GROMOV'S BETTI NUMBER BOUND

RAMAN ALIAKSEYEU

Abstract. This paper presents a beautiful result of Gromov, exhibiting a bound on the sum of Betti numbers of a complete Riemannian manifold M of non-negative sectional curvature which depends only on the dimension of M.

Contents

1. Introduction

Many important results of Riemannian geometry have come about out of attempts to understand how different restrictions on the curvature tensor of a Riemannian manifold affect its topology. One more recent such result, first published in [7], will be the topic of this paper.

Theorem 1.1. *[Gromov, 1981] Let (M, g) be a Riemannian n-manifold with sectional curvature $K(\sigma, p) \geq 0$ for any choice of $p \in M$ and plane $\sigma \subset T_qM$. Denote by b_i the ith Betti number of M. Then, there exists a constant $C(n)$, dependent only on the dimension of M, such that*

$$\sum_{i=0}^{n} b_i \leq C(n)$$

Petersen, in [9], has described this result as "one of the deepest and most beautiful results in Riemannian geometry". Previously, Bochner's technique was used to bound b_i of closed Riemannian manifolds with non-negative curvature by $\binom{n}{i}$. More famously, the bound $b_1 \leq n$ requires only that M is compact and has non-negative Ricci curvature (see [9], chapter 9). The result Theorem 1.1 above, however, does not require any compactness assumptions, and has a weaker curvature assumption.

Bounds on sums of Betti numbers are also significant in light of the famous Hopf conjecture, which asks if $S^2 \times S^2$ admits a metric of positive sectional curvature.

Date: August 2024.

Theorem 1.1 shows that the connect sum

$$(S^2 \times S^2)\# \dots \#(S^2 \times S^2)$$

with sufficiently many terms does not admit a metric of non-negative sectional curvature. Indeed, since Sha-Yang ([10]) exhibited a metric of positive Ricci curvature on this space, it follows that the sectional curvature hypothesis in Gromov's result cannot be weakened to Ricci curvature.

The proposed bound $C(n)$ above is on the order of $(n2^{2^n a})^{2^n b}$, for some constants a, b. It is likely very far from optimal. Gromov, in [7], conjectured that the optimal $C(n)$ is 2^n, achieved by the n-torus $S^1 \times \dots \times S^1$. It is known that the 2^n bound holds for manifolds which are rationally elliptic, and there is an unresolved conjecture of Bott that states that all simply-connected non-negatively curved manifolds are rationally elliptic ([1]). There appears to be little recent progress on this conjecture.

The proof of Theorem 1.1 can be minimalistically summarized like this. First, by developing a critical point theory for distance functions on complete Riemannian manifolds, we reduce the problem of estimating the sum of Betti numbers of M to estimating the sum of Betti numbers of a ball B in M of sufficiently large radius. We can bound this topological 'content' of the ball by a factor of the content of smaller balls by using the Mayer-Vietoris double-complexes of a system of covers of B. The maximum number of times this 'shrink'-bound can be iteratively applied to B is a topological invariant we will call the *rank* of B. We will use our critical point theory again to conclude that the rank of B is bounded by an expression depending only on the dimension of M.

My goal with this paper is to provide a clear and detailed proof of Gromov's result, and to present some insights and intuitions I came across while trying to understand it. I also hope that my writing will be useful to others who try to study similar expositions of this result, such as [7], [9], and [5]. These are my primary sources, along with [3] for section 4. Some preliminary results and techniques are collected in sections 2, 3, and 4, while the body of the proof is contained in section 5.

2. Distance and Volume Comparison

We will need two lemmas that, as the title of this section suggests, allow us to obtain bounds on geometric quantities of objects in M using the geometry of flat manifolds.

First, recall how the Riemannian metric g on M gives rise to a notion of an angle between vectors.

Definition 2.1. Let (M, g) be a Riemannian manifold, and let $p \in M$. If $v, w \in T_pM$ are non-zero vectors tangent to M at p, then the **angle** between v and w is given by

$$\angle(v, w) := \frac{g(v, w)}{|v||w|}$$

It makes sense to draw a further parallel with Euclidean geometry by importing some of its language:

Definition 2.2. Let $p, q \in M$. A **segment** $\overline{pq} : [0, l] \to M$ is a *minimizing* unit-speed geodesic such that $\gamma(0) = p, \gamma(1) = q$. The **length** l of $\overline{pq}$ is denoted $l(\overline{pq})$.

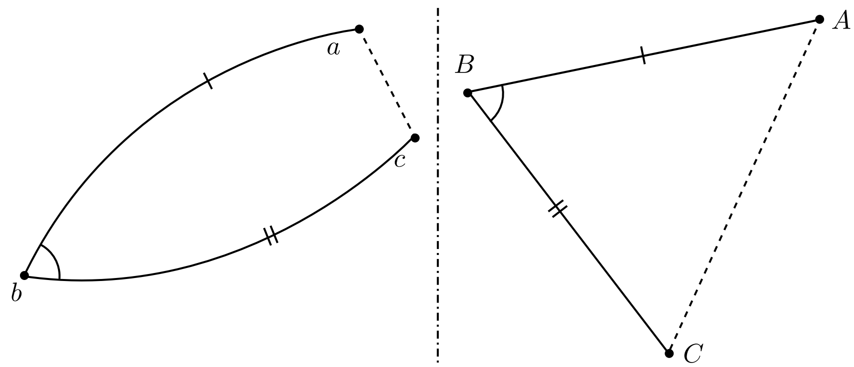

FIGURE 1. A rough diagram demonstrating the principle of Lemma 2.4. Diagram similar to one in [9].

