So the only possibility is $D_\infty(q_0)$ is close to a time slice of a strong $\epsilon$ neck.

Therefore the limit manifold $(M_\infty, g_\infty(0), x_\infty)$, as a union of $3\epsilon$ necks, is diffeomorphic to $S^2 \times (0,1)$. The sectional curvature is non-negative due to Hamilton-Ivey pinching Theorem 5.2.4 and the scalar curvature tends to infinity as a point in $B_\infty$ tends to the end point $x_\infty$.

The rest of the proof is identical to that of step 3 in Theorem 7.5.1. One scales the limit flow in $P_\infty$ to get a metric cone, which has to split by the maximum principle. This leads to contradiction with the assumption that $\rho_0 < \infty$. □

The next result is a surgery version of the Hamilton compactness theorem for Ricci flow. The reader can also consult [MT] Theorem 11.1 for its proof.

**Proposition 9.1.2** *Let $(M^\alpha, g^\alpha, (x^\alpha, t^\alpha))$ be a sequence of marked 3-dimensional Ricci flows with surgeries, which satisfy the a priori assumption. Suppose that:*

*(i) For each $y^\alpha \in M^\alpha$ and $t \le t^\alpha$ such that $R(y^\alpha, t) \ge 4R(x^\alpha, t^\alpha)$, there exists a strong canonical neighborhood with accuracy $\epsilon$.*

*(ii)* $\lim_{\alpha\to\infty} Q_\alpha = \infty$ *where $Q_\alpha = R(x^\alpha, t^\alpha)$.*

*(iii) For each $A < \infty$, if $\alpha$ is sufficiently large, then the ball $B(x^\alpha, AQ_\alpha^{-1/2}, t^\alpha)$ has a compact closure in $M^\alpha$.*

*(iv) For uniform constants $\kappa$ and $r$, the flow $(M^\alpha, g^\alpha)$ is $\kappa$ noncollapsed on scales less than $r$ at each point of the ball $B(x^\alpha, AQ_\alpha^{-1/2}, t^\alpha)$.*

*(v) There is $\mu > 0$ such that for each $A > 0$, any point $y^\alpha \in B(x^\alpha, AQ_\alpha^{-1/2}, t^\alpha)$ is unscathed by surgery in the time interval $[t^\alpha - (max(Q_\alpha, R(y^\alpha, t^\alpha)))^{-1}\mu, t^\alpha]$.*

*Then, after passing to a subsequence and shifting the time $t^\alpha$ to 0, the sequence*

$$(M^\alpha, Q_\alpha g^\alpha, (x^\alpha, t^\alpha))$$

*converges in $C^\infty_{loc}$ topology to a complete Ricci flow $(M^\infty, g^\infty, (x^\infty, 0))$. The limiting Ricci flow exists in a time interval $[-t_0, 0]$ for some $t_0 > 0$. Moreover the limiting flow has bounded and nonnegative curvature.*

Proof. By assumption (v), the space time domain

$$\begin{aligned} P^\alpha_A \equiv &\{(x,t) \,|\, d(x^\alpha, x, g^\alpha(t^\alpha)) \\ &< AQ_\alpha^{-1/2},\, t^\alpha - (max(Q_\alpha, R(y^\alpha, t^\alpha)))^{-1}\mu \le t \le t^\alpha\} \end{aligned}$$

is unscathed by surgery. The previous proposition tells us that there exists a positive constant $K = K(A)$ such that

$$R(x, t^\alpha) \le K(A) Q_\alpha, \quad d(x^\alpha, x, g^\alpha(t^\alpha)) < AQ_\alpha^{-1/2}.$$

By the noncollapsing assumption (iv), we know, from Theorem 5.3.5 again, that after passing to a subsequence and shifting the time $t^\alpha$ to 0, the sequence

$$(M^\alpha, Q_\alpha g^\alpha, (x^\alpha, t^\alpha))$$

converges in $C^\infty_{loc}$ topology to a Ricci flow $(M^\infty, g^\infty, (x^\infty, 0))$. This Ricci flow exists in a space time subregion of $M^\infty \times (-\infty, 0]$, which is relatively open and contains the time slice $M^\infty \times \{0\}$.

By step 4 of the proof of Theorem 7.5.1, the scalar curvature of $M^\infty$ under $g^\infty(0)$ is bounded. Hence assumption (v) implies that the limit flow actually exists in a fixed time interval $[-t_0, 0]$. The curvature is nonnegative due to Hamilton Ivey pinching. This proves the proposition. □

**Remark 9.1.1** *It is convenient and desirable that condition (iv) holds, i.e. every point in the ball $B(x^\alpha, AQ_\alpha^{-1/2}, t^\alpha)$ is $\kappa$ noncollapsed for a fixed $\kappa$ and under a fixed scale. One may think that it suffices to assume only the point $(x^\alpha, t^\alpha)$ is $\kappa$ noncollapsed. Then one uses the classical volume comparison theorem and a priori curvature upper bound from the previous proposition to obtain lower bound for the volume of balls centered elsewhere. However, the constant $\kappa$ obtained in this manner is* *no longer uniform with respect to A. Moreover there could be complications when surgeries are present.*

The next result ( [P2] Lemma 4.5.) gives a description of the evolution of surgery caps. It is a key lemma in proving the long time existence of Ricci flow with surgery. In fact it can be regarded as a warm-up run for the canonical neighborhood Theorem 9.2.1 below. The proof here is quite different from the other ones available in the literature. One reason is the strong noncollapsing property proven in the last chapter. The other is the proof of Claim 9.1.1 below, which seems to need further justification in the proof outlined in [P2].

In order to present the lemma, we introduce one more notation. Let $(M, g(t))$ be a Ricci flow. Given $x_0 \in M$, $r > 0$ and two moments $t_0, t_1$ in time, we denote

$$P(x_0, t_0, r, t_0, t_1) \equiv \{(x,t) \mid x \in B(x_0, r, t), \ t \in [t_0, t_1]\}, \qquad (9.1.1)$$

if $t_1 > t_0$.

**Lemma 9.1.1** *(behavior of Ricci flow near a surgery cap) For any sufficiently small $\epsilon > 0$, any $A \geq 1$ and $\theta \in (0,1)$, there exists $\delta_0 = \delta_0(A, \theta, \epsilon) > 0$ such that the following result is valid.*

*Let $(M, g(t))$ be a Ricci flow with compact, orientable initial value and finitely many $(r, \delta)$ surgeries in the time interval $[0, T]$, which satisfies the following requirements.*

*(1) The strong canonical neighborhood property with parameter $r_0$ and accuracy $\epsilon$ holds in $[0, T]$ for some $r_0 > 0$.*

*(2) $\delta \leq \delta_0$ for all the $(r, \delta)$ surgeries incurred in $[0, T]$.*

*(3) One $(r, \delta)$ surgery with surgery radius $h$ occurs at time $T_0 \in (0, T)$.*

*(4) $x_0$ is any fixed point in the surgery cap in (3).*

*Then,*

*(i) There exists a time $T_1 \in (T_0, \min(T_0 + \theta h^2, T)]$ such that the Ricci flow in the space time region*

$$P(x_0, T_0, Ah, T_0, T_1) \equiv \{(x,t) \mid x \in B(x_0, Ah, t), \ t \in [T_0, T_1]\},$$

*is unscathed.*

*Moreover, after scaling by factor $h^{-2}$ and shifting time $T_0$ to 0, the Ricci flow in*

$$P(x_0, T_0, Ah, T_0, T_1)$$

*is $A^{-1}$ close, in $C^{[A]}_{loc}$ topology, to the corresponding portion of the standard solution which starts at time* 0 *and which includes the tip of the cap;*

*(ii) If $T_1 < \min(T_0+\theta h^2, T)$ and $T_1$ is the supremum of those times such that (i) holds, then the Ricci flow after time $T_1$ is undefined for all points in $B(x_0, Ah, T_0)$, i.e. the whole ball is cut off at time $T_1$ by a surgery.*

PROOF. The proof is divided into a few steps.

*Step 1.* Proof of (i).

By Lemma 8.1.2, at the surgery time $T_0$, the metric $g(T_0^+)$ immediately after the surgery in the ball $B(x_0, \delta^{-1/2}h, T_0)$ is, after scaling by $h^{-2}$, $\delta^{1/2}$ close to the metric in the corresponding ball of the standard capped cylinder. Recall that the standard solution is the Ricci flow whose initial value is the standard capped cylinder. We can choose $\delta$ sufficiently small so that $A << \delta^{-1/2}$. By continuity of Ricci flow in small time, conclusion (i) of the lemma holds for $P(x_0, T_0, Ah, T_0, t_1)$, for some time $t_1 \in (T_0, \min(T_0 + \theta h^2, T)]$.

Let $Q$ be the maximum of the scalar curvature of the standard solution up to time $\theta$. Since $\epsilon$ is sufficiently small, we know that

$$R(x,t) \le 2Qh^{-2}, \quad (x,t) \in P(x_0, T_0, Ah, T_0, t_1).$$

*Step 2.* We start the proof of (ii).

Let $T_1$ be the supremum of those $t_1$s in Step 1, i.e. the Ricci flow satisfies the requirements (1)–(4) and conclusion (i) of the lemma. Moreover $T_1$ is the supremum time with these properties.

If $T_1 = \min\{T_0 + \theta h^2, T\}$, then we are done with the proof. Thus we assume $T_1 < \min\{T_0 + \theta h^2, T\}$.

The task of this step is to prove the following:

**Claim 9.1.1** *Suppose $0 < \delta \le \delta_0(A, \theta, \epsilon)$ which is sufficiently small. Then there must be a surgery happening at $T_1$, which cuts off at least some point in the closed ball $\bar{B}(x_0, Ah, T_0)$.*

We caution that before the time $T_1$, many other surgeries could have occurred elsewhere. But they do not cut off any points in the ball $B(x_0, Ah, T_0)$. In step 3, we will show that the whole ball is cut off.

In the proof of the lemma outlined in [P2] page 10, it seems that this claim is taken for granted with little proof. However, there could be a case when no surgery appears at $T_1$ or shortly after $T_1$ but the $A^{-1}$ closeness to a standard solution fails. One may use a limiting method outlined in [P2] Lemma 4.5 to rule out the case. It is suggested that a sequence of flows violating this property will converge to a standard solution. However there seems to be a need to justify the limiting process, due to the possibility of surgery and lack of curvature control at any given distance from the tip of the cap. Here we present a detailed proof for the claim. The new ingredient in the proof is to show that if the scalar curvature is very high at a space time point, then it will not drop off very quickly. So canonical neighborhood persists over certain time interval.

