A THEOREM OF KATOK

WILL J. MERRY

ABSTRACT. These are notes for two lectures given by the author at the 2009 Neuchatel Winter School on Closed Geodesics. The aim of the lectures was to present the proof of the following remarkable theorem of Katok [3, 4]: consider a closed surface Σ with Euler characteristic $\chi(\Sigma) < 0$, and let g denote a Riemannian metric on Σ, normalized so as to have volume one. Let $n_g(T)$ denote the number of closed geodesics of length at most T. Then the exponential growth of n_g is bounded below by the purely topological quantity $\sqrt{-2\pi\chi(\Sigma)}$, and equality is attained if and only if g has constant curvature. The proof of this theorem is the content of the second lecture. In the first lecture we sketch the proof that if (M^n, g_0) is a closed Riemannian manifold with constant nonpositive curvature $K \leq 0$ then the exponential growth of $n_g(T)$ is given by $(n-1)\sqrt{|K|}$. In an appendix we collect some elementary material on the geometry of the tangent bundle and the structure of the geodesic flow.

Lecture One: Introducing the counting function

The aim of this first lecture is to introduce the key players in our story; namely the counting function and the topological entropy. We conclude by calculating the exponential growth of the counting function for any closed Riemannian manifold of constant negative curvature. In the next lecture we build on this by sketching the proof of a remarkable theorem of Katok, which gives a lower bound on the exponential growth of the counting function for a surface *without* any assumptions on the curvature.

1. INTRODUCTION

Let M denote a closed Riemannian manifold. Given a Riemannian metric g on M, let $\Lambda(g)$ denote the set of all *closed geodesics* on M. Given a closed geodesic γ, let $\lambda(\gamma)$ denote the length of γ, that is, the smallest positive real number T such that $\gamma(0) = \gamma(T)$ and $\dot{\gamma}(0) = \dot{\gamma}(T)$. Given $T \in \mathbb{R}^+$ let $\Lambda(T; g)$ denote the set of all closed geodesics γ satisfying $\lambda(\gamma) \leq T$.

Definition 1.1. Define the *counting function* $n_g : \mathbb{R}^+ \to \mathbb{Z}$ by

$$n_g(T) := \#\Lambda(T; g).$$

Thus $n_g(T)$ is the number of closed geodesics on (M, g) with length $\leq T$.

Let us introduce some more standard notation. M will always denote an n dimensional oriented closed Riemannian manifold. Σ will always denote an oriented closed surface. g will always denote a Riemannian metric, and g_0 will always denote a Riemannian metric of negative curvature.

For a Riemannian manifold (M, g), let $\pi^g : S^gM \to M$ denote the unit sphere bundle and $\phi_t^g : S^gM \to S^gM$ the geodesic flow. Let μ_g denote the *Liouville measure* on S^gM; by definition this is the unique smooth probability measure on S^gM that is ϕ_t^g-invariant. Explicitly, μ_g is the measure obtained from the volume form $\alpha_g \wedge d\alpha_g^{n-1}$, where $\alpha_g \in \Omega^1(TM)$ is the form obtained by pulling back the canonical

Liouville form on T^*M via the Riemannian metric. For the convenenience of the reader this background material is summarized in detail in the Appendix.

Let vol_g denote the volume form on (M, g), and

$$\mathrm{Vol}(g) := \int_M \mathrm{vol}_g$$

denote the volume of (M, g). Let ν_g denote the *Riemannian measure* of (M, g); by definition

$$\nu_g(U) = \frac{\int_U \mathrm{vol}_g}{\mathrm{Vol}(g)}.$$

If $f \in C^\infty(M, \mathbb{R})$ then we can view f as a function on S^gM by setting $f(x, v) := f(x)$. In this case we then have

$$\int_{S^gM} f \, d\mu_g = \int_M f \, d\nu_g.$$

2. Topological entropy

An excellent reference for much of this material is [5]. Let $d = d_{\hat{g}}$ denote the distance function on SM obtained from the Sasaki metric $\hat{g}$ (see the Appendix). Given $T > 0$ define a new metric d_T by

$$d_T((x, v), (y, w)) := \max_{t \in [0,T]} d(\phi_t(x, v), \phi_t(y, w)).$$

Since SM is compact, we can define $N^g(T, \varepsilon)$ to be the minimal (finite) number of balls of radius $\varepsilon > 0$ required to cover all of SM under the d_T metric.