If M is complete, by the Hopf-Rinow theorem, a segment $\overline{pq}$ exists for any $p, q \in M$ but is not necessarily unique. When referring to *the* segment $\overline{pq}$ in the future, we will mean some specific chosen segment which should be clear from context.

Definition 2.3. Let $a, b, c \in M$. Then the **angle** between segments $\overline{ab}$ and $\overline{bc}$ is the angle between their velocity vectors at b:

$$\angle abc := \angle(\overline{ba}'(0), \overline{bc}'(0))$$

We are ready to state the distance comparison lemma.

Lemma 2.4. *[Toponogov, 1959] Suppose (M, g) is a complete, non-negatively curved Riemannian n-manifold. Let a, b, c be points in M. Suppose A, B, C are points in $\mathbb{R}^n$ such that*

$$l(\overline{ab}) = l(\overline{AB}) \quad l(\overline{bc}) = l(\overline{BC})$$

$$\angle abc = \angle ABC$$

Then

$$l(\overline{ac}) \leq l(\overline{AC})$$

Proofs can be found in [4], chapter 2, or [9], chapter 12. For some graphical intuition, imagine 'peeling' a wedge made of $\overline{ab}, \overline{bc}$ away from a positively curved surface. In order to flatten $\overline{ab}, \overline{bc}$, one would have to stretch $\overline{ac}$ (see Figure 1). As a result of this lemma, we obtain the following generalization of the Law of Cosines to non-negatively curved Riemannian manifolds, which we will make ample use of.

Corollary 2.5. *For $a, b, c \in M$ as above, we have*

$$l(\overline{ac})^2 \leq l(\overline{ab})^2 + l(\overline{bc})^2 - 2l(\overline{ab})l(\overline{bc})\cos(\angle abc)$$

The second result involves volume comparison.

Lemma 2.6 (Bishop-Gromov)**.** *Let M be a complete n-manifold with positive sectional curvature. Denote by $V(p, r)$ the volume of the geodesic ball of radius r centered at p in M. If $v(r)$ is the volume of a Euclidean n-ball of radius r, then* *the ratio $V(p,r)/v(r)$ is non-increasing in r. In particular, the following inequality holds for $0 < r < R$:*

$$\frac{V(p,R)}{V(p,r)} \leq \frac{v(R)}{v(r)}$$

A proof of this fact can also be found in [9], chapter 7. We will use this fact only to give a bound on the greatest number $N(s,r)$ of disjoint balls of radius s that can be contained in $B_r(p) \subset M$, which is:

$$N(r,s) \leq \frac{\text{vol}B(p,r)}{\text{vol}B(p,s)} \leq (r/s)^n \tag{2.7}$$

Note that this also bounds the number of balls of radius $2s$ with centers in $B_r(p)$ needed to cover $B_r(p)$.

3. Grove-Shiohama Critical Point Theory

One of the things Morse theory tells us is that on a smooth manifold, only neighborhoods of critical points 'add topology' to the manifold. More precisely, if the region $M_a^b := f^{-1}([a,b])$ (for f a Morse function, $a < b$) is compact and contains no critical points, we have a (smooth) deformation retraction from the sublevel set $M^b := f^{-1}((-\infty, b))$ to M^a (see [8]).

One can develop a Morse-like theory for distance functions $f = \text{dist}_p$ to $p \in M$ where 'criticality' of a point x depends not on the gradient/Hessian of f, but on the spread, around x, of minimizing geodesics leading back to p. This theory still gives us a retraction lemma for compact regions free of critical points, which we will use in conjunction with results from section 2 to show that any complete Riemannian manifold of non-negative curvature has the homotopy type of a compact manifold with boundary. I believe the first use of this theory was by Berger in [2] while proving a variant of the Sphere theorem, but Grove and Shiohama were the ones who proved the retraction lemma in [6].

Let's start with a precise definition of a critical point.

Definition 3.1. Let M be a complete Riemannian manifold. Let $f = \text{dist}_p$ be a distance function from x to p, i.e. the length of a segment $\overline{px}$. Let $m(x)$ be the set of 'minimizing directions' in T_xM, i.e. the set of all $w = \overline{xp}'(0)$ ranging over all choices of segment $\overline{xp}$. A point $x \in M$ is then α-**critical** if for any $v \in T_xM$ there is a $w \in m(p)$ such that $\angle(v,w) \leq \alpha$. Accordingly, a point is α-**regular** if it is not α-critical, i.e. if there is a unit $v \in T_xM$ such that $m(x)$ is contained in an α-ball on $S^{n-1} \subset T_xM$ centered at v.

Practically, only $\pi/2$-critical points matter, but α-criticality will be useful for proving parts of the retraction lemma. Critical points, without α specified, should be assumed to be $\pi/2$-critical. See Figure 2 for an example of a critical and regular point.

The statement we want to prove is:

Theorem 3.2. *[Grove-Shiohama] Let $f = \text{dist}_p$. If the region $f^{-1}([a,b])$ for $0 < a < b$ is compact and free of critical points, then there is a deformation retraction from $B_b(p)$ to $B_a(p)$.*

We prove this similarly to how we prove the corresponding lemma in Morse theory, by constructing a non-zero vector field in the neighborhood of $f^{-1}([a,b])$ such

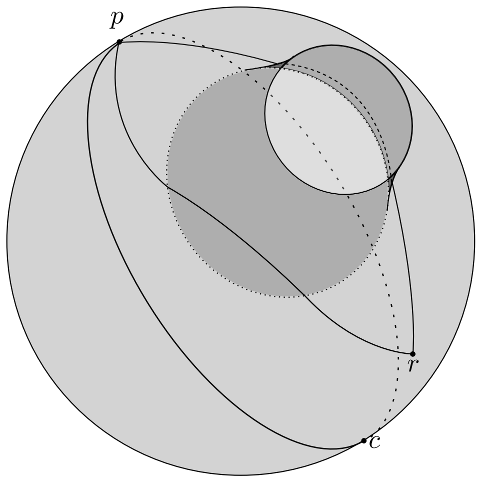

FIGURE 2. The point c is $\pi/2$-critical with respect to p, point r is $\pi/2$ regular. Minimizing geodesics to p shown.

that f is strictly decreasing (negative derivative bounded above) along its integral lines. That way, by going along the flow of this vector field, we will reach $B_a(p)$ in finite time. In order to do this, we will need some properties of α-regular/critical points.