We use the method of contradiction. Suppose the claim is false. Then, for each $\alpha = 1, 2, \ldots$, there exists $\delta_0 = \delta_0(\alpha)(\to 0)$, and a marked Ricci flow

$$(M^\alpha, g^\alpha, x_0^\alpha, T_0) \tag{9.1.2}$$

satisfying the requirements (1)–(4) and conclusion (i) of the lemma in a maximal interval $[T_0, T_1]$ where $T_0 = T_0(\alpha)$ and $T_1 = T_1(\alpha)$. But there is no surgery at time $T_1$, which scathes the closed ball $\bar{B}(x_0^\alpha, Ah(\alpha), T_0(\alpha))$. Here $h(\alpha)$ is the surgery radius at time $T_0(\alpha)$. By continuity, the following statement is true:

there exists $T_2(\alpha) > T_1(\alpha)$ such that the ball

$$B(x_0^\alpha, Ah(\alpha), T_0(\alpha)) \quad \text{is unscathed in the interval} \quad [T_0(\alpha), T_2(\alpha)]. \tag{9.1.3}$$

*Step 2.1.* We prove the following claim

**Claim 9.1.2** *With the assumption (9.1.3), the balls* $B(x_0^\alpha, \delta_0^{-1/2}(\alpha)$ $h(\alpha), T_0)$ *are also unscathed for all* $t \in [T_0(\alpha), T_2(\alpha)]$*, provided that* $T_2(\alpha)$ *is sufficiently close to* $T_1(\alpha)$*.*

To simplify the presentation, we will drop the parameter $\alpha$ in the next paragraphs, unless stated otherwise. An useful observation is that we are free to make $T_2(\alpha)(> T_1(\alpha))$ as close to $T_1(\alpha)$ as we wish.

Suppose for contradiction Claim 9.1.2 at this sub-step is false in this interval $[T_0, T_2]$ i.e. the ball $B(x_0, \delta_0^{-1/2} h, T_0)$ is scathed by a surgery at a time $t$ in the above interval. There are only two ways this can happen. One is that the whole ball is cut off by a surgery. The other is that a surgery two sphere intersects with the ball. But the first way is impossible since $x_0$ would be cut off too, contradicting the definition of $T_2$ in (9.1.3) just before Step 2.1. Therefore, there exists **a first time** $t_3 \in [T_0, T_2]$, when a surgery two sphere intersects with $B(x_0, \delta_0^{-1/2} h, T_0)$. Consequently

$$\text{there exists a point} \quad x_1 \in B(x_0, \delta_0^{-1/2} h, T_0), \tag{9.1.4}$$

which is at the center of a $\delta$ neck of radius $h_3$. Here $h_3$ is the surgery radius of the surgery at time $t_3$. Note $h_3$ is independent of $h$, the surgery radius at time $T_0$.

We show that there exists a positive constant $C = C(\eta)$, depending only on the coefficient $\eta$ in the gradient bounds of the strong canonical neighborhood property, such that

$$R(x,t) \geq C(\eta)h^{-2}, \quad x \in B(x_0, \delta_0^{-1/2}h, T_0),\ t \in [T_0, t_3]. \tag{9.1.5}$$

Here $t_3$ is regarded as the time right before the surgery takes place. To prove this, we recall from Step 1 first paragraph that, for small $\delta_0$,

$$R(x, T_0) > \frac{\sigma_0}{2}h^{-2}, \quad x \in B(x_0, \delta_0^{-1/2}h, T_0).$$

Here $\sigma_0 > 0$ is the lowest scalar curvature of the standard solution. By continuity of Ricci flow, there exists a time $t_4 \in [T_0, t_3]$ such that

$$R(x,t) \geq \frac{\sigma_0}{4}h^{-2}, \quad x \in B(x_0, \delta_0^{-1/2}h, T_0),\ t \in [T_0, t_4].$$

Here we recall that the Ricci flow is smooth in the space time region $B(x_0, \delta_0^{-1/2}h, T_0) \times [T_0, t_3^-]$, by definition of $t_3$ as the first surgery time in the ball.

Since $h$ is much smaller than the canonical neighborhood scale $r_0$, we see that the point $(x,t)$ specified above satisfies the canonical neighborhood property. In particular the gradient bound holds:

$$\partial_t R(x,t) \geq -\eta R^2(x,t).$$

After integration, this implies

$$R(x,t) \geq \frac{R(x,T_0)}{1 + \eta(t - T_0)R(x,T_0)}, \quad t \in [T_0, t_4].$$

Since $t - T_0 \leq \theta h^2 < h^2$ and $R(x, T_0) \geq \frac{\sigma_0}{2}h^{-2}$, we deduce

$$R(x,t) \geq C(\eta)h^{-2}, \quad t \in [T_0, t_4].$$

Note this bound is independent of $t_4$, as long as $t_4 - T_0 < h^2$. Hence we can iterate this procedure as many times as necessary to deduce (9.1.5).

Recall $h \leq \delta^2 r_0$. Hence $R(x, t_3) \geq C(\eta)h^{-2} \geq r_0^{-2}$ when $\delta$ is small. Thus every point $(x, t_3)$ with $x \in B(x_0, \delta_0^{-1/2}h, T_0)$ has a canonical neighborhood of accuracy $\epsilon$. So the ball $B(x_0, \delta_0^{-1/2}h, T_0)$ at time $t_3$ is entirely covered by canonical neighborhoods. By the definition of surgery procedure, we have thrown away all compact components with positive scalar curvature. So only $\epsilon$ caps and $\epsilon$ necks are present. By definition of $T_1$, the balls $B(x_0, Ah, T_0)$ at time $t \in [T_0, T_1]$ intersect a cap. So the ball $B(x_0, Ah, T_0)$ at $t_3$ also intersects a cap since either $t_3 \le T_1$ or it can be chosen close to $T_1$. Therefore, at time $t_3$, the ball $B(x_0, \delta_0^{-1/2}h, T_0)$ is covered by a $\epsilon$ cap and $\epsilon$ necks.

By (9.1.4), there exists $x_1 \in B(x_0, \delta_0^{-1/2}h, T_0)$, which lies in the surgery two sphere which is the center of a $\delta$ neck. Note $\delta$ is smaller than $\epsilon$. So a center of every $\delta$ neck is a part of an $\epsilon$ neck. Hence, at the moment after surgery, the smaller ball $B(x_0, Ah, T_0)$ at $t_3$ would lie in a compact manifold without boundary, which has positive scalar curvature. But this means the ball $B(x_0, Ah, T_0)$ at $t_3$ would be thrown out, contradicting (9.1.3). This contradiction proves Claim 9.1.2.

*Step 2.2.* Thus the Ricci flow (9.1.2), with index $\alpha$ omitted, remains smooth in the space time region

$$B(x_0, \delta_0^{-1/2}h, T_0) \times [T_0, T_2]$$

for some $T_2 > T_1$. Since we can always choose a slightly smaller $T_2$ if necessary, we can assume without loss of generality

$$T_2 \le \min(T_1 + (12\eta)^{-1}Q^{-1}h^2, T_0 + \theta h^2, T). \tag{9.1.6}$$

Here $\eta$ is the constant in the gradient bound of Definition 8.1.6.

By our assumption on $T_1$, when restricted to $B(x_0, Ah, T_1)$, the metric $g(T_1)$ is, after scaling by $h^{-2}$ and shifting time $T_0$ to 0, $A^{-1}$ close to the metric of the standard solution in the corresponding ball at time $(T_1 - T_0)h^{-2}$. Thus we know that $R(x_0, T_1) \le 2Qh^{-2}$. We show that a similar bound can be extended up to time $T_2$, i.e.

$$R(x_0, t) \le 3Qh^{-2}, \qquad t \in [T_1, T_2]. \tag{9.1.7}$$

This inequality is a quick consequence of the canonical neighborhood property. For $t$ specified in (9.1.7), if $R(x_0, t) \le 2Qh^{-2}$, then there is nothing to prove. So we assume $R(x_0, t) \ge 2Qh^{-2}$. Let

$$\bar{t} = \inf\{l \;|l \in [T_0, t],\; R(x_0, l) \ge 2Qh^{-2}\}.$$

Note $h$, as the surgery radius, is much smaller than $r_0$, the parameter in the canonical neighborhood property. Hence we know that each $(x_0, l)$,

$l \in [\bar{t}, t]$ has a canonical neighborhood. Therefore the gradient estimate holds:

$$\partial_l R(x_0, l) \le 2\eta R(x_0, l)^2.$$

Integrating this from $\bar{t}$ to $t$ and noticing $R(x_0, \bar{t}) = 2Qh^{-2}$, we obtain

$$R(x_0, t) \le \frac{1}{0.5Q^{-1}h^2 - 2\eta(t - \bar{t})} \le \frac{1}{0.5Q^{-1}h^2 - 2\eta(T_2 - T_1)}.$$

Therefore, when $T_2$ satisfies (9.1.6), the upper bound (9.1.7) holds, i.e.

$$R(x_0, t) \le 3Qh^{-2}, \quad t \in [T_1, T_2].$$

By the argument in Step 2.1, there is $c > 0$ such that

$$R(x_0, t) \ge cQh^{-2}, \quad t \in [T_0, T_2].$$

If $T_2 \le T_0 + (12\eta)^{-1}Q^{-1}h^2$, the above argument can be applied in the space time region

$$B(x_0, \delta_0^{-1/2}h, T_0) \times [T_0, T_2].$$

Hence, for these $T_2$, it holds

$$R(x, t) \le CQh^{-2}, \quad (x, t) \in B(x_0, \delta_0^{-1/2}h, T_0) \times [T_0, T_2]. \tag{9.1.8}$$

If $T_2 \ge T_0 + (12\eta)^{-1}Q^{-1}h^2$, by the upper and lower bound for $R(x_0, t)$, we have the necessary surgery free time interval to apply Proposition 9.1.1. It tells us that for any fixed $D > 0$, there exists $K(D, \epsilon, r_0) > 0$, such that the uniform bound

$$R(x, t) \le K(D, \epsilon, r_0)R(x_0, t), \quad \text{on} \quad \mathrm{B}(x_0, Dh, t) \times \{t\}, \quad t \in [T_0, T_2] \tag{9.1.9}$$

holds when the (hidden) $\alpha$ is sufficiently large. Here the uniformity is with respect to the parameter $\alpha$. We mention that Proposition 9.1.1 is stated for complete Ricci flow. But both the statement and proof is of local nature. Hence the result holds for local Ricci flows too. It is the local version that we are using here.

*Step 2.3.* Recover the index $\alpha$ and consider the scaled metric

$$\tilde{g}^\alpha(s) = h^{-2}(\alpha)g^\alpha(T_0(\alpha) + h^2(\alpha)s)$$

in the space time region

$$\begin{aligned}\tilde{P}(\alpha) \equiv &\{(x, s) \mid d(x, x_0^\alpha, \tilde{g}^\alpha(0)) < \delta_0^{-1/2}(\alpha), \\ &s \in [0, h^{-2}(\alpha)(T_2(\alpha) - T_0(\alpha))]\}.\end{aligned}$$

For any fixed $D > 0$, we want to show that there exists a positive constant $K_1(D, \epsilon, r_0)$ such that the uniform bound on the scalar curvature

$$\tilde{R}^\alpha \le K_1(D, \epsilon, r_0) \tag{9.1.10}$$

holds on the compact sets

$$\{(x, s) \mid d(x, x_0^\alpha, \tilde{g}^\alpha(0)) < D, \quad s \in [0, h^{-2}(\alpha)(T_2(\alpha) - T_0(\alpha))]\}.$$

This bound is a consequence of (9.1.9) and Proposition 5.1.1 (3). Here is the reason. Let $s_0 \in [0, 1]$ be a small number close to 0 so that

$$B(x_0^\alpha, D, \tilde{g}^\alpha(0)) \subset B(x_0^\alpha, 2D, \tilde{g}^\alpha(s)), \quad s \in [0, s_0].$$

For $x \in B(x_0^\alpha, 2D, \tilde{g}^\alpha(s))$, $s \in [0, s_0]$, by Hamilton Ivey pinching, we know from (9.1.9) that

$$|Ric_{\tilde{g}^\alpha}(x, s)| \le cK(2D, \epsilon, r_0)$$

for some $c > 0$. Therefore Proposition 5.1.1 (3) implies

$$\partial_s d(x, x_0^\alpha, \tilde{g}^\alpha(s)) \le cK(2D, \epsilon, r_0) d(x, x_0^\alpha, \tilde{g}^\alpha(s)).$$

Since $s \le 1$, we can integrate to get

$$d(x, x_0^\alpha, \tilde{g}^\alpha(s)) \le d(x, x_0^\alpha, \tilde{g}^\alpha(0)) e^{cK(2D, \epsilon, r_0)}, \quad x \in B(x_0^\alpha, D, \tilde{g}^\alpha(0)).$$

Thus, we know

$$B(x_0^\alpha, D, \tilde{g}^\alpha(0)) \subset B(x_0^\alpha, De^{cK(2D, \epsilon, r_0)}, \tilde{g}^\alpha(s)), \quad s \in [0, s_0].$$

This process can be repeated starting from $s_0$ and so on. Hence the above inclusion holds for all $s \in [0, 1]$. Now (9.1.10) follows from (9.1.9) by scaling.