Definition 2.1. Define the *topological entropy of* g to be the quantity $h_{\mathrm{top}}(g)$ given by

$$h_{\mathrm{top}}(g) := \lim_{\varepsilon \to 0} \limsup_{T \to \infty} \frac{\log N^g(T, \varepsilon)}{T}.$$

The limit as $\varepsilon \to 0$ exists as the function $\varepsilon \mapsto \limsup_{T\to\infty} \frac{1}{T} \log N^g(T, \varepsilon)$ is monotone decreasing as $\varepsilon \to 0$.

Exercise 2.2. Here are two alternative characterizations of $h_{\mathrm{top}}(g)$. Say that a set $Y \subseteq S^gM$ is (T, ε)*-separated* if for any two points $(x, v), (y, w) \in Y$ it holds that

$$d_T((x, v), (y, w)) > \varepsilon.$$

Then if $S^g(T, \varepsilon)$ denotes the maximum cardinality of a (T, ε)-separated set,

$$h_{\mathrm{top}}(g) = \lim_{\varepsilon \to 0} \limsup_{T \to \infty} \frac{\log S^g(T, \varepsilon)}{T}. \tag{2.1}$$

Similarly, say that a set $Z \subseteq S^gM$ is a (T, ε)*-spanning set* if for any point $(x, v) \in S^gM$ there exists a point $(y, w) \in Z$ such that

$$d_T((x, v), (y, w)) < \varepsilon.$$

Now let $P^g(T, \varepsilon)$ denote the minimal cardinality of a (T, ε)-spanning set. Then

$$h_{\mathrm{top}}(g) = \lim_{\varepsilon \to 0} \limsup_{T \to \infty} \frac{\log P^g(T, \varepsilon)}{T}.$$

The key result we will need that relates the counting function n_g and the topological entropy is the following, which was originally proved by Margulis [7] in his dissertation.

Theorem 2.3. *On a closed Riemannian manifold* (M, g_0) *of negative curvature*

$$\lim_{T\to\infty} \frac{\log n_{g_0}(T)}{T} = h_{\mathrm{top}}(g_0).$$

I will try and sketch the idea behind a particularly elegant proof due to Parry and Pollicott of this theorem. First let us give a suggestive reformulation of the problem. Define

$$\pi_{g_0}(T) := n_g(\log T/h_{\mathrm{top}}(g_0)),$$

that is,

$$\pi_{g_0}(T) = \#\left\{\gamma \in \Lambda(g_0) \,:\, e^{\lambda(\gamma)\cdot h_{\mathrm{top}}(g_0)} \leq T\right\}.$$

To see the point of this, recall that the *prime number theorem* from number theory states that if $\pi(x)$ denotes the numbers of primes less than or equal to x then $\pi(x)$ is asymptotically equivalent to $x/\log x$,

$$\pi(x) \sim \frac{x}{\log x}. \tag{2.2}$$

We shall outline the proof that

$$\pi_{g_0}(T) \sim \frac{T}{\log T}. \tag{2.3}$$

Note that this is in fact a much stronger statement than that of Theorem 2.3.

Returning to the classical number theory setup, to prove (2.2) one introduces the *zeta function* $\zeta : \mathbb{C} \to \mathbb{C}$ defined by

$$\zeta(s) = \prod_p \frac{1}{1 - p^{-s}},$$

where the product is taken of all primes p. A priori, ζ is only defined for $\Re(s) > 1$. However one can show that ζ admits an analytic extension to $\{\Re(s) \geq 1\} \setminus \{1\}$. In fact, ζ is meromorphic on $\Re(s) \geq 1$ with a simple pole at $s = 1$. The Wiener-Ikehara proof of the prime number theorem proceeds from here; see [11].

Definition 2.4. Given a closed geodesic γ, define the *norm* of γ to be $N(\gamma) := e^{\lambda(\gamma)}$.

Define the *zeta function* $\zeta_{g_0}(s)$ for the flow $\phi_t^{g_0}$ to be

$$\zeta_{g_0}(s) := \prod_{\gamma\in\Lambda(g_0)} \frac{1}{1 - N(\gamma)^{-s}},$$

at least where this sum makes sense. In fact, as long as

$$\sum_{\gamma\in\Lambda(g_0)} \sum_{k=1}^{\infty} \frac{1}{k} N(\gamma)^{-sk} \tag{2.4}$$

converges absolutely then $\zeta_{g_0}(s)$ is well defined and

$$\zeta_{g_0}(s) = \exp \sum_{\gamma\in\Lambda(g_0)} \sum_{k=1}^{\infty} \frac{1}{k} N(\gamma)^{-sk},$$

as some elementary computations show.