Proposition 3.3. *The α-regular points form an open subset of M.*

Proof. We will show that the set of α-critical points in M is closed. Suppose $\{x_i\}$ is a sequence of α-critical points that converges to $x \in M$ (without loss of generalization, $\{x_i\}$ is contained in a geodesically convex neighborhood of x). Take any vector $v \in T_xM$ and identify it with its parallel transports v_i in $T_{x_i}M$. Because every point in $\{x_i\}$ is critical, we can select a w_i in $T_{x_i}M$ such that w_i is the initial velocity vector of a length-minimizing geodesic from x_i to p, and $\angle(v_i, w_i) \geq \alpha$. Pick a subsequence $\{x_j\}$ of $\{x_i\}$ such that corresponding w_j as identified with their transports in T_xM converge to some w. Then $\angle(v, w) \geq \alpha$. We now show that w points in the direction of p. Let $d_j = \text{dist}(p, x_j)$ and let d be their limit. Because $\exp(x, v, t) = \exp_p(tv)$ is continuous, the limit $\exp_x(dw)$ of (some subsequence of) $\exp_{x_j}(d_j w_j)$ is equal to p. Hence, $\exp_x(tw) : [0, d] \to M$ is a geodesic. Because it is a limit of minimizing geodesics, it is forced to be minimizing also. We have just shown that x must be a critical point. □

For any point $q \neq p$ in M, denote by $c_\alpha(q)$ the set of all unit vectors $v \in T_qM$ that serve as 'centers of α-hemispheres' which contain $m_p(q)$. More precisely, $v \in c_\alpha(q)$ is such that for any $w \in m_p(q)$ we have $\angle(v, w) \leq \alpha$. This set is convex, since $c_\alpha(q)$ at any $q \in M$ is an intersection of the convex balls on $S^n \subset T_qM$ of radius α centered at vectors in $m_p(q)$.

Proposition 3.4. *Given the same assumptions as in Theorem 3.2, there is a vector field X supported on a compact neighborhood of $f^{-1}([a, b])$ such that if $c(t)$ is an integral curve of X contained in $f^{-1}([a, b])$, then*

$$\frac{d(f \circ c)(t)}{dt} \leq \delta < 0$$

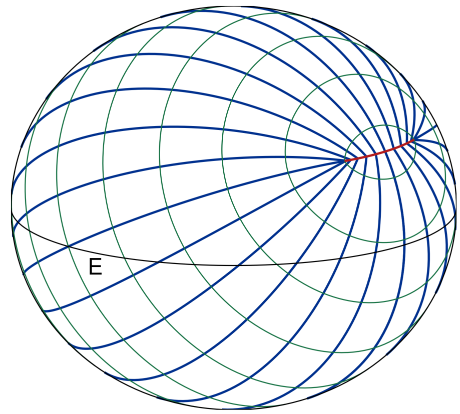

Figure 3. Cut locus on an ellipsoid, shown in red. Diagram by Cffk from https://en.wikipedia.org/wiki/Cut_locus

Proof. Note that if a point $q \in M$ is $\pi/2$ regular, since $S^n \subset T_qM$ is compact, we have that q is also α_q-regular, for some $\alpha_q < \pi/2$. We show that in fact $f^{-1}([a,b])$ is composed not of just $\pi/2$-regular, but indeed α-regular points. We will then use this to construct a vector field X with flow ψ compactly supported on a neighborhood of $f^{-1}([a,b])$ such that for $0 \leq \tau_1 < \tau_2$:

$$r \circ c(\tau_2) - r \circ c(\tau_1) < (\tau_1 - \tau_2)\cos\alpha$$

Since c_α is convex, $V = \sum \phi_i V_{q_i}$ as defined above has the property that $V(x) \subset c_\alpha(x)$ for all $x \in f^{-1}([a,b])$ so it is in particular non-zero. Denote by X the vector field $V/|V|$.

Note that f is smooth precisely on the complement of the *cut locus* of M with respect to p (the set of points in M past which a minimizing geodesic cannot be continued while still remaining minimizing). This is a closed nowhere dense set which contains, but in general is not entirely composed of, critical points of r. (see diagram Figure 3). On the points where f is smooth, we can compute the derivative of $f \circ c$ to be negative, which gives the result. On those points where r is not differentiable, we can use the first variation formula to express the derivative of $f \circ c$ in terms of the energy functional.

Let $x = c(t)$ and consider the segment $\gamma = \overline{px}$, so $\gamma(0) = p$ and $\gamma(1) = x$. Consider a variation $v(s,t)$ of γ, such that its variational field $V = \frac{dv}{ds}$ at $t = 0$ is 0, and the curve $v(s,1)$ starting at x follows an integral line of X (see diagram Figure 4). Note that $\left|\frac{dv}{dt}\right|$ must be constant and hence is equal to the length of the curve $v(s,\bullet)$.