By Theorem 8.2.1 in the previous chapter, the flows $\tilde{g}^\alpha$ are strong $\kappa$ noncollapsed under scale $h^{-1}(\alpha)$ in the region $\tilde{P}(\alpha)$. We mention that, due to the possible presence of surgeries elsewhere before $T_2$, we can not use Perelman's $\kappa$ noncollapsing result with surgeries here. In fact the proof of that result hinges on this lemma.

There are two cases to deal with.

Case 1 is when $\limsup_{\alpha \to \infty} h^{-2}(\alpha)(T_2(\alpha) - T_0(\alpha)) = a > 0$.

By (9.1.10) and the strong $\kappa$ noncollapsing property, we can use the localized version of Proposition 9.1.2 to conclude that a subsequence of $\{\tilde{g}^\alpha(s)\}$ converges, in $C^\infty_{loc}$ sense, to a complete Ricci flow $\tilde{g}^\infty$ in the time interval $[0, a]$. The completeness comes from the assumption that $\delta_0(\alpha) \to 0$ and each time slice of the region $\tilde{P}(\alpha)$ is a ball of radius $\delta_0^{-1/2}(\alpha)$ under $\tilde{g}^\alpha(0)$. It is easy to see that $\tilde{g}^\infty(0)$ is the initial metric of the standard solution. Following the proof of Theorem 7.5.1, Step 4, we know that the scalar curvature $\tilde{R}^\infty$ is bounded at each time slice. Note in doing so we need the $\kappa$ noncollapsing property again. The uniqueness theorem of Chen and Zhu (Theorem 5.1.3 here) then tells us that $\tilde{g}^\infty(s)$ is the standard solution in $[0, a]$.

**Remark 9.1.2** *Due to the strong $\kappa$ noncollapsing property from Theorem 8.2.1, the injectivity radii $\tilde{g}^\infty$ are bounded away from 0. Hence the uniqueness actually can be proven easily by following the proof of the compact case. The main work in Chen and Zhu's paper [ChZ2] is to overcome the lack of uniform lower bound of the injectivity radii in the noncompact case.*

Thus, for any $D > 1$, there exists arbitrarily large $\alpha$ such that $\tilde{g}^\alpha(s)$ in

$$\{(x, s) \mid d(x, x_0^\alpha, \tilde{g}^\alpha(0)) < D, \quad s \in [0, h^{-2}(\alpha)(T_2(\alpha) - T_0(\alpha))]\}.$$

is $A^{-1}$ close to a portion of the standard solution. For the standard solution in the time interval $[0, \theta]$, $\theta < 1$, the distances are comparable. Hence for each $A \geq 1$, we can find $D > 1$ such that

$$\begin{aligned} &\{(x, s) \mid d(x, x_0^\alpha, \tilde{g}^\alpha(s)) < A, \quad s \in [0, h^{-2}(\alpha)(T_2(\alpha) - T_0(\alpha))]\} \\ &\subset \{(x, s) \mid d(x, x_0^\alpha, \tilde{g}^\alpha(0)) < D, \quad s \in [0, h^{-2}(\alpha)(T_2(\alpha) - T_0(\alpha))]\}. \end{aligned}$$

Therefore the unscaled flow $g^\alpha$ in the region $P(x_0, T_0, Ah, T_0, T_2)$ is $A^{-1}$ close to the standard solution after scaling. But $T_2(\alpha) > T_1(\alpha)$, contradicting the assumption that $T_1$ is the supremum time this can happen. Hence, in this case, we have proven Claim 9.1.1 made at the beginning of Step 2.

Case 2 is when $\limsup_{\alpha\to\infty} h^{-2}(\alpha)(T_2(\alpha) - T_0(\alpha)) = 0$.

By (9.1.8), for large $\alpha$, the scalar curvature $\tilde{R}^\alpha$ corresponding to the metric $\tilde{g}^\alpha$ are uniformly bounded in the region

$$\{(x, s) \mid d(x, x_0^\alpha, \tilde{g}^\alpha(0)) < \delta_0^{-1/2}(\alpha), \quad s \in [0, h^{-2}(\alpha)(T_2(\alpha) - T_0(\alpha))]\}.$$

Then we can use Proposition 5.3.1 which is an improved version of Shi's local gradient estimate. It tells us that all derivatives of the scalar curvature are bounded uniformly on compact sets even though the time gap tends to zero. By the strong $\kappa$ noncollapsing property, we can find a subsequence of $\{\tilde{g}^\alpha\}$ that converges, in $C^\infty_{loc}$ sense, to the initial metric of the standard solution. Recall that $h^{-2}(\alpha)(T_2(\alpha) - T_0(\alpha)))$ goes to 0 when $\alpha \to \infty$. Hence there exists arbitrarily large $\alpha$ such that $g^\alpha(s)$, restricted in $P(x_0, T_0, Ah, T_0, T_2)$, and after scaling and shifting as before is $A^{-1}$ close to a portion of the standard solution. But $T_2(\alpha) > T_1(\alpha)$, which again contradicts the assumption that $T_1(\alpha)$ is the supremum time this can happen.

Therefore the Claim 9.1.1 is valid, i.e. a surgery at $T_1$ scathes the ball $\bar{B}(x_0, Ah, T_0)$.

*Step 3.* Let $(M, g(t))$ be the Ricci flow in the statement of the lemma. By definition of $T_1$, for any $s < T_1$, $(M, g(t))$ in the parabolic region $P(x_0, T_0, Ah, T_0, s)$, after scaling by factor $h^{-2}$ and shifting time $T_0$ to 0, is $A^{-1}$ close to the corresponding portion of the standard solution that includes the tip of the cap.

Therefore, for $s$ close to $T_1$, there exists a point $q \in B(x_0, Ah, T_0)$, which, as a point in the cap, is far from the center of a $\delta$ neck. In fact the distance is comparable to $\delta^{-1}h$. Note that the distance functions at times $s_1, s_2 \in [0, \theta]$ in a standard solution are comparable. Consequently the diameter of the ball $B(x_0, Ah, T_0)$ measured by $g(s)$ is bounded from above by $Q\delta^{-1/2}h$. Hence, at time $s$, the distance between every point in this ball $B(x_0, Ah, T_0)$ and a center of a $\delta$ neck is comparable to $\delta^{-1}h$.

By Step 2 (Claim 9.1.1), there is a point in $\bar{B}(x_0, Ah, T_0)$ that is cut off by a surgery at time $T_1$. Recall that the cut in any $(r, \delta)$ surgery happens at the center of a $\delta$ neck. By the last paragraph, the center of the $\delta$ neck can not intersect with the ball $B(x_0, Ah, T_0)$. Thus the whole ball $B(x_0, Ah, T_0)$ is cut off. □

## 9.2 Canonical neighborhood property for Ricci flow with surgeries

The following theorem states that the strong canonical neighborhood property with suitable surgeries still holds, with parameters depending only on the initial manifold and time, but not depending on the number of surgeries. This is a key fact in proving there are only finitely many surgeries in finite time. The proof is similar in spirit to that of Theorem 7.5.1, except that surgeries make the matter more complicated. The theorem is a combination of Proposition 5.1 [P2] and Theorem 8.2.1 here.

**Theorem 9.2.1** *(canonical neighborhood property with surgeries) Let* $(\mathbf{M}, g(0))$ *be a compact, orientable, normalized* 3 *manifold. For any sufficiently small* $\epsilon > 0$*, there exist nonincreasing, positive numbers* $\delta_i$*,* $r_i$ *and* $\kappa_i$*,* $i = 0, 1, 2 \ldots$*, such that the Ricci flow with* $(\mathbf{M}, g(0))$ *as the initial value satisfies the following:*

*(1) On the time interval* $[i\epsilon, (i+1)\epsilon]$*, the Ricci flow with finitely many* $(r_i, \delta_i)$ *surgeries, performed on strong* $\delta_i$ *necks, satisfies the strong canonical neighborhood property with accuracy* $\epsilon$ *and parameter* $r_i$*.*

*(2) On the time interval* $[i\epsilon, (i+1)\epsilon]$*, the Ricci flow is strong* $\kappa$ *noncollapsed with constant* $\kappa_i$*, under scale* 1*.*

PROOF. Once (1) is proven, then (2) follows immediately from Theorem 8.2.1 in the previous chapter. So we just need to give a proof of (1), which is divided into several steps. Recall that the strong canonical neighborhood property is defined as part of the a priori assumption (Definition 8.1.6).

*Step 1.* Setup of induction process.

The proof is by induction and contradiction. Since the initial manifold is normalized, the Ricci flow is smooth in a fixed amount of time. Therefore the theorem is true for $i = 0$ by the theory in smooth case.

Assume for induction that the theorem holds in the time interval $[(i-1)\epsilon, i\epsilon]$ for a positive integer $i$ and surgery parameters $r_i, \delta_i$ and surgery radius $h_i \le \delta_i^2 r_i$.

Now suppose for contradiction that part (1) of the theorem is false in the next time interval $[i\epsilon, (i+1)\epsilon]$. Then, there exists a sequence of Ricci flows $\{(M^\alpha, g^\alpha)\}$ with the following conditions:

(i) Each flow in the sequence satisfies the theorem up to the time $i\epsilon$,

(ii) There exist positive numbers $r^\alpha \to 0$, $\delta^\alpha \to 0$, $\alpha \to \infty$, such that $(M^\alpha, g^\alpha)$ undergoes finitely many $(r^\alpha, \delta^\alpha)$ surgeries in the time interval $[i\epsilon, T^\alpha]$ where $i\epsilon \le T^\alpha \le (i+1)\epsilon$.

Note $r^\alpha$ and $\delta^\alpha$ depend on two indices: $\alpha$ and $i$. However, if no confusion arises, we just use one index and ignore the index $i$ which defines the time interval. We may also ignore both indices all together if no confusions occur.

(iii) There exist $x^\alpha \in M^\alpha$, such that the strong canonical neighborhood property with accuracy $\epsilon$ and parameter $r^\alpha$ fails at the space time point $(x^\alpha, T^\alpha)$.

(iv) $T^\alpha$ is the smallest time such that (ii) and (iii) happen.