What is much less easy to see is exactly where the cutoff point is for (2.4) to converge absolutely. For reasons I don't have time to go into fully, it turns out that ζ is well defined, analytic and nonzero precisely when $\Re(s) > h_{\mathrm{top}}(g_0)$. The following theorem due to Parry and Pollicott ([9], Theorem 1) is highly non-trivial.

Theorem 2.5. *ζ_{g_0} has a nowhere vanishing analytic extension to an open neighborhood of $\Re(s) \geq h_{\mathrm{top}}(g_0)$ apart from a simple pole at $s = h_{\mathrm{top}}(g_0)$.*

Having got this theorem, essentially the same proof as the (Wiener-Ikehara) proof of the prime number theorem goes through, and this proves (2.3).

3. Volume growth of balls

An excellent reference for this section is [10]. Here is the main theorem we wish to prove in this lecture.

Theorem 3.1. *Suppose (M, g_0) is a closed n-dimensional Riemannian manifold of constant non-positive curvature K. Then*

$$\lim_{T\to\infty} \frac{\log n_{g_0}(T)}{T} = (n-1)\sqrt{-K}.$$

Let $\widetilde{M}$ denote the universal cover of M, and let $B(x,r)$ denote a ball with centre x and radius r in $\widetilde{M}$. Let

$$V_g(x,r) := \mathrm{Vol}(B(x,r), g).$$

Then a result of Manning [6] tells us that

Theorem 3.2. *For any $x \in \widetilde{M}$, the limit*

$$b_g := \lim_{r\to\infty} \frac{\log V_g(x,r)}{r}$$

exists and is independent of $x \in \widetilde{M}$.

Proof. Choose a compact fundamental domain[1] $K \subseteq \widetilde{M}$ for the fundamental group $\pi_1(M)$ acting by isometries on $\widetilde{M}$. Let $k := \operatorname{diam} K$. Then for any $r > k$ and $x, y \in K$ it holds that

$$B(x, r-k) \subseteq B(y,r) \subseteq B(x, r+k),$$

and thus we deduce that

$$V_g(x, r-k) \leq V_g(y,r) \leq V_g(x, r+k), \tag{3.1}$$

since the covering transformations that bring x into y are isometries. Note that if we knew the limit $b_g(x)$ existed, (3.1) would tell us that the limit is independent of the choice of x.

Fix $x \in K$. It remains to check that $\lim_{r\to\infty} \frac{1}{r} \log V_g(x,r)$ exists. We may assume $V_g(x,r)$ is unbounded as $r \to \infty$, as otherwise there is nothing to prove.

Here is the key observation. Suppose $S \subseteq B(x,r)$ is such that for all $p, q \in S$ it holds that $d(p,q) = \varepsilon$. Then it is easily seen that

$$\#S \leq \frac{V_g(x, r+\varepsilon/2)}{\inf_{y\in B(x,r)} V_g(y, \varepsilon/2)}.$$

Suppose S is maximal with this property: then every point of $B(x,r)$ lies within ε of of some point of S. It thus holds that for any $\delta > 0$,

$$B(x, r+\delta) \subseteq \bigcup_{p\in S} B(p, \delta+\varepsilon).$$

---

[1] A *compact fundamental domain* K is defined as follows. Let d denote the distance function on $\widetilde{M}$. Fix a point $x_0 \in \widetilde{M}$ and put

$$U := \left\{ x \in \widetilde{M} \,:\, d(x, x_0) < d(x, \varphi x_0) \text{ for all } \varphi \neq \mathrm{Id} \in \pi_1(M) \right\}.$$

Then U is open, and contains no two points that are equivalent under the operation of $\pi_1(M)$. Let $K := \overline{U}$ denote the closure of U. Then K is compact, and, given any $x \in \widetilde{M}$ there exists $\varphi \in \pi_1(M)$ such that $\varphi x \in K$.