Note that

$$\frac{1}{2}|r \circ c(s)|^2 = E(c(s))$$

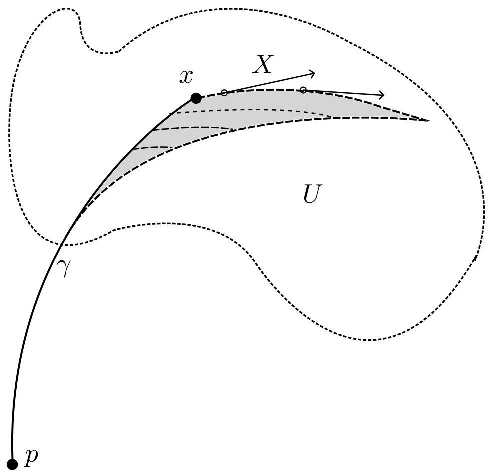

FIGURE 4. A possible variation v we can use in the proof of Proposition 3.4. We would like v to be zero outside of the neighborhood U where X is defined. The image of v is shaded.

Where $E(c(s))$ denotes the energy of the curve $c : [0,1] \to M$. Taking the derivatives of both sides and applying the first variation formula, we get:

$$\begin{aligned} r(x)\,\frac{d(f\circ c)}{ds}\bigg|_{s=0} &= \frac{dE}{ds}(0) \\ &= g\left(\frac{dv}{dt}(0), \frac{dv}{ds}(0)\right) \\ &= \left|\frac{dv}{dt}(0,1)\right| \angle\left(X, \frac{dv}{dt}\right) \\ &= -f(x)\angle\left(X, -\frac{dv}{dt}\right) \end{aligned}$$

Note that $-\frac{dv}{dt} \in m_p(x)$, and hence the angle between it and $V(x)$ is less than α. This proves the proposition. □

Proof of Theorem 3.2. Let F^t be the flow generated by X as given in the proposition above. Because the derivative $\frac{f\circ c(t)}{dt} \leq \delta < 0$, we have that for every point $x \in f^{-1}([a,b])$ there is a time t_x such that $F^{t_x}(x) \in f^{-1}((-\infty, a))$. Since the assignment $x \mapsto t_x$ is continuous, we get the desired retraction. □

The primary advantage of our definition of $\pi/2$-critical points (with respect to some point p) is that sequences $\{x_i\}$ of critical points cannot get too far away from p without $\{\overline{px_i}'(0)\}$ getting further away from each other in terms of angle. This idea is made precise with the following lemma.

Lemma 3.5. *[Gromov] Let M be a complete manifold with non-negative sectional curvature. Suppose x_1 is a critical point, and suppose x_2 is such that*

$$l(\overline{px_2}) \geq \nu l(\overline{px_1})$$

with $\nu > 1$. Then

$$\theta := \angle x_1 p x_2 \geq \arccos(1/\nu)$$

Proof. Because q_1 is a critical point, for any choice of segment $\overline{q_1q_2}$ it is possible to find a segment $\overline{q_1p}$ such that $\angle(\overline{q_1q_2}'(0), \overline{q_1p}'(0)) \leq \pi/2$. Apply Lemma 2.4 to q_2, q_1, p (in this order) to get

$$l(\overline{q_2p})^2 \leq l(\overline{q_1q_2})^2 + l(\overline{q_1p})^2$$

Also apply Lemma 2.4 to q_1, p, q_2 to get

$$l(\overline{q_1q_2})^2 \leq l(\overline{pq_1})^2 + l(\overline{pq_2})^2 - 2l(\overline{pq_1})l(\overline{pq_2})\cos(\theta)$$

Plugging in the first inequality into the second, and doing some algebra, gives us the result. □

Corollary 3.6. *If $x_0, \ldots, x_n$ are critical with respect to p, and*

$$l(\overline{px_i}) \geq \nu l(\overline{px_{i-1}})$$

for $\nu > 1$ and $i = 1, 2, ..., n$, then $n \leq \frac{2\pi}{\arccos(1/\nu)}^{n-1}$.

Proof. Lemma 3.5 says that $\angle(\overline{px_i}'(0), \overline{px_j}'(0)) \geq \arccos(1/\nu)$ for all $i \neq j$. The greatest number of unit vectors in $S^{n-1} \subset T_pM$ that are at least α apart in angle is the greatest number of balls on S^{n-1} of radius α that are pairwise disjoint. Thus the bound is given by (2.7). □

Notably, this is the only time the curvature condition is used while proving Theorem 1.1.

Corollary 3.7. *If M is a complete manifold with non-negative sectional curvature, then for any $p \in M$ all $\pi/2$-critical points must be contained in some sufficiently large ball $B_R(p)$ around p. Thus M has the homotopy type of a compact manifold with boundary.*

4. Mayer-Vietoris Double Comlpex

We will consider Betti numbers b_i of M in terms of (deRham) cohomology with $\mathbb{R}$ coefficients.

Definition 4.1. For $A \subset M$, denote by $H^*(A)$ the sum of all the cohomology vector spaces:

$$H^*(M) = \bigoplus_{i=0}^{n} H^i(A, \mathbb{R})$$

Denote the Betti numbers of A by $b_i(A)$.

If $A \subset B$, denote by $b_i(A \subset B)$ the rank (as a linear map of vector spaces) of the map $i^* : H^i(B) \to H^i(A)$ induced by the inclusion $i : A \to B$. Similarly,

$$b_*(A \subset B) = \sum_{i=0}^{n} b_i(A \subset B)$$

Particularly important is the following 'in-between' lemma.