Let $(\tilde{M}^\alpha, \tilde{g}^\alpha, (\tilde{x}^\alpha, 0))$ be the Ricci flow obtained by scaling $(M^\alpha, g^\alpha)$

around the space time point $(x^\alpha, T^\alpha)$ by the factor $R^\alpha(x^\alpha, T^\alpha)$, and by shifting time by $T^\alpha$, i.e.

$$\tilde{g}^\alpha(\tilde{t}) = Q^\alpha g^\alpha(T^\alpha + \tilde{t}(Q^\alpha)^{-1})$$

where $Q^\alpha = R^\alpha(x^\alpha, T^\alpha)$ is the scalar curvature under the metric $g^\alpha$. We will show that a subsequence of $\{(\tilde{M}^\alpha, \tilde{g}^\alpha)\}$ converges to a $\kappa$ solution, contradicting the induction assumption (iii), and proving the theorem.

Let us collect some properties of $(M^\alpha, g^\alpha)$ and the scaled Ricci flow $(\tilde{M}^\alpha, \tilde{g}^\alpha)$.

Property 1. Since $T^\alpha$ is the first time for $(M^\alpha, g^\alpha)$ such that the canonical neighborhood property with accuracy $\epsilon$ breaks down, by continuity, we know that the strong canonical neighborhood property with accuracy $2\epsilon$ holds up to $T^\alpha$.

Property 2. By Theorem 8.2.1, $(M^\alpha, g^\alpha(t))$ is strong $\kappa$ noncollapsed in the interval $t \in [0, T^\alpha]$ under scale 1 with some constant $k_{i+1} > 0$.

Property 3. Denote by $\tilde{R}^\alpha$ the scalar curvature of $(\tilde{M}^\alpha, \tilde{g}^\alpha)$. Then $\tilde{R}^\alpha(\tilde{x}^\alpha, 0) = 1$ and

$$\tilde{r}^\alpha = r^\alpha \sqrt{R^\alpha(x^\alpha, T^\alpha)} \geq 1,$$

where $\tilde{r}^\alpha$ and $r^\alpha$ are the canonical neighborhood scale for $\tilde{g}^\alpha$ and $g^\alpha$ respectively. The reason is that the strong canonical neighborhood property for $(M^\alpha, g^\alpha)$ with scale $r^\alpha$ and accuracy $\epsilon$ breaks down at $(x^\alpha, T^\alpha)$. This implicitly says that $R^\alpha(x^\alpha, T^\alpha) \geq 1/(r^\alpha)^2 \to \infty$.

When attempting to take the limit, we face two possibilities.

Case 1. For every $A > 0$ and $b > 0$, there exists $\alpha$ sufficiently large, such that all points in the ball $B(\tilde{x}^\alpha, A, \tilde{g}^\alpha(0))$ are defined on the time interval $[-b, 0]$, i.e. there is no surgery cutting off any part of the ball.

Case 2. There exist $A > 0$ and $b = b(\alpha) > 0$ such that for each large $\alpha$, there is a point $y^\alpha$ in the ball $B(\tilde{x}^\alpha, A, \tilde{g}^\alpha(0))$, which is not defined before the time interval $[-b, 0]$, i.e. the point $y^\alpha$ is added as a point in a surgery cap at time $-b(\alpha)$.

*Step 2.* Case 1 is easy to deal with. Recall that the unscaled $(M^\alpha, g^\alpha(t))$ is strong $\kappa$ noncollapsed in the interval $t \in [0, T^\alpha]$ under scale 1 with constant $k_{i+1}$. Hence $(\tilde{M}^\alpha, \tilde{g}^\alpha(\tilde{t}))$ is strong $\kappa$ noncollapsed when $\tilde{t} \leq 0$ under scale $\sqrt{R^\alpha(x^\alpha, T^\alpha)}$ with the same constant $k_{i+1}$. Note we assumed the strong canonical neighborhood property breaks down at $(x^\alpha, T^\alpha)$. This implicitly says that $R(x^\alpha, T^\alpha) \geq 1/(r^\alpha)^2 \to \infty$. Thus the scale under which $(\tilde{M}^\alpha, \tilde{g}^\alpha(\tilde{t}))$ is $\kappa$ noncollapsed tends to $\infty$ when $\alpha \to \infty$. Also there is no surgery in the space time region of concern.

Hence we can just duplicate the proof of Theorem 7.5.1 (singularity structure theorem) to prove that a subsequence of $\{(\tilde{M}^\alpha, \tilde{g}^\alpha)\}$ converges to a $\kappa$ solution. Therefore, for sufficiently large $\alpha$, the unscaled flow $(M^\alpha, g^\alpha)$ at the point $(x^\alpha, T^\alpha)$ has a strong canonical neighborhood with accuracy $\epsilon$ and parameter $r^\alpha$, reaching a contradiction with induction condition (iii) in Step 1, proving the theorem.

*Step 3.* Thus we just need to deal with Case 2 from now on, i.e. there exist $A > 0$ and $b = b(\alpha) > 0$ such that for each large $\alpha$, there is a point $y^\alpha$ in the ball $B(\tilde{x}^\alpha, A, \tilde{g}^\alpha(0))$, which is not defined before the time interval $[-b(\alpha), 0]$, i.e. the point $y^\alpha$ is added as a point in a surgery cap at time $-b(\alpha)$.

In this step we follow the approach in [P2] as explained in [CZ], [KL] and [MT].

**Lemma 9.2.1** *Suppose for some $A > 0$ and $b = b(\alpha) > 0$, there is a point $y^\alpha$ in the ball $B(\tilde{x}^\alpha, A, \tilde{g}^\alpha(0))$, which is added as a point in a surgery cap at time $-b(\alpha)$. Suppose also the scalar curvature satisfies, for some constant $J > 0$,*

$$\tilde{R}^\alpha(y^\alpha, \tilde{t}) \le J, \quad \textit{for all large} \quad \alpha, \quad \tilde{t} \in [-b(\alpha), 0].$$

*Then for sufficiently large $\alpha$, the space time point $(\tilde{x}^\alpha, 0)$ in $(\tilde{M}^\alpha, \tilde{g}^\alpha, (\tilde{x}^\alpha, 0))$ has a strong canonical neighborhood with accuracy $\epsilon$.*

Proof. (of the lemma) During the proof, all quantities are for the scaled manifold $(\tilde{M}^\alpha, \tilde{g}^\alpha)$.

First we show that there is $\theta_1 = \theta_1(\epsilon, A, J, b(\alpha)) < 1$, depending on the upper bound of $b = b(\alpha)$, but otherwise independent of $\alpha$, such that for all large $\alpha$,

$$b(\alpha) \le (\tilde{h}_0^\alpha)^2 \theta_1. \tag{9.2.1}$$

Here $\tilde{h}_0^\alpha$ is the surgery radius of the surgery which added the point $y^\alpha$.

Recall for any $(r, \delta)$ surgery, the parameter $\delta$ is scaling invariant. By the induction condition (ii) in Step 1, the parameter $\delta_\alpha$ in the $(r, \delta)$ surgery tends to 0 when $\alpha \to \infty$. Also, for a fixed large $\alpha$ and fixed $\theta \in (0, 1)$, if

$$-b(\alpha) + (\tilde{h}_0^\alpha)^2 \theta \le 0, \tag{9.2.2}$$

then the space time point $(y^\alpha, -b(\alpha) + (\tilde{h}_0^\alpha)^2\theta)$ is unscathed by surgery. Hence we can apply Lemma 9.1.1, which implies that, the flow $(\tilde{M}^\alpha, \tilde{g}^\alpha)$

in the region

$$\begin{aligned} P(y^\alpha, -b(\alpha), A\tilde{h}_0^\alpha, -b(\alpha), -b(\alpha) + (\tilde{h}_0^\alpha)^2\theta) \\ = \{(z,l) \,|\, d(z, y^\alpha, \tilde{g}^\alpha(l)) < A\tilde{h}_0^\alpha, \\ - b(\alpha) < l < -b(\alpha) + (\tilde{h}_0^\alpha)^2\theta\} \end{aligned}$$

is, after scaling by $(\tilde{h}_0^\alpha)^{-2}$ and time shifting, $A^{-1}$ close in $C_{loc}^{[A]}$ topology, to a corresponding region of the standard solution. In particular the scalar curvature satisfies

$$\tilde{R}^\alpha(y^\alpha, -b(\alpha) + (\tilde{h}_0^\alpha)^2\theta) \geq \frac{1}{2} \min R_s(\cdot, \theta)(\tilde{h}_0^\alpha)^{-2}$$

where $R_s$ is the scalar curvature of the standard solution. By Lemma 8.1.3 (iii), we have

$$\min R_s(\cdot, \theta) \geq C/(1-\theta)$$

for some $C > 0$. Therefore, for any fixed $\theta \in (0, 1)$, when $\alpha$ is sufficiently large, it holds

$$\tilde{R}^\alpha(y^\alpha, -b(\alpha) + (\tilde{h}_0^\alpha)^2\theta) \geq \frac{C}{2(1-\theta)(\tilde{h}_0^\alpha)^2}. \tag{9.2.3}$$

For any $\theta_1 \in (1/2, 1)$, it holds $0 < 2\theta_1 - 1 < 1$. Hence we can find such a $\theta_1$ so that

$$\tilde{R}^\alpha(y^\alpha, -b(\alpha) + (2\theta_1 - 1)(\tilde{h}_0^\alpha)^2) \geq 4b(\alpha)J(\tilde{h}_0^\alpha)^{-2}. \tag{9.2.4}$$

In fact any $\theta_1 \in (1/2, 1)$ satisfying

$$\theta_1 \geq 1 - \frac{C}{16b(\alpha)J}$$

suffices. Therefore $\theta_1$, which depends on the upper bound of $b(\alpha)$, can be chosen independently of $\alpha$ otherwise.

If $b(\alpha) > \theta_1(\tilde{h}_0^\alpha)^2$, then $\theta_1 < b(\alpha)(\tilde{h}_0^\alpha)^{-2}$. Thus

$$\tilde{R}^\alpha(y^\alpha, -b(\alpha) + (2\theta_1 - 1)(\tilde{h}_0^\alpha)^2) \leq J \leq 2\theta_1 J < 2b(\alpha)J(\tilde{h}_0^\alpha)^{-2}.$$

This contradicts with the bound for the scalar curvature in (9.2.4). So

$$b(\alpha) \leq \theta_1(\tilde{h}_0^\alpha)^2 \tag{9.2.5}$$

as long as (9.2.2) holds. Let $\theta$ be the one given in (9.2.2). If $\theta > \theta_1$, then (9.2.2) and (9.2.5) can not hold at the same time. Therefore $\theta \le \theta_1$ which implies that (9.2.5) always hold, proving (9.2.1) i.e.

$$0 \le -b(\alpha) + (\tilde{h}_0^\alpha)^2 \theta_1.$$

This allows us to use Lemma 9.1.1 to deduce: For any large $A_1 > 0$, when $\alpha$ is sufficiently large, the flow $(\tilde{M}^\alpha, \tilde{g}^\alpha)$ in

$$P(y^\alpha, -b(\alpha), A_1 \tilde{h}_0^\alpha, -b(\alpha), 0)$$

is, after scaling by $(\tilde{h}_0^\alpha)^{-2}$ and shifting time by $-b(\alpha)$, $A_1^{-1}$ close, in $C_{loc}^{[A_1]}$ topology, to the corresponding portion of the standard solution in the time interval $[0, b(\alpha)(\tilde{h}_0^\alpha)^{-2}]$.