Now choose ε such that

$$\inf_{y \in B(x,r)} V_g(y, \varepsilon/2) = 1,$$

and set $c := k + \frac{3}{2}\varepsilon$. Then

$$\begin{aligned} V_g(x, r+\delta) &\leq \sum_{p \in S} V_g(p, \delta + \varepsilon/2) \\ &\leq \#S \cdot V_g(x, \delta + \varepsilon/2 + k) \\ &\leq V_g(x, \delta + \varepsilon/2) \cdot V_g(x, \delta + \varepsilon/2 + k), \end{aligned}$$

where the second inequality used (3.1). But now if $r' = r - \varepsilon/2$ and $\delta' = \delta + \varepsilon/2$ then $r + \delta = r' + \delta'$ and we obtain

$$V_g(x, r+\delta) = V_g(x, r' + \delta') \leq V_g(x, r) \cdot V_g(x, \delta + c).$$

Now fix $\delta > 0$, and select $n \in \mathbb{N}$ such that $n\delta \leq r < (n+1)\delta$. Iterating the above tells us that

$$\begin{aligned} V_g(x, r) &\leq V_g(x, (n+1)\delta) \\ &\leq V_g(x, n\delta) \cdot V_g(x, \delta + c) \\ &\leq \dots \\ &\leq V_g(x, \delta) \cdot (V_g(x, \delta + c))^n . \end{aligned}$$

Thus

$$\begin{aligned} \frac{\log V_g(x,r)}{r} &\leq \frac{\log V_g(x,\delta)}{r} + \frac{n \log V_g(x, \delta + c)}{r} \\ &\leq \frac{\log V_g(x,\delta)}{r} + \frac{\log V_g(x, \delta + c)}{\delta}. \end{aligned}$$

Hence

$$\limsup_{r \to \infty} \frac{\log V_g(x,r)}{r} \leq \frac{\log V_g(x, \delta + c)}{\delta}.$$

This holds for all $\delta > 0$, and so taking the limit inferior as $\delta \to \infty$ on the right-hand side gives

$$\begin{aligned} \limsup_{r \to \infty} \frac{\log V_g(x,r)}{r} &\leq \liminf_{\delta \to \infty} \frac{\log V_g(x, \delta + c)}{\delta} \\ &= \liminf_{\delta \to \infty} \frac{\log V_g(x, \delta)}{\delta}, \end{aligned}$$

and hence the limit exists. ☐

The next result, *Manning's inequality* (also from [6]), links b_g and $h_{\text{top}}(g)$.

Theorem 3.3. *It holds that*

$$h_{top}(g) \geq b_g. \tag{3.2}$$

Proof. Since certainly, $h_{\text{top}}(g) \geq 0$, we may assume $b_g > 0$ or else there is nothing to prove. Fix a point $x \in \widetilde{M}$, and consider the annuli

$$A_\delta(r) := B(x, r + \delta/2) \backslash B(x, r).$$

Fix an arbitrary $\varepsilon \in (0, b_g)$ and choose $\delta > 0$ small. I will leave it as an easy exercise to check that there exists a sequence $(r_n) \subseteq \mathbb{R}^+$ with $r_n \to \infty$ such that

$$\text{Vol}_g(A_\delta(r_n)) \geq e^{(b_g - \varepsilon) r_n}$$

for all $n \in \mathbb{N}$ (essentially, if this fails then the growth rate of $V_g(x,r)$ cannot be b_g; see [10], p70 if you get stuck).

Now for each $n \in \mathbb{N}$, choose a subset $Q_\delta(r_n) \subseteq A_\delta(r_n)$ such that for all $p, q \in Q_\delta(r_n)$ it holds that $d(p,q) > 2\delta$, and that $Q_\delta(x, r_n)$ is maximal with this property. Then

$$\# Q_\delta(r_n) \geq \frac{\operatorname{Vol}_g(A_\delta(r_n))}{\sup_{y \in \widetilde{M}} V_g(y, 2\delta)} \geq c_\delta e^{(b_g - \varepsilon) r_n},$$

where $c_\delta > 0$ is a constant depending on δ.

Here is the key observation for this proof. $\widetilde{M}$ is complete and thus any two points of $\widetilde{M}$ can be joined by a minimizing geodesic. In particular, x can be joined to each $q \in Q_\delta(r_n)$ by a geodesic of length between r_n and $r_n + \delta/2$. Call this geodesic γ_q. Now consider the set

$$Y_\delta(r_n) := \bigcup_{q \in Q_\delta(r_n)} (\gamma_q(0), \dot{\gamma}_q(0)) \subseteq S\widetilde{M}.$$