Lemma 4.2. *Let $A \subset B \subset C \subset D$, then*

$$b_i(A \subset D) \leq b_i(B \subset C)$$

This is only a consequence of linear algebra. Since for any bounded, open A, B with $\overline{A} \subset B$ we can find some compact manifold with boundary Y with $A \subset Y \subset B$, we have

$$b_i(A \subset B) \leq b_i(Y) < \infty$$

We have proven

Corollary 4.3. *If A, B are bounded, open, and $\overline{A} \subset B$, then $b_i(A \subset B)$ is finite.*

Recall the Mayer-Vietoris sequence for cohomology: if M is covered by open U_1, U_2, we have

$$\cdots \to H^{i-1}(U_1 \cap U_2) \to H^i(U_1 \cup U_2) \to H^i(U_1) \bigoplus H^i(U_2) \to \dots$$

Then, we have

$$b_i(M) \leq b_i(U_1) + b_i(U_2) + b_{i-1}(U_1 \cap U_2)$$

Notation 4.4. Denote by $U_{i_1,\dots,i_n}$ the intersection $U_{i_1} \cap \cdots \cap U_{i_n}$.

Applying induction on the number N of the open sets in the cover, we get the following inequality.

Lemma 4.5. *If $U_1, \dots, U_N$ form an open cover of some subset $A \subset M$, we have*

$$b_i(A) \leq \sum_{j=0}^{N} \sum_{U_{i_1,\dots,i_{N-j}} \neq \emptyset} b_j(U_{i_1,\dots,i_{N-j}})$$

We will need the following variation of the above lemma.

Lemma 4.6. *Let U_i^j be sets such that $\overline{U_i^j} \subset U_i^{j+1}$. For $j = 0, 1, \dots, n+1$ let $X^j := \cup_i U_i^j$. Then*

$$b_k(X^0 \subset X^{n+1}) \leq \sum_{j=0}^{k} \sum_{U_{i_1,\dots,i_{k-j}} \neq \emptyset} b_j(U^j_{i_1,\dots,i_{k-j}}, U^{j+1}_{i_1,\dots,i_{k-j}})$$

To prove this, we need a topological tool known as the Mayer-Vietoris double complex. For details in the material to follow, see [3], chapter 2.

Let $U_1, \dots, U_N$ be a cover of M, with $N > 2$. Consider the following chain of inclusions

$$X^j \leftarrow \prod_{0 \leq i_0 \leq n} U^j_{i_0} \leftarrow \prod_{0 \leq i_0 < i_1 \leq n} U^j_{i_0,i_1} \leftarrow \dots$$

where the arrow $\prod U^j_{i_0,\dots,i_m} \to \prod U^j_{i_0,\dots,i_{m-1}}$ represents $m+1$ maps $\delta_k : \prod U_{i_0,\dots,i_m} \to \prod U_{i_0,\dots,i_{m-1}}$ $(k = 0, \dots, m)$ which on $U_{i_0,\dots,i_m}$ 'ignore the kth set in the intersection', so $\delta_j : U_{i_0,\dots,i_m} \to U_{i_0,\dots,\hat{i_k},\dots,i_m}$ is the natural inclusion. The first arrow is composed of the natural inclusions $U_{i_0} \to M$. This sequence produces the following induced sequence of maps on products of spaces of forms:

$$\Omega^*(M) \to \bigoplus \Omega^*(U_{i_0}) \to \bigoplus \Omega^*(U_{i_0,i_1}) \to \dots$$

We combine the arrows δ_k^* into one boundary operator δ for $\omega \in \bigoplus \Omega^*(U_{i_0,\dots,i_m})$ as such:

$$(\delta\omega)_{i_0,\dots,i_{m+1}} = \sum (-1)^k \delta_k \omega_{i_0,\dots,i_{m+1}}$$

It can be checked that indeed $\delta^2 = 0$, so

$$0 \to \Omega^*(M) \xrightarrow{\delta} \bigoplus \Omega^*(U_{i_0}) \xrightarrow{\delta} \bigoplus \Omega^*(U_{i_0,i_1}) \to \dots$$

is a chain complex. An additional calculation tells us that it is exact.

Expand this complex into the following lattice:

$$
\begin{array}{ccccccccc}
\vdots & & \vdots & & \vdots & & \vdots & & \iddots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
\Omega^3(X^j) & \longrightarrow & \bigoplus \Omega^3(U^j_{i_0}) & \longrightarrow & \bigoplus \Omega^3(U^j_{i_0,i_1}) & \longrightarrow & \bigoplus \Omega^3(U^j_{i_0,i_1,i_2}) & \longrightarrow & \dots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
\Omega^2(X^j) & \longrightarrow & \bigoplus \Omega^2(U^j_{i_0}) & \longrightarrow & \bigoplus \Omega^2(U^j_{i_0,i_1}) & \longrightarrow & \bigoplus \Omega^2(U^j_{i_0,i_1,i_2}) & \longrightarrow & \dots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
\Omega^1(X^j) & \longrightarrow & \bigoplus \Omega^1(U^j_{i_0}) & \longrightarrow & \bigoplus \Omega^1(U^j_{i_0,i_1}) & \longrightarrow & \bigoplus \Omega^1(U^j_{i_0,i_1,i_2}) & \longrightarrow & \dots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
\Omega^0(X^j) & \longrightarrow & \bigoplus \Omega^0(U^j_{i_0}) & \longrightarrow & \bigoplus \Omega^0(U^j_{i_0,i_1}) & \longrightarrow & \bigoplus \Omega^0(U^j_{i_0,i_1,i_2}) & \longrightarrow & \dots
\end{array}
$$

$\uparrow d$ $\qquad$ $\xrightarrow[\delta]{}$

Here the (non-zero) horizontal arrows are given by the δ operator and the (non-zero) vertical arrows are given by the d operator. Both are differential, and commute with each other, giving this complex a 'double'-differential structure.