Since $A$ and $\epsilon$ are fixed here, we can choose $A_1$, which is much larger than $A$ and $\epsilon^{-1}$. As $y^\alpha$ at time $-b(\alpha)$ lies in the surgery cap, by definition of surgery, there exists a universal constant $c > 0$ such that the surgery radius $\tilde{h}_0^\alpha$ satisfies

$$c^{-1} \tilde{R}^\alpha(y^\alpha, -b(\alpha))^{-1/2} \le \tilde{h}_0^\alpha \le c \tilde{R}^\alpha(y^\alpha, -b(\alpha))^{-1/2}.$$

From the last paragraph, $(\tilde{h}_0^\alpha)^2 \tilde{R}^\alpha(y^\alpha, 0)$ is $A_1^{-1}$ close to the scalar curvature at a point in the standard solution at a time $\theta_2$, where $\theta_2 \le \theta_1$ which is bounded away from 1. Thus, there is a constant $C > 0$ such that

$$(\tilde{h}_0^\alpha)^2 \tilde{R}^\alpha(y^\alpha, 0) \ge \frac{C}{1 - \theta_1}.$$

By the assumed bound of $\tilde{R}^\alpha(y^\alpha, 0) \le J$, we know that $\tilde{h}_0^\alpha$ is bounded from below by a positive constant. Therefore the ball $B(y^\alpha, A_1 \tilde{h}_0^\alpha, \tilde{g}^\alpha(0))$ contains the point $\tilde{x}^\alpha$, since $d(\tilde{x}^\alpha, y^\alpha, \tilde{g}^\alpha(0)) < A$ and $A << A_1 \tilde{h}_0^\alpha$.

By Lemma 8.1.4, there are a few situations to deal with.

Situation 1. $(y^\alpha, 0)$ is in the tip of an evolving $A_1^{-1}$ cap and $(\tilde{x}^\alpha, 0)$ is in the tip of an evolving $2A_1^{-1}$ cap. Because $A_1 >> \epsilon^{-1}$ by choice, the later cap is a strong canonical neighborhood of accuracy (at least) $\epsilon$. Recall there is no requirement for the life span of a cap in the definition. The lemma is proven in this situation.

Situation 2. $(y^\alpha, 0)$ is in the tip of an evolving $A_1^{-1}$ cap and $(\tilde{x}^\alpha, 0)$ is a center of evolving strong $2A_1^{-1}$ neck. The lemma is also proven since $2A_1^{-1} < \epsilon$.

Situation 3. $(y^\alpha, 0)$ is in the tip of an evolving $A_1^{-1}$ cap and $(\tilde{x}^\alpha, 0)$ is a center of evolving $2A_1^{-1}$ neck which is not strong. Then we will deal with it together with the situation 5.

Situation 4. $(y^\alpha, 0)$ is a center of an evolving $A_1^{-1}$ neck.

Since $A_1^{-1}$ is much smaller than $A^{-1}$ or $\epsilon$, and $d(\tilde{x}^\alpha, y^\alpha, \tilde{g}^\alpha(0)) \le A$, we know that $(\tilde{x}^\alpha, 0)$ is a center of evolving $2A_1^{-1}$ neck. If the neck is strong, the lemma is proven. So there is only one situation left, which also includes Situation 3:

Situation 5. $(\tilde{x}^\alpha, 0)$ is at the center of an evolving $2A_1^{-1}$ neck which is not strong, i.e. the scaled life span is less than one.

However, we can extend the neck to earlier time to form a strong $\epsilon$ neck by virtue of the following:

Claim*(Proposition 15.2 [MT])*

*There is $\beta \in (0, 1/2)$ such that the following holds for any $\epsilon \in (0, 1)$:*

*Let $(N_1 \times [-t_1, 0], g_1(t))$ be an evolving $\beta\epsilon$ neck centered at $x$ with $R(x, 0) = 1$. Let $(N_2 \times [-t_2, -t_1], g_2(t))$ be a strong $\beta\epsilon/2$ neck centered at $(x, -t_1)$. Suppose $N_1 \times \{-t_1\} \subset N_2 \times \{-t_1\}$. Then the union*

$$(N_2 \times [-t_2, -t_1]) \cup (N_1 \times [-t_1, 0])$$

*with the induced metric contains a strong $\epsilon$ neck centered at $(x, 0)$.*

Taking the claim for granted, we proceed to prove the lemma. We choose $A_1$ so that $2A_1^{-1} < \beta\epsilon$ where $\beta$ is in the above claim. Observe that for sufficiently large $\alpha$, the space time point $(\tilde{x}^\alpha, -b(\alpha))$ lies in the center of a $\beta\epsilon/2$ neck (actually in a much finer strong $\delta$ neck) where a surgery takes place which added the point $y^\alpha$. Recall $y^\alpha$ is defined in Case 2 at the end of Step 1. Thus the neck must be strong $\beta\epsilon/2$ neck. The above claim then shows that $(\tilde{x}^\alpha, 0)$ is at the center of a strong $\epsilon$ neck. Now a contradiction is reached in all situations. This proves the lemma except for the claim.

For completeness we give a proof of the claim. Suppose the claim is not true. Then there is a sequence $\beta_i \to 0^+$ and sequences of counterexamples: $\beta_i\epsilon$ necks $(N_1^{(i)} \times [-t_1^{(i)}, 0], g_1^{(i)}(t))$ and strong $\beta_i\epsilon/2$ necks $(N_2^{(i)} \times [-t_2^{(i)}, -t_1^{(i)}], g_2^{(i)}(t))$, which satisfy the conditions of the claim but not the conclusion. Consider the unions

$$U^i \equiv N_2^{(i)} \times [-t_2^{(i)}, -t_1^{(i)}] \cup N_1^{(i)} \times [-t_1^{(i)}, 0],$$

These are incomplete Ricci flows which are glued together at time $-t_1^i$. By the assumption that $N_1^{(i)} \times \{-t_1\} \subset N_2^{(i)} \times \{-t_1\}$ and the smoothness of the necks in their respective time interval, we know, for $x$ in the interior of $N_1^{(i)}$ and $k = 0, 1, 2, \ldots$,

$$\lim_{s \to (-t^{(i)})^-} \partial_s \nabla^k g_2^{(i)}(x, s) = \lim_{s \to (-t^{(i)})^+} \partial_s \nabla^k g_1^{(i)}(x, s).$$

Hence the union $U^i$ is still a smooth solution of the Ricci flow in the interior of the space time domain. Observe that the scalar curvature at each space time point in $U^i$ are uniformly bounded by $1 + o(1)$. The reason is that the scalar curvature of a $\beta_i \epsilon$ neck is essentially nondecreasing in time when $\beta_i$ is sufficiently small, and at the largest time $t = 0$ the scalar curvature is a perturbation of 1 by assumption. Using Hamilton's compactness Theorem 5.3.5, we can pick a subsequence of $\{U^i, (x, 0)\}$ which converge, in $C^\infty_{loc}$ topology, to a limit Ricci flow. Since $\beta_i \to 0$, the limit flow is a complete portion of the standard shrinking cylinder: $S^2 \times R \times [-D, 0]$. Recall that $(N_2^{(i)} \times [-t_2^{(i)}, -t_1^{(i)}], g_2(t))$ are strong $\beta_i \epsilon$ necks. Also the scalar curvature is less than or equal to $1 + o(1)$. Therefore $|t_2^{(i)} - t_1^{(i)}| \geq 1 - o(1)$. Thus $D > 1$ and hence $U^i$ contains a strong $\epsilon$ neck when $i$ is large. This contradiction with the assumption proves the claim and the lemma. □

*Step 4.* We prove the following statement:

**Claim 9.2.1** *For any $A > 1$, there exists $b(A) > 0$ such that for all $\alpha$ sufficiently large, the space time region*

$$B(\tilde{x}^\alpha, A, \tilde{g}^\alpha(0)) \times [-b(A), 0]$$

*is unscathed by surgery.*

The key is that $b(A)$ is independent of $\alpha$.

*Step 4.1.* We prove the following lemma.

**Lemma 9.2.2** *Given any number $J \geq 1$, suppose for a point $y^\alpha \in B(\tilde{x}^\alpha, A, \tilde{g}^\alpha(0))$, it holds*

$$\tilde{R}^\alpha(y^\alpha, 0) \leq J.$$

*Then for large $\alpha$, there is no surgery scathing the segment*

$$\{y^\alpha\} \times [-cJ^{-1}, 0].$$

*i.e. $y^\alpha$ is not added as a point of a surgery cap in the time interval $[cJ^{-1}, 0]$. Here $c > 0$ is a universal constant.*

Proof. We use the method of contradiction again. Suppose the lemma is false. Then there is a sequence of $\{\alpha_k\}$, $\alpha_k \to \infty$ when $k \to \infty$ and a sequence of times $\{-b(\alpha_k)\}$ such that:

$$|b(\alpha_k)| \le 1/(kJ), \tilde{R}^{\alpha_k}(y^{\alpha_k}, 0) \le J \tag{9.2.6}$$

and the point $y^{\alpha_k}$ is added as a point in a surgery cap at time $-b(\alpha_k)$. Therefore, there exists a universal positive constant $c_1 > 0$ such that

$$\tilde{R}^{\alpha_k}(y^{\alpha_k}, -b(\alpha_k)) \ge c_1/(\tilde{h}^{\alpha_k})^2.$$

Here $\tilde{h}^{\alpha_k}$ is the scaled surgery radius. We consider two cases.

Case 1. $\tilde{R}^{\alpha_k}(y^{\alpha_k}, -b(\alpha_k)) \ge 4J$.

In this case, there exists a time $-b_1(\alpha_k) \in (-b(\alpha_k), 0)$, which is the closest time to 0 such that

$$\tilde{R}^{\alpha_k}(y^{\alpha_k}, -b_1(\alpha_k)) = 4J. \tag{9.2.7}$$

By continuity, there exists a time $-b_2(\alpha_k) > -b_1(\alpha_k)$ such that

$$\tilde{R}^{\alpha_k}(y^{\alpha_k}, s) \ge J, \qquad s \in (-b_1(\alpha_k), -b_2(\alpha_k)).$$

By Property 3 in Step 1, it holds

$$J \ge 1 \ge 1/(\tilde{r}^{\alpha_k})^2,$$

where $\tilde{r}^{\alpha_k}$ is the scaled parameter of the canonical neighborhood property with accuracy $2\epsilon$. Hence the following gradient bound holds for all the above $s$:

$$\partial_s \tilde{R}^{\alpha_k}(y^{\alpha_k}, s) \ge -\eta[\tilde{R}^{\alpha_k}(y^{\alpha_k}, s)]^2.$$

Integrating from $-b_1(\alpha_k)$ to $-b_2(\alpha_k)$ and using (9.2.7), we find

$$\tilde{R}^{\alpha_k}(y^{\alpha_k}, -b_2(\alpha_k)) \ge \frac{4J}{1 + \eta|b_1(\alpha_k) - b_2(\alpha_k)|4J}.$$

Note

$$|b_1(\alpha_k) - b_2(\alpha_k)| \le b(\alpha_k) \le 1/(kJ) \le 1/(4\eta J),$$

when $k$ is large. Hence

$$\tilde{R}^{\alpha_k}(y^{\alpha_k}, -b_2(\alpha_k)) \ge 2J.$$

This argument can be repeated as long as $-b_2(\alpha_k) < 0$. Therefore we know that

$$\tilde{R}^{\alpha_k}(y^{\alpha_k}, 0) \ge 2J,$$

which contradicts with the assumption that $\tilde{R}^{\alpha_k}(y^{\alpha_k}, 0) \le J$ in the lemma. Thus the lemma is proven in this case.