It is not hard to see straight from the definition that $Y_\delta(r_n)$ is a (r_n, δ)-separated set for the lifted flow $\widetilde{\phi}_t$ on $S\widetilde{M}$. Recall this means that if $d_n = d_{r_n}$ denotes the metric on $S\widetilde{M}$ defined as above by

$$d_{r_n}((y,v),(z,w)) := \max_{t \in [0, r_n]} d(\widetilde{\phi}_t(y,v), \widetilde{\phi}_t(z,w)),$$

then for all $(y,v),(z,w) \in Y_\delta(r_n)$ it holds that

$$d_{r_n}((y,v),(z,w)) > \delta.$$

Let ι_g denote the injectivity radius of (M,g). If $P : \widetilde{M} \to M$ denotes the covering projection then given $y, z \in \widetilde{M}$ with $d(y,z) < \iota_g$ it holds that

$$d(P(y), P(z)) = d(y,z).$$

It follows that $dP[Y_\delta(r_n)] \subseteq SM$ is $(r_n, \frac{1}{2}\min\{\delta, \iota_g\})$-separated set for $\phi_t : SM \to SM$. Then by our alternative characterization (2.1) of $h_{\text{top}}(g)$ it holds that

$$h_{\text{top}}(g) \geq \limsup_{n \to \infty} \frac{\log \# dP[Y_\delta(r_n)]}{r_n} \geq b_g - \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows $h_{\text{top}}(g) \geq b_g$ as required. □

Finally, to complete this circle of ideas it is known that if g has no conjugate points (which is certainly the case for $g = g_0$ a negatively curved metric) then we have actually have equality in (3.2) (see [10], Lemma 3.39 for a proof). Putting this together with Theorem 2.3 we deduce the following corollary.

Corollary 3.4. *Let (M, g_0) be a negatively curved Riemannian manifold. Then*

$$\lim_{T \to \infty} \frac{\log n_{g_0}(T)}{T} = b_{g_0}.$$

We can now complete the proof of Theorem 3.1. We simply need to compute b_{g_0} for g_0 a metric of non-positive curvature K on M. For $K = 0$, since the volume of a ball in Euclidean space grows polynomially we have $b_g = 0$. For $K < 0$, since the volume of a ball in hyperbolic space of constant curvature K grows exponentially with exponent $(n-1)\sqrt{-K}$, we have in this case that $b_g = (n-1)\sqrt{-K}$.

Lecture Two: A Theorem of Katok

4. Metric entropy

Let (M,g) denote a Riemannian manifold and $\phi_t^g : S^gM \to S^gM$ the geodesic flow. Recall that a measure m on S^gM is *ϕ_t^g-invariant* if for any Borel set $B \subseteq S^gM$, we have $m(\phi_t^g[B]) = m(B)$. Let $\mathcal{M}(\phi^g)$ denote the set of all invariant Borel probability measures m on S^gM. It is known that $\mathcal{M}(\phi^g)$ is nonempty. As an example of an invariant Borel probability measure, if $\gamma \in \Lambda(T;g)$ then $\dot\gamma$ is a closed orbit of ϕ_t^g with period T, and we can define a measure $m_{\dot\gamma} \in \mathcal{M}(\phi^g)$ by

Now fix $m \in \mathcal{M}(\phi^g)$.

$$\int_{S^gM} f\, dm_{\dot\gamma} := \frac{1}{T}\int_0^T f(\phi_t v)dt,$$

where $v \in \dot\gamma$. We call $m_{\dot\gamma}$ the *δ-measure* of the closed orbit $\dot\gamma$.

Definition 4.1. A *partition* of S^gM is a disjoint finite collection $\alpha = \{A_1,\dots,A_k\}$ of Borel sets whose union is all of S^gM. If $\alpha = \{A_1,\dots,A_k\}$ and $\beta = \{B_1,\dots,B_n\}$ are two partitions then the *join* $\alpha \vee \beta$ is the partition given by

$$\alpha \vee \beta = \{A_i \cap B_j \,:\, i = 1,\dots,k,\ j = 1,\dots,n\}.$$

Similarly we can form the join $\bigvee_i \alpha_i$ of multiple partitions $\alpha_1,\dots,\alpha_\ell$.

Definition 4.2. Given a Borel probability measure m, the *entropy of a partition* $\alpha = \{A_1,\dots,A_k\}$ *with respect to* m is given by

$$H_m(\alpha) := -\sum_{i=1}^{k} m(A_i)\log m(A_i),$$

where $0\log 0$ is interpreted to be 0.