This lattice admits another operator which makes it into a differential complex. Denote $\Omega^l(X^j) = C^j_{0,l}$ and $\bigoplus \Omega^l(U^j_{i_1,\dots,i_k}) = C^j_{k,l}$. Then consider the 'sum-diagonals' $S_m = \bigoplus_{p=0}^m C^j_{p,m-p}$. We define the *total differential* operator $D : S_m \to S_{m+1}$ to map the pth component of $(c_1, \dots, c_m) \in S_m$ by

$$D(c_p) = \delta + (-1)^p d$$

This makes $0 \to S_1 \to S_2 \to \dots$ into a differential complex. It is then possible to show that the map $H^*(X^j) \to \bigoplus H^*(U^j_{i_0})$ induced by the inclusions of the cover elements into X^j is an isomorphism of the cohomology $H^*(X^j)$ into the cohomology $H_D(S_m, D)$ of the total differential.

We can now prove the topological fact we need.

Proof of Lemma 4.6: We will follow Cheeger's proof from [5], section 1.5. Let Z_{n+1} be a vector space of representative cocycles in $C^{n+1}_{0,n}$, isomorphic to $H^n(X^{n+1})$. We will find a filtration $Z_n = Z^n_{n+1} \supset Z^n_n \supset \dots \supset Z^n_0$ such that

$$\dim(Z^n_{j+1}/Z^n_j) \leq \sum_{i_0,\dots,i_{n-j}} b_j(U^{j+1}_{i_0,\dots,i_{n-j}} \subset U^j_{i_0,\dots,i_{n-j}})$$

And so that if $r_s^* : H^k(X^{s+1}) \to H^k(X^s)$ are the maps induced on the cohomology by the restrictions $r_s : X^s \to X^{s+1}$, and $z \in Z^j_0$, then $r_0^* \dots r_n^*(z)$ is d-exact. This will give us the result, since

$$\begin{aligned} b_k(X^0 \subset X^{n+1}) &= \dim(\operatorname{im}(H^k(X^{n+1}) \xrightarrow{i_*} H^k(X^0))) \\ &= \dim(Z^n_n/\ker(H^k(X^{n+1}) \xrightarrow{r_0^*\cdots r_n^*} H^k(X^0))) \\ &\leq \dim(Z^n_{n+1}/Z^n_0) \\ &= \sum_{j=0}^{n} \dim(Z^n_{j+1}/Z^n_j) \end{aligned}$$

We start the construction with

$$Z^n_n := \{z \in Z^n_{n+1} \mid r_n^*(z) \in C^n_{0,n} \text{ is } d\text{-exact}\}$$

Where d is the map induced on cohomology. Then,

$$\begin{aligned} \dim(Z^n_{n+1}/Z^n_n) &= \dim(r_n^*(Z^n_{n+1}) \\ &= \dim(b_n(X^{n+1} \subset X^n)) \\ &= \sum_i b_n(U_i^{n+1}, U_i^n) \end{aligned}$$

Choose a linear map $d^{-1} : r^*(Z^n_{n+1}) \to C^n_{0,n-1}$ such that $dd^{-1}(z) = z$. Note that because d and δ anti-commute, and $\delta z = 0$ by the fact that Z^n is made up of D-cocycles, we have that $d\delta d^{-1}(z) = 0$. Also, $\delta(\delta d^{-1})(z) = 0$. We have shown that $\delta d^{-1} z$ sends D-cocycles (that are d-exact) to D-cocycles.

Define

$$Z^n_{n-1} := \{z \in Z^n_n \mid r_{n-1}^* \delta d^{-1} r_n^*(z) \text{ is } d\text{-exact}\}$$

Proceeding in a similar fashion, define the rest of the filtration. Then,

$$\begin{aligned} \dim(Z^n_{j+1}/Z^n_j) &\leq \dim\left(\operatorname{im}\left(\bigoplus_{i_0,\ldots,i_{n-j}} H^j(U^{j+1}_{i_0,\ldots,i_{n-j}}) \xrightarrow{r_j^*} \bigoplus_{i_0,\ldots,i_{n-j}} H^j(U^j_{i_0,\ldots,i_{n-j}})\right)\right) \\ &= \sum_{i_0,\ldots,i_{n-j}} b_j(U^{j+1}_{i_0,\ldots,i_{n-j}}, U^j_{i_0,\ldots,i_{n-j}}) \end{aligned}$$

Now we have to show that $r_0^* \ldots r_n^*(z)$ is d-exact. Note that it is D-exact, as

$$r_0^* \ldots r_n^*(z) = D(a_{n-1} + \cdots + a_0)$$

where

$$a_i = (-1)^{(n-1)-i} r *_0 \ldots r_{i-1}^*(d^{-1} r^* i)(\delta d^{-1} r_{i+1}^* \ldots d^{-1} r_n^*(z)$$

Then, using the fact that the double complex above is δ-exact, choose $b_0 \in C^0_{n-1,0}$ with $\delta b_0 = a_0$, and set $a_1' = a_1 + db_0$. Then

$$r_0^* \ldots r_n^*(z) = D(a_{n-1} + \cdots + a_1' + (d+\delta)(b_0)) = D(a_{n-1} + \cdots + a_1')$$

Proceeding similarly to this, we get

$$r_0^* \ldots r_n^*(z) = D(a'_{n-1}) \quad a'_{n-1} \in C^0_{0,n-1}$$

By construction, $\delta a'_{n-1} = 0$, so we are done. □

5. Proof of Theorem 1.1

We will reason about the dimension of $H^*(M)$ in terms of the content of a sufficiently large ball $B_r(p) \subset M$.