Case 2. $\tilde{R}^{\alpha_k}(y^{\alpha_k}, -b(\alpha_k)) < 4J$.

Then we can assume that

$$\tilde{R}^{\alpha_k}(y^{\alpha_k}, s) \le 4J, \quad s \in [-b(\alpha_k), 0]. \tag{9.2.8}$$

Otherwise there would be a time $-b_1(\alpha_k) \in (-b(\alpha_k), 0]$ such that

$$\tilde{R}^{\alpha_k}(y^{\alpha_k}, -b_1(\alpha_k)) = 4J.$$

Then we are in the same situation as Case 1.

By (9.2.8) and (9.2.6), we can apply Lemma 9.2.1 to make the following statement:

when $k$ is large, the space time point $(x^{\alpha_k}, T^{\alpha_k})$ in the unscaled flow $(M^{\alpha_k}, g^{\alpha_k})$ has a strong canonical neighborhood of accuracy $\epsilon$. This is a contradiction to the induction assumption (iii) in Step 1 of the proof of the theorem. So we have finished the proof of Lemma 9.2.2. □

*Step 4.2.* By Lemma 9.2.2 in Step 4.1, we can apply Proposition 9.1.1 to deduce:

For sufficiently large $\alpha$, there exists $K = K(\epsilon, A) > 0$ such that

$$\tilde{R}^{\alpha}(z^{\alpha}, 0) \le K(\epsilon, A), \qquad z^{\alpha} \in B(\tilde{x}^{\alpha}, A, \tilde{g}^{\alpha}(0)). \tag{9.2.9}$$

Here we note that Proposition 9.1.1 does not care about the largeness of the canonical neighborhood scale, which is not smaller than 1 here.

Lemma 9.2.2 and (9.2.9) then imply Claim 9.2.1 at the beginning of Step 4, i.e. for some $b(A) > 0$, independent of large $\alpha$,

$$B(\tilde{x}^{\alpha}, A, \tilde{g}^{\alpha}(0)) \times [-b(A), 0]$$

is unscathed by surgery.

*Step 5.* Finding a limit flow $(M^{\infty}, g^{\infty}, x^{\infty})$.

By the result in Step 4 and strong $\kappa$ noncollapsing, we may apply the compactness result Proposition 9.1.2. Hence there exists a subsequence $\alpha_k$ such that the flow $(\tilde{M}^{\alpha_k}, \tilde{x}^{\alpha_k}, \tilde{g}^{\alpha_k})$ converges, in $C^{\infty}_{loc}$ topology, to a smooth limit flow $(M^{\infty}, g^{\infty}, x^{\infty})$ which exists in an open set of the space time $M^{\infty} \times (-\infty, 0]$, containing $M^{\infty} \times \{0\}$. Now we are in the situation of the canonical neighborhood Theorem 7.5.1 in the smooth case. By Step 4 of that theorem, we know that the scalar curvature of $(M^{\infty}, x^{\infty}, g^{\infty}(0))$ is bounded.

*Step 6.* We show the limit flow $(M^\infty, g^\infty, x^\infty)$ exists in a time interval $[-B, 0]$ where $B$ is a positive constant.

Suppose the above statement is false. Then there exists a sequence $\{\alpha_k\}$, going to infinity with the following property:

For any small $\rho > 0$, there exists fixed $A > 1$ and arbitrarily large $k$ such that $y^{\alpha_k} \in \tilde{M}^{\alpha_k}$ satisfies $d(\tilde{x}^{\alpha_k}, y^{\alpha_k}, \tilde{g}^{\alpha_k}(0)) < A$. Moreover $y^{\alpha_k}$ is added during a surgery at time $-b(\alpha_k)$ such that $b(\alpha_k) < \rho$.

By Step 5, we can find a positive constant $Q$ which bounds the scalar curvature of $(M^\infty, g^\infty(0), x^\infty)$. Hence for all large $k$,

$$\tilde{R}^{\alpha_k}(y^{\alpha_k}, 0) \le 2Q.$$

By Lemma 9.2.2, there is no surgery scathing $\{y^{a_k}\} \times [-\rho, 0]$ when $\rho < cQ^{-1}$ for a positive constant $c$. This is a contradiction with the assumption that $y^{\alpha_k}$ is added at time $-b(\alpha_k) \in [-\rho, 0]$. Step 6 is complete.

*Step 7.* We show the limit flow $(M^\infty, g^\infty, x^\infty)$ exists in the time interval $(-\infty, 0]$.

Suppose not. Let $B$ be the maximum number such that the flow $(M^\infty, g^\infty, x^\infty)$ exists in a time interval $[-B, 0]$. By Hamilton Ivey pinching, the curvature operator of the limit flow is nonnegative. By the trace Harnack inequality Corollary 5.3.1,

$$R^\infty(x,t) \le \frac{B}{t+B} Q, \quad t \in [-B, 0].$$

Hence we only need to bound the scalar curvature when $t$ approaches $-B$. By the same proof as Step 5 of Theorem 7.5.1, we know that $R^\infty(x, -B)$ is finite for each $x \in M^\infty$. Now we can repeat Step 5 of this theorem to show that the limit flow $(M^\infty, g^\infty, x^\infty)$ exists in an open set of the space time $M^\infty \times (-\infty, 0]$, containing $M^\infty \times [-B, 0]$. Note the existance of the necessary surgery free region is guaranteed by lemma 9.2.1. Using Step 6 of the theorem, we know that $(M^\infty, g^\infty, x^\infty)$ exists in $M^\infty \times [-B - s, 0]$ for some $s > 0$. This contradicts with the definition that $-B$ is the earliest time that the limit flow can exist.

Thus the limiting flow $(M^\infty, g^\infty, x^\infty)$ is a $\kappa$ solution. This contradicts the induction condition (iii) in Step 1: for large $\alpha$, the flow $(M^\alpha, g^\alpha)$ at the space time point $(x^\alpha, T^\alpha)$ does not have a canonical neighborhood property of accuracy $\epsilon$. This finally proves the theorem. $\square$

## 9.3 Summary and conclusion

Let us close by presenting the flow chart of a simplified proof of the Poincaré conjecture without reduced distance or volume.

*Step 1.* $W$ entropy and its monotonicity [P1]. See also [Cetc], [CZ], [KL], [MT] and Chapter 6 of this book.

*Step 2.* Local noncollapsing result via Step 1 [P1]. See also [Cetc], [CZ], [KL], [MT] and Chapter 6 here.

*Step 3.* Getting ancient $\kappa$ solutions by blowing up of singularity using Step 2 and Hamilton's compactness theorem [P1]. See also [Cetc], [CZ], [KL], [MT] and Chapter 7 here.

*Step 4.* (i) Showing the backward limits of ancient $\kappa$ solutions are gradient shrinking solitons. Earlier work of Hamilton [Ha2] for type II case and Chapter 7 of this book for type I case.

(ii) Universal noncollapsing of ancient, nonround $\kappa$ solutions; noncompact case in Section 3.2 of [ChZ1]; both noncompact and compact case in Chapter 7 of the book.

(iii) Curvature and volume estimates for ancient solutions [P1], [P2]. See also [Cetc], [CZ], [KL], [MT] and Chapter 7 here.

*Step 5.* Classification of gradient shrinking solitons [P1], [P2]. See also [Cetc], [CZ], [KL], [MT] and Section 5.4 here.

*Step 6.* Canonical neighborhood property [P1]. That is: regions of high scalar curvature resemble the ancient $\kappa$ solution after appropriate scaling.

See also [Cetc], [CZ], [KL], [MT] and Chapter 7 here.

*Step 7.* Surgery procedure, including properties of the standard solution [P2]. See also [CZ], [KL], [MT] and Chapter 8 here.

*Step 8.* Finite time strong $\kappa$ noncollapsing with surgeries [Z4] or Chapter 8 here.

*Step 9.* Canonical neighborhood property with surgeries [P2]. See also [CZ], [KL], [MT] and this chapter.

*Step 10.* Existence of Ricci flow with surgeries, i.e. proving there are finitely many surgeries within finite time: this chapter. See also [P2], [CZ], [KL], [MT].

**Theorem 9.3.1** *(existence of Ricci flow with surgeries) Let* $(\mathbf{M}, g(0))$ *be a compact, orientable, normalized* 3 *manifold. For any sufficiently small* $\epsilon > 0$*, there exist nonincreasing, positive functions* $\delta = \delta(t)$*,* $r = r(t)$ *and* $\kappa = \kappa(t)$ *on* $\mathbf{R}^+$ *such that the Ricci flow with* $(\mathbf{M}, g(0))$ *as the initial value satisfies the following properties.*

*(i) The Ricci flow is smooth on the time interval* $[0, t_1)$ *for some* $t_1 > 0$*, and* $(\mathbf{M}, g(t))$ *with* $t \in [0, t_1)$*, satisfies the strong canonical neighborhood condition with accuracy* $\epsilon$ *and parameter* $r(t)$*.*

*(ii) Finitely many surgeries are done at* $t_1$*, with parameters* $(r(t_1), \delta(t_1))$*, performed on strong* $\delta_1$ *necks. With the post surgery manifold as the initial value at* $t_1$*, the Ricci flow is smooth on the time interval* $[t_1, t_2)$ *for some* $t_2 > t_1$*. Also* $(\mathbf{M}, g(t))$ *with* $t \in [t_1, t_2)$*, satisfies the canonical neighborhood condition with accuracy* $\epsilon$ *and parameter* $r(t)$*.*

*(iii) The Ricci flow can be extended in the above manner to a maximal time* $T_0$ *with surgeries at discrete times* $\{t_i\}$ *with parameters* $(r(t_i), \delta(t_i))$*,* $i = 1, 2, \ldots$*. The number of surgeries in a finite time interval is finite.*

*(iv) At any existence time* $t$*, the Ricci flow is strong* $\kappa$ *noncollapsed with constant* $\kappa(t)$*, under scale* $1$*.*

*(v) If* $T_0$ *is finite, the initial manifold is diffeomorphic to a connected sum of a finite copies of* $S^2 \times S^1$ *and metric quotients of the round three sphere* $S^3$*.*

Proof. Finally, only the last statement of (iii) and (v) remain to be proven. Let $[0, T]$ be a finite time interval when the Ricci flow with surgery exists.

By the maximum principle and the equation

$$\Delta R - \partial_t R + 2|Ric|^2 = 0,$$

when the Ricci flow is smooth, it holds

$$\inf_{x \in \mathbf{M}} R(x, t) \geq -3/[2(t + (3/2)].$$

Here we also used the assumption that the initial metric is normalized. Since a surgery is performed at a high scalar curvature region, this inequality survives surgeries. Let $V = V(t)$ be the volume of $(\mathbf{M}, g(t))$ at a nonsurgery time. Using the formula

$$\frac{d}{dt} V(t) = - \int R d\mu(g(t))$$

and the fact that surgery does not increase volume, we have

$$V(t) \leq C(1 + t)^{3/2}.$$

Here $C > 0$ is a constant. Let $h(T)$ be the infimum of surgery radii over the time period $[0, T]$. We have shown in Theorem 9.2.1

that the canonical neighborhood parameter (radius) $r(T)$ depends only on $T$, $\epsilon$ and the initial metric. Therefore we can choose $h(T) > 0$, which is independent of the number of surgeries. After each surgery, the manifold loses at least $ch(T)^3$ in volume. Hence, there are only finitely many surgeries in the interval $[0, T]$. This proves (iii).