Definition 5.1. The **content** of the ball $B_r(p)$ in M is given by

$$\mathrm{cont}(p, r) = b_*(B_r(p) \subset B_{5r}(p))$$

Many constants in this section, including the 5 in the definition above, are chosen somewhat arbitrarily. Nevertheless, their choices make the necessary triangle inequalities work out.

There is a way to estimate the content of large balls $B_r(p)$ using the content of smaller balls inside it.

Lemma 5.2. *If for every $p' \in B_r(p)$ and $j = 0, 1, \dots, n+1$ we have*

$$\mathrm{cont}(p', 10^{-j} r) \leq c$$

then

$$\mathrm{cont}(p, r) \leq (n+1) 2^{N(10^{-(n+1)}r, r)} c$$

where $N(s, r)$ is the maximal number of open balls of radius s required to cover a ball of radius r.

Proof. Let $U_i^j = B_{10^{j-(n+1)}r}(p_i)$ be a cover as in Lemma 4.6. Notice that there are at most $(n+1)2^{N(10^{-(n+1)r,r})}$ terms on the right hand side of the inequality given by the lemma (where $N(10^{-(n+1)r,r})$ is defined as at the end of section 2). Since for $1 \leq k \leq l$

$$U^{j+1}_{i_0,\dots,i_l} \subset B_{10^{j-(n+1)}r}(p_k) \subset B_{5\cdot 10^{j-(n+1)r}}(p_k) \subset U^{j}_{i_0,\dots,i_l}$$

Thus by Lemma 4.2,

$$b_j(U^j_{i_0,\dots,i_{k-j}}, U^{j+1}_{i_0,\dots,i_{k-j}}) \leq c$$

Since

$$B_r(p) \subset X^0 \subset X^{n+1} \subset B_{5r}(p)$$

we get the claimed content bound. □

Definition 5.3. The ball $B_r(p)$ **compresses** into $B_s(q)$ if $B_{5s}(q) \subset B_{5r}(p)$ and there exists a homotopy $f_t : B_r(p) \to B_{5r}(p)$ such that $\mathrm{im} f_0 = B_r(p)$ and $\mathrm{im} f_1 \subset B_s(q)$.

Using a variant of Lemma 4.2 we see that

Proposition 5.4. *If $B_r(p)$ compresses into $B_s(q)$ then*

$$\mathrm{cont}(p, r) \leq \mathrm{cont}(q, s)$$

See figure Figure 5 for an illustration.

Definition 5.5. A ball $B_r(p)$ is called **incompressable** if any ball $B_s(q)$ it compresses into has $s > r/2$.

We can now define an invariant which will, together with Lemma 5.2, give us a bound on the content of M.

Definition 5.6. The **rank** $\mathrm{rank}(r, p)$ of a ball $B_r(p)$ is defined inductively as follows:

(1) Contractible balls are assigned the rank 0.

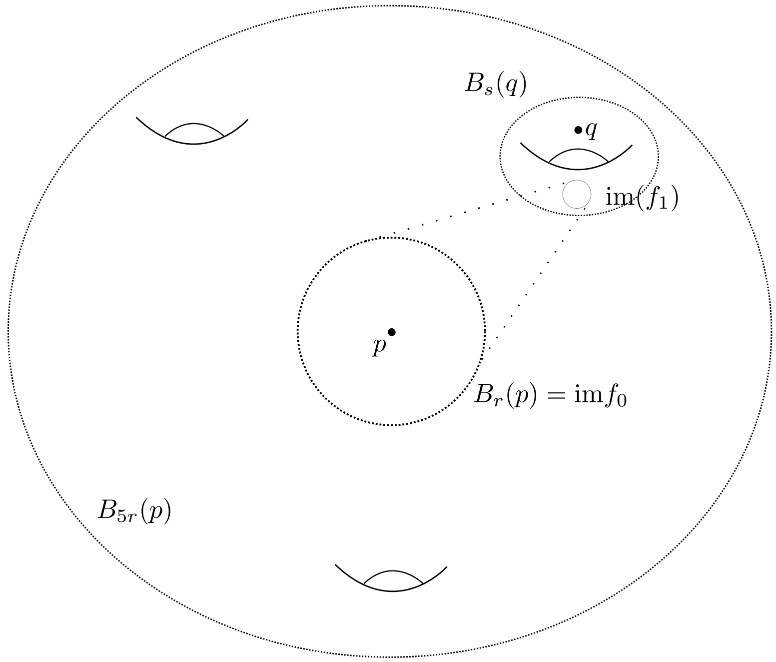

FIGURE 5. A schematic illustration of how $B_r(p)$ can compress into $B_s(q)$. $B_r(p)$ is contractible, $B_s(q)$ has content 2.

(2) Ball $B_r(p)$ that compresses into an incompressible $B_s(q)$, which is such that for all $q' \in B_s(q)$ the balls $B_{s/10}(q')$ are contractible, is given rank 1.
(3) In general, $B_r(p)$ that compresses into an incompressible $B_s(q)$, which is such that for all $q' \in B_s(q)$ the balls $B_{s/10}(q')$ have rank at most k, is given rank $k+1$.

The quantity $\mathrm{rank}(r,p)$ is bounded, since balls can only be shrunk by a factor of 10 so many times before they become smaller than the injectivity radius of M (recall that the injectivity radius of M is the infimum over all $p \in M$ of the radii R_p of the largest balls $B_{R_p}(0)$ which $\exp_p$ maps diffeomorphically into M).