Now we prove (v). Suppose the Ricci flow becomes extinct at time $T_0$. Then after finitely many surgeries the manifold is entirely covered by canonical neighborhoods. Hence the initial manifold is diffeomorphic to the connected sum of finite copies of metric quotients of $S^3$ and $S^2 \times S^1$. □

The final hurdle in proving the Poincaré conjecture is

*Step 11.* Finite time extinction of Ricci flow on simply connected manifolds [P3].

**Theorem 9.3.2** *Let* $(\mathbf{M}, g(0))$ *be a compact, orientable, normalized* 3 *manifold. Suppose the fundamental group of* $\mathbf{M}$ *is a free product of finite groups and infinite cyclic groups, then the Ricci flow with* $(\mathbf{M}, g(0))$ *as the initial value, given by Theorem 9.3.1, becomes extinct in finite time* $T_0$.

The idea of the proof is to use curve shortening flow and certain minimax functional. See [MT] Chapter 18 for a detailed exposition of [P3] and [CM] for a different proof using minimal surfaces.

# Conclusion

When $\mathbf{M}$ is simply connected, the Ricci flow becomes extinct in finite time. By Theorem 9.3.1 (v), the manifold $\mathbf{M}$ is diffeomorphic to $S^3$, as conjectured by Henri Poincaré. □

# Bibliography


►[Ak] Akutagawa, Kazuo. *Yamabe metrics of positive scalar curvature and conformally flat manifolds.* Differential Geom. Appl. 4 (1994), no. 3, 239–258.

[Ad] Adams, Robert A. *Sobolev spaces.* Pure and Applied Mathematics, Vol. 65. Academic Press New York, 1975.

[Au] Aubin, Thierry. *Problèmes isopérimétriques et espaces de Sobolev.* (French) J. Differential Geometry 11 (1976), no. 4, 573–598.

[Au2] Aubin, T. *Equations différentielles non linéaires et probléme de Yamabe concernant la courbure scalaire.* J. Math. Pures Appl. 55 (1976), 269–296.

[AB] Aronson, Donald G. and Bénilan, Philippe. *Régularité des solutions de l'équation des milieux poreux dans $R^N$.* (French. English summary) C. R. Acad. Sci. Paris Sér. A–B 288 (1979), no. 2, A103–A105.

[Br] Brendle, Simon. *A generalization of Hamilton's differential Harnack inequality for the Ricci flow.* J. Differential Geom. 82 (2009), no. 1, 207–227.

[BBI] Burago, D.; Burago, Y.; Ivanov, S. *A course in metric geometry*, AMS, Providence, RI, 2001.

►[BCLS] Bakry, D.; Coulhon, T.; Ledoux, M.; Saloff-Coste, L. *Sobolev inequalities in disguise.* Indiana Univ. Math. J. 44 (1995), no. 4, 1033–1074.

[BSSG] Bai, Z. G.; Shen, Y. B.; Shui, N. S.; Guo, X. Y. *An introduction to Riemann geometry,* China High Education Press, 2004, in Chinese.

[Ca] Carron, Gilles. *Inégalités isopérimétriques de Faber-Krahn et conséquences.* (French) Actes de la Table Ronde de Géométrie Différentielle (Luminy, 1992), 205–232, Sémin. Congr., 1, Soc. Math. France, Paris, 1996.

[CC] Carleson, L. and Chang, S. Y. A. *On the existence of an extremal function for an inequality of J. Moser.* Bull. Sc. Math. 110 (1986), 113–127.

[ChZ1] Chen, Bing-Long. Zhu, Xi-Ping; *Ricci flow with surgery on four-manifolds with positive isotropic curvature.* J. Differential Geom. 74 (2006), no. 2, 177–264.

[ChZ2] Chen, Bing-Long; Zhu, Xi-Ping. *Uniqueness of the Ricci flow on complete noncompact manifolds.* J. Differential Geom. 74 (2006), no. 1, 119–154.

▶[Cx] Cao, Xiaodong. *Dimension reduction under the Ricci flow on manifolds with nonnegative curvature operator.* Pacific J. Math. 232 (2007), no. 2, 263–268.

▶[Cx2] Cao, Xiaodong. *Differential Harnack estimates for backward heat equations with potentials under the Ricci flow.* J. Funct. Anal. 255 (2008), no. 4, 1024–1038.

[CaH] Cao, Xiaodong and Hamilton, Richard. *Differential Harnack estimates for time-dependent heat equations with potentials*, arXiv:0807.0568, GAFA 19 4 (2009), 989–1000.

[Cetc] Chow, B.; Chu, S-C; Glickenstein, D.; Guenter, C.; Isenberg, J.; Ivey, T.; Knopf, D.; Lu, P.; Luo, F.; and Ni, L. *The Ricci flow, techniques and applications. Part I, II*, AMS, 2007.

[Cha] Chavel, I. *Isoperimetric inequalities*, Cambridge U. Press, 2001.

[Cha2] Chavel, Isaac. *Riemannian geometry. A modern introduction. Second edition.* Cambridge Studies in Advanced Mathematics, 98. Cambridge University Press, Cambridge, 2006.

[CE] Cheeger, Jeff; Ebin, David G. *Comparison theorems in Riemannian geometry.* North-Holland Mathematical Library, Vol. 9. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, 1975.

[CG] Cheeger, Jeff; Gromoll, Detlef. *The splitting theorem for manifolds of nonnegative Ricci curvature.* J. Differential Geometry 6 (1971/72), 119–128.

▶[CG2] Cheeger, J. and Gromoll, Detlef. *On the structure of complete manifolds of nonnegative curvature*, Ann. Math. 46 (1972), 413–433.

[CGT] Cheeger, Jeff; Gromov, Mikhail; Taylor, Michael. *Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds.* J. Differential Geom. 17 (1982), no. 1, 15–53.

▶[ChLi] Chen, Wen Xiong; Li, Congming. *Classification of solutions of some nonlinear elliptic equations.* Duke Math. J. 63 (1991), no. 3, 615–622.

[CH] Chen, Xiuxiong and He, Weiyong. *The Calabi flow on Kähler surface with bounded Sobolev constant–(I)* arXiv:0710.5159.

[CwLx] Chen, Weiheng and Li, Xingxiao. *Riemanniann geometry, an introduction*, Beijing University Press, 2002, in Chinese.

[CK] Chow, B. and Knopf, D. *The Ricci flow, an introduction*, AMS, 2004.

[CL] Chow, Bennett; Lu, Peng. *On the asymptotic scalar curvature ratio of complete type I-like ancient solutions to the Ricci flow on noncompact 3-manifolds.* Comm. Anal. Geom. 12 (2004), no. 1–2, 59–91.

▶[CL2] Chow, Bennett; Lu, Peng. *The maximum principle for systems of parabolic equations subject to an avoidance set.* Pacific J. Math. 214 (2004), no. 2, 201–222.

[CaLo] Carlen, E. A. and Loss, M. *Sharp constant in Nash's inequality.* Internat. Math. Res. Notices (1993), no. 7, 213–215.

[CLN] Chow, B. Lu, P. and Ni, L. *Hamilton's Ricci flow*, Science Press Beijing and AMS, 2006.

▶[CLY] Cheng, Siu Yuen; Li, Peter; Yau, Shing Tung. *On the upper estimate of the heat kernel of a complete Riemannian manifold.* Amer. J. Math. 103 (1981), no. 5, 1021–1063.

▶[CM] Colding, Tobias H.; Minicozzi, William P., II. *Estimates for the extinction time for the Ricci flow on certain 3-manifolds and a question of Perelman*. J. Amer. Math. Soc. 18 (2005), no. 3, 561–569.

[CTY] Chau, Albert; Tam, Luen-Fai; Yu, Chengjie. *Pseudolocality for the Ricci flow and applications*, arXiv:math/0701153.

[CY] Cheng, S. Y. and Yau, S. T. *Differential equations and Riemannian manifolds and their geometric applications*. CPAM, vol. 28 (1975), 333–354.

[CZ] Cao, Huai-Dong and Zhu, Xi-Ping. *A complete proof of Poincare and geometrization conjectures-application of the Hamilton-Perelman theory of the Ricci flow*, Asian J. Math. 10, no. 2, (2006), 165–492.

▶[Da] Davies, E. B. *Heat kernel and spectral theory*, Cambridge University Press, 1989.

[DT] DeTurck, Dennis M. *Deforming metrics in the direction of their Ricci tensors*. J. Differential Geom. 18 (1983), no. 1, 157–162.

[DiTi] Ding, W. and Tian, G. *The generalized Moser-Trudinger inequality*, in Proceedings of the International Conference at Nankai Institute of Math. World Scientific, 1992, 57–70.

▶[Fo] Fontana, Luigi. *Sharp borderline Sobolev inequalities on compact Riemannian manifolds.* Comment. Math. Helv. 68 (1993), no. 3, 415–454.

[Gro] Guenther, C. *The fundamental solution on manifolds with time-dependent metrics*. J. Geom. Anal. 12 (2002), 425–436.

[Gr] Grigoryan, Alexander. *Gaussian upper bounds for the heat kernel on arbitrary manifolds*. J. Differential Geom. 45 (1997), no. 1, 33–52.

[Gr2] Grigor'yan, A. A. *The heat equation on noncompact Riemannian manifolds*. (Russian) Mat. Sb. 182 (1991), no. 1, 55–87; translation in Math. USSR-Sb. 72 (1992), no. 1, 47–77.

▶[Gro] Gross, L. *Logarithmic Sobolev inequalities*. Amer. J. Math. 97 (1975), no. 4, 1061–1083.

[GHL] Gallot, Sylvestre; Hulin, Dominique; Lafontaine, Jacques. *Riemannian geometry.* Third edition. Universitext. Springer-Verlag, Berlin, 2004.

[GNN] Gidas, B.; Ni, Wei Ming; Nirenberg, L. *Symmetry of positive solutions of nonlinear elliptic equations in* $R^n$. Mathematical analysis and applications, Part A, pp. 369–402, Adv. in Math. Suppl. Stud., 7a, Academic Press, New York-London, 1981.

[GT] Gilbarg, David; Trudinger, Neil S. *Elliptic partial differential equations of second order.* Second edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 224. Springer-Verlag, Berlin, 1983.

[Ha1] Hamilton, Richard S. *Three-manifolds with positive Ricci curvature.* J. Differential Geom. 17 (1982), no. 2, 255–306.

[Ha2] Hamilton, Richard S. *Four manifold with positive curvature operator.* J. Differential Geom. 24 (1986), 153–179.

[Ha3] Hamilton, Richard S. *Ricci flow on surfaces* Contemporary Math. 71 (1988), 237–261.

[Ha4] Hamilton, Richard S. *Eternal solutions to the Ricci flow.* J. Differential Geom. 38 (1993), no. 1, 1–11.

[Ha5] Hamilton, R. *A matrix Harnack estimate for the heat equation.* Comm. Analysis Geom. 1 (1993), 113–126.

[Ha6] Hamilton, Richard S. *The Harnack estimate for the Ricci flow.* J. Differential Geom. 37 (1993), no. 1, 225–243.