Lemma 5.7. *For any ball $B_r(p)$ we have*

$$\mathrm{cont}(p,r) \leq ((n+1)2^{N(10^{-(n+1)}r,r)})^{\mathrm{rank}(p,r)}$$

where n is the same as in Lemma 5.2.

Proof. We proceed by induction. If $\mathrm{rank}(p,r) = 1$ then $B_r(p)$ is homotopy equivalent to a subset of a ball $B_s(q)$ for which all balls $B_{10^{-j}s}(q')$ are contractible, and hence have content 1. Thus by Lemma 5.2 this case follows. The induction step is similar. □

At this point all we need is a bound on rank that is independent of anything except the dimension of M. We do this by relating the rank of a ball to the number of critical points in its vicinity.

Lemma 5.8. *Consider a ball $B_r(p)$, and suppose*

$$5s + l(\overline{py}) \leq 5r \quad l(\overline{py}) \leq 2r$$

Then, if $B_r(p)$ does not compress into $B_s(y)$, there is a critical point x (with respect to y) in $B_{r+l(\overline{py})}(y) \setminus B_s(y)$.

Proof. If there is no critical point in $B_{r+l(\overline{py})}(y) \setminus B_s(y)$, then Theorem 3.2 tells us that there is a deformation retraction of $B_{r+l(\overline{py})}(y)$ into $B_s(y)$. However, by using this deformation retraction we can homotope $B_r(p) \subset B_{r+l(\overline{py})}(y)$ to $B_s(y)$. Since the first equation says $B_{5s}(y) \subset B_{5r}(p)$ and the second one says $B_{r+l(\overline{py})}(y) \subset B_{5r}(p)$, we have that $B_r(p)$ compresses into $B_s(y)$, which is a contradiction. □

Lemma 5.9. *Let $B_r(p)$ have rank k. Then there exist k critical points $x_k, \dots, x_0 \in B_{3r/2}(p)$, critical with respect to some point $y \in B_{3r/2}(p)$, such that*

$$l(\overline{yx_i}) \geq \frac{5}{4} l(\overline{yx_{i-1}})$$

Proof. Assume $B_r(p)$ is incompressable, by homotoping to a smaller ball if necessary. Let $r_k = r$, $p_k = p$. By definition of rank, there exists some $r_{k-1} \leq r_k/10$ and $p_{k-1} \in B_r(p)$ such that $B_{r_{k-1}}(p_{k-1})$ is incompressible and has rank $k-1$. Note that

$$B_{5r_{k-1}}(p_{k-1}) \subset B_{3r/2}(p_k)$$

Construct r_i, p_i for $i = k, \dots, 0$ inductively in this way. This implies, for $y = p_0$:

$$l(\overline{p_i y}) + \frac{5r_i}{2} \leq 5r_i \quad l(\overline{p_i y}) \leq \frac{3r_i}{2} < 2r_i$$

Because $B_{r_i}(p_i)$ are incompressible, they do not compress into $B_{r_i/2}(y)$. Then by Lemma 5.8 we have a critical point x_i such that

$$\frac{1}{2} r_i \leq l(\overline{yx_i}) \leq \frac{4}{r\,i}$$

From which follows

$$l(\overline{yx_i}) \geq \frac{r_i}{2} \geq 5r_{i-1} \geq \frac{5}{4} l(\overline{yx_{i-1}})$$

□

Corollary 5.10.

$$\operatorname{rank}(p, r) \leq \left(\frac{2\pi}{\arccos(4/5)} \right)^{n-1}$$

Proof. See 3.6. □

Proof of Theorem 1.1: Recall that by Corollary 3.7 we have that $\dim(H^*(M)) = \operatorname{cont}(p, R)$, for some sufficiently large R. Applying Lemma 5.7 and the corollary above gives us the result. □

6. Acknowledgments

I would like to thank Ben Lowe for being a phenomenal mentor this summer, and Peter May for organizing yet another wonderful UChicago Math REU. I would also like to thank both of them for thoroughly reviewing drafts of this paper and giving me many helpful suggestions.

References

[1] Igor Belegradek, Progress on Gromov's Conjecture of the bound of total Betti numbers, URL (version: 2024-04-30): https://mathoverflow.net/q/470204

[2] Berger, M. Les variétés riemanniennes (1/4)-pincées. Annali della Scuola Normale Superiore di Pisa - Scienze Fisiche e Matematiche, Serie 3, Volume 14 (1960) no. 2, pp. 161-170. http://www.numdam.org/item/ASNSP_1960_3_14_2_161_0/

[3] Bott, Raoul, and Loring W Tu. Differential Forms in Algebraic Topology. Springer Science & Business Media, 2013.

[4] Cheeger, Jeff, and David Ebin. Comparison Theorems in Riemannian Geometry. Newnes, 2009.

[5] Cheeger, Jeff, Mikhail Gromov, Christian Okonek, and Pierre Pansu. Geometric Topology: Recent Developments. Springer, 2006.

[6] K. Grove, K. Shiohama, A generalized sphere theorem. Ann. Math. (2) 106(2), 201–211 (1977)

[7] Gromov, Michael. "Curvature, Diameter and Betti Numbers." Commentarii Mathematici Helvetici 56, no. 1 (December 1, 1981): 179–95. https://doi.org/10.1007/bf02566208.

[8] Milnor, John W, R Wells, and M Spivak. Morse Theory. Princeton: Princeton University Press, 1973.

[9] Petersen, Peter. Riemannian Geometry. Third Edition. Springer, 2016.

[10] Ji-Ping Sha, Dagang Yang, Positive Ricci Curvature on Compact Simply Connected 4-Manifolds. From Proceedings of Symposia in Pure Mathematics Volume 54 (1993), Part 3, pg. 529.