[Ha7] Hamilton, Richard S. *The formation of singularities in the Ricci flow.* Surveys in differential geometry, Vol. II (Cambridge, MA, 1993), 7–136, Int. Press, Cambridge, MA, 1995.

[Ha8] Hamilton, Richard S. *A compactness property for solutions of the Ricci flow.* Amer. J. Math. 117 (1995), no. 3, 545–572.

[Ha9] Hamilton, Richard S. *Nonsingular solutions of the Ricci flow on three manifolds*, Comm. Analysis and Geometry 7 (1999), 695–729.

▶[Heb1] Hebey, Emmanuel. *Optimal Sobolev inequalities on complete Riemannian manifolds with Ricci curvature bounded below and positive injectivity radius.* Amer. J. Math. 118 (1996), no. 2, 291–300.

[Heb2] Hebey, E. *Nonlinear Analysis on Manifolds: Sobolev Spaces and Inequalities*, Courant Lecture Notes, 1999.

[HV] Hebey, Emmanuel and Vaugon, Michel. *Meilleures constantes dans le théoréme d'inclusion de Sobolev.* (French) Ann. Inst. H. Poincaré Anal. Non Lin. 13 (1996), no. 1, 57–93.

[HL] Han and Lin. *Elliptic Differential Equations*, Courant Lecture Notes, 1997.

[Hs] Hsu, Shu-Yu. *Uniform Sobolev inequalities for manifolds evolving by Ricci flow,* arXiv:0708.0893.

▶[Iv] Ivey, T. *Ricci solitons on compact three manifolds.* Diff. Geom. Appl. 3 (1993), 301–307.

▶[Je] Jerison, D. *The Poincaré inequality for vector fields satisfying Hörmander's condition.* Duke Math. J. 53 (1986), 503–523.

[Jo] Jost, Jürgen. *Riemannian geometry and geometric analysis.* Fourth edition. Universitext. Springer-Verlag, Berlin, 2005. xiv+566.

[KaLi] Karp and P. Li, *unpublished.*

[KL] Kleiner, Bruce and Lott, John. *Notes on Perelman's papers*, http://arXiv.org/math.DG/0605667 v1(May 25, 2006).

[Ko] Kotschwar, Brett. *Hamilton's gradient estimate for the heat kernel on complete manifolds*, arXiv.org:math/0701335.

[KS] Kusuoka, S. and Stroock, D. *Applications of the Malliavin calculus, Part II.* J. Fac. Sci. Univ. Tokyo Sect. IA Math. 32 (1985), 1–76.

▶[KZ] Kuang, Shilong; Zhang, Qi S. *A gradient estimate for all positive solutions of the conjugate heat equation under Ricci flow.* J. Funct. Anal. 255 (2008), no. 4, 1008–1023. 53C44.

[Lj] Li, Jun-Fang. *Eigenvalues and energy functionals with monotonicity formulae under Ricci flow.* Math. Ann. 338 (2007), no. 4, 927–946.

[LL] Lieb, Elliott H.; Loss, Michael. *Analysis.* Second edition. Graduate Studies in Mathematics, 14. American Mathematical Society, Providence, RI, 2001.

[LS] Li, Peter; Schoen, Richard. *$L^p$ and mean value properties of subharmonic functions on Riemannian manifolds.* Acta Math. 153 (1984), no. 3–4, 279–301.

[LT] Lu, Peng and Tian, Gang. *Uniqueness of standard solutions in the work of Perelman*, preprint, 2005.

[LX] Li, Junfang; Xu, Xiangjin. *Notes on Perelman's differential Harnack inequality*, preprint.

[LY] Li, Peter; Yau, Shing-Tung. *On the parabolic kernel of the Schrödinger operator.* Acta Math. 156 (1986), no. 3–4, 153–201.

[LZ] Li, Junfang; Zhu, Meijun. *Sharp local embedding inequalities.* Comm. Pure Appl. Math. 59 (2006), no. 1, 122–144.

[Mo1] Moser, J. *A Harnack inequality for parabolic differential equations.* Comm. Pure Appl. Math. 17 (1964), 101–134. Correction in vol 20, 1967, 231–236.

[Mo2] Moser, J. *A sharp form of an inequality by N. Trudinger.* Indiana Univ. Math. J. 20 (1970/71), 1077–1092.

[Maz] Mazya, Vladimir G. *Sobolev spaces.* Translated from the Russian by T. O. Shaposhnikova. Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1985.

[MT] Morgan, John W. and Tian, Gang. *Ricci Flow and the Poincare Conjecture*, Clay Mathematics Monographs, 3. American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2007.

[Nab] Naber, A. *Noncompact shrinking 4-solitons with nonnegative curvature*, 2007, arXiv:0710.5579

[Ni] Ni, Lei. *Ricci flow and nonnegativity of sectional curvature.* Math. Res. Lett. 11 (2004), no. 5–6, 883–904.

[Ni2] Ni, Lei. *A note on Perelman's Li-Yau-Hamilton inequality.* Comm. Anal. Geom. 14 (2006), 883–905.

[Ni3] Ni, Lei. *Ancient solutions to Kähler-Ricci flow.* Math. Res. Lett. 12 (2005), no. 5–6, 633–653.

[NW] Ni, Lei; Wallach, Nolan. *On a classification of gradient shrinking solitons.* Math. Res. Lett. 15 (2008), no. 5, 941–955.

▶[On] Onofri, E. *On the positivity of the effective action in a theory of random surfaces.* Comm. Math. Phys. 86 (1982), no. 3, 321–326.

[P1] Perelman, Grisha. *The entropy formula for the Ricci flow and its geometric applications*, 11 Nov. 2002, http://arXiv.org/math.DG/0211159v1

[P2] Perelman, Grisha. *Ricci flow with surgery on three manifolds*, arXiv.org/ math.DG/0303109.

[P3] Perelman, Grisha. *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, arXiv:math/0307245

[P4] Perelman, Grisha. *Proof of the soul conjecture of Cheeger and Gromoll.* J. Differential Geom., 40 (1994), 209–212.

[Pet] Petersen, P. *Riemannian geometry*, Springer, 1997.

[PW] Petersen, P. and Wylie, W. *On the classification of gradient Ricci solitons*, preprint, 2007. arXiv 0712.1298

▶[Ro] Rothaus, O. S. *Logarithmic Sobolev inequalities and the spectrum of Schrödinger operators.* J. Funct. Anal. 42 (1981), no. 1, 110–120.

▶[Ru1] Rubinstein, Yanir A. *On the construction of Nadel multiplier ideal sheaves and the limiting behavior of the Ricci flow*, arXiv:0708.1590, Trans. Ams. 361(2009), 5839–5850.

▶[Ru2] Rubinstein, Yanir A. *Some discretizations of geometric evolution equations and the Ricci iteration on the space of Kähler metrics.* Adv. Math. 218 (2008), no. 5, 1526–1565.

[Ru3] Rubinstein, Yanir A. *On energy functionals, Kähler-Einstein metrics, and the Moser-Trudinger-Onofri neighborhood.* J. Funct. Anal. 255 (2008), no. 9, 2641–2660.

[Sal] Saloff-Coste, Laurent. *Aspects of Sobolev-type inequalities.* London Mathematical Society Lecture Note Series, 289. Cambridge University Press, Cambridge, 2002.

[Sal2] Saloff-Coste, L. *A note on Poincaré, Sobolev, and Harnack inequalities.* Internat. Math. Res. Notices (1992), no. 2, 27–38.

[Sc] Schoen, Richard. *Conformal deformation of a Riemannian metric to constant scalar curvature.* J. Differential Geom. 20 (1984), no. 2, 479–495.

[Se1] Sesum, Natasa. *Limiting behaviour of the Ricci flow*, arXiv:math/0402194

[Se2] Sesum, Natasa. *Compactness results for the Kähler-Ricci flow,* arXiv:0707.2974

[Shi] Shi, Wan-Xiong. *Deforming the metric on complete Riemannian manifolds.* J. Differential Geom. 30 (1989), no. 1, 223–30.

[SZ] Souplet, Philippe; Zhang, Qi S. *Sharp gradient estimate and Yau's Liouville theorem for the heat equation on noncompact manifolds.* Bull. London Math. Soc. 38 (2006), no. 6, 1045–1053.

[SY] Schoen, R.; Yau, S.-T. *Lectures on differential geometry.* Lecture notes prepared by Wei Yue Ding, Kung Ching Chang [Gong Qing Zhang], Jia Qing Zhong and Yi Chao Xu. Translated from the Chinese by Ding and S. Y. Cheng. Preface translated from the Chinese by Kaising Tso. Conference Proceedings and Lecture Notes in Geometry and Topology, I. International Press, Cambridge, MA, 1994.

[Tal] Talenti, Giorgio. *Best constant in Sobolev inequality.* Ann. Mat. Pura Appl. (4) 110 (1976), 353–372.

[Tao] Tao, Terence. *Poincaré conjecture, course Blog at UCLA*, terrytao.wordpress.com/2008

[Tao2] Tao, Terence. *Perelman's proof of the Poincaré conjecture: a nonlinear PDE perspective*, arXiv:math/0610903

[Topo] Toponogov, V. A. *The metric structure of Riemannian spaces of nonnegative curvature containing straight lines.* (Russian) Sibirsk. Mat. Z. 5 (1964), 1358–1369.

[To] Topping, Peter. *Lectures on the Ricci flow.* London Mathematical Society Lecture Note Series, 325. Cambridge University Press, Cambridge, 2006.

[Tr1] Trudinger, Neil S. *On imbeddings into Orlicz spaces and some applications.* J. Math. Mech. 17 (1967), 473–483.

[Tr2] Trudinger, Neil S. *Remarks concerning the conformal deformation of Riemannian structures on compact manifolds.* Ann. Scuola Norm. Sup. Pisa (3) 22 (1968), 265–274.

[Ya] Yamabe, Hidehiko. *On a deformation of Riemannian structures on compact manifolds.* Osaka Math. J. 12 (1960), 21–37.

[Y] Ye, Rugang. *The Logarithmic Sobolev inequality along the Ricci flow*, arXiv: 0707.2424v4, 2007.

[Yu] Yudovich, V. I. *Some estimates connected with integral operators and with solutions of elliptic equations.* (Russian) Dokl. Akad. Nauk SSSR 138 (1961), 805–808.

[Z1] Zhang, Qi S. *Some gradient estimates for the heat equation on domains and for an equation by Perelman.* IMRN, vol. 2006, Article ID 92314, 1–39, 2006.

[Z2] Zhang, Qi S. *A uniform Sobolev inequality under Ricci flow.* International Math. Research Notices, Vol. 2007, article id: rnm056, 1–17. ibid. erratum 2007.

[Z3] Zhang, Qi S. *Addendum to: A uniform Sobolev inequality under Ricci flow.* International Math. Research Notices, 2007.

▶[Z4] Zhang, Qi S. *Strong noncollapsing and uniform Sobolev inequalities for Ricci flow with surgeries.* Pacific J. Math. 239 (2009), no. 1, 179–200. Announcement, CRAS, 2008.

▶[Z5] Zhang, Qi S. *Heat kernel bounds, ancient $\kappa$ solutions and the Poincaré conjecture.* J. Func. Analysis 258 (2010), 1225–1246. arXiv:0812.2460

[Zi] Ziemer, William P. *Weakly differentiable functions. Sobolev spaces and functions of bounded variation.* Graduate Texts in Mathematics, 120. Springer-Verlag, New York, 1989.