Let $\alpha = \{A_1,\dots,A_k\}$ a partition. Define $\phi_{-n}^g(\alpha)$ to be the partition

$$\phi_{-n}^g(\alpha) = \left\{\phi_{-n}^g(A_1),\dots,\phi_{-n}^g(A_k)\right\}.$$

Definition 4.3. Let $m \in \mathcal{M}(\phi^g)$. The *metric entropy of ϕ_t^g with respect to m* is defined to be

$$h_m(\phi^g) := \sup_\alpha \lim_{n\to\infty} \frac{1}{2} H_m\left(\bigvee_{i=0}^{n-1} \phi_{-n}^g(\alpha)\right),$$

where the supremum is taken over all partitions α.

How are the metric and topological entropy related? The metric entropy $h_m(\phi^g)$ should be thought of as a quantitative estimate of the complexity of the flow ϕ_t^g from the point of view of the invariant measure m. It can be shown that the metric entropy of the union of two invariant sets is the sum (suitably weighted by m) of the entropy of each of the sets individually, and so in some sense the metric entropy measures the average complexity of ϕ_t^g. Meanwhile the topological entropy measures the global maximal complexity of ϕ_t^g. It thus makes sense to assume that the metric entropy is at most the topological entropy, with the metric entropy maximized for measures that assign the most weight to areas of high complexity.

This is indeed what happens; this is called the *Variational Principle*:

Theorem 4.4. *Let (M,g) denote a closed Riemannian manifold. Then*

$$h_{\text{top}}(g) = \sup_{m\in\mathcal{M}(\phi^g)} h_m(\phi^g).$$

Moreover **measures of maximum entropy***, that is, measures $m \in \mathcal{M}(\phi^g)$ such that $h_m(\phi^g) = h_{\mathrm{top}}(g)$ always exist.*

The first statement of the theorem is actually true for any continuous flow on a compact metric space; for a proof of this see Theorem 4.5.3 on p181 of [5]. The second statement is due to Newhouse; see [8]. Section 4.5 of [5] also contains a much fuller discussion on the difference between metric entropy and topological entropy. On a surface of constant negative curvature one can actually say substantially more, namely, the following theorem holds.

Theorem 4.5. *If (Σ, g_0) is a surface of constant negative curvature then there exists a unique measure of maximal entropy in $\mathcal{M}(\phi^{g_0})$, and this is precisely the Liouville measure μ_{g_0}.*

Hence

$$h_{\mathrm{top}}(g_0) = h_{\mu_{g_0}}(\phi^{g_0}),$$

meanwhile for any other $m \in \mathcal{M}(\phi^{g_0})$ it holds that

$$h_{\mathrm{top}}(g_0) > h_m(\phi^{g_0}).$$

Recall that Theorem 2.3 told us that on a manifold of negative curvature, the exponential growth of the counting function is given by the topological entropy. Katok ([3], Theorem 1.8 or [2], Section 4) has extended this theorem as follows.

Theorem 4.6. *Suppose (Σ, g) is a closed surface. Then* **without assuming negative curvature** *it holds that*

$$\liminf_{T\to\infty} \frac{\log n_g(T)}{T} \geq h_{\mathrm{top}}(g).$$

We will not discuss the details of this proof. Despite the similarity in the statement of Theorem 4.6 to that of Theorem 2.3, the proof of Theorem 4.6 requires several new ideas, which would unfortunately take us too far afield to cover.

5. A Theorem of Katok

Instead, we will devote the remainder of this second lecture to proving the following theorem (also due to Katok [3, 4]).

Theorem 5.1. *Let Σ be an orientable surface of Euler characteristic $\chi(\Sigma) < 0$. Then for any Riemannian metric g on Σ with normalized volume $\mathrm{Vol}(g) = 1$, it holds that*

$$\liminf_{T\to\infty} \frac{\log n_g(T)}{T} \geq \sqrt{-2\pi\chi(\Sigma)},$$

with equality if and only if g has constant curvature.

The first tool we will use is the following classical result of Koebe.

Theorem 5.2. *(Conformal equivalence theorem)*

Let Σ be an orientable surface of Euler characteristic $\chi(\Sigma) < 0$. Then for any Riemannian metric g on Σ there exists a unique function $\rho \in C^\infty(M, \mathbb{R})$ with

$$\int_\Sigma \rho \, d\nu_g = 1 \tag{5.1}$$

such that the metric $g_0 := \rho g$ has constant negative curvature.

Note that the Riemannian measures of g and g_0 are related by

$$\nu_{g_0} = \frac{\mathrm{Vol}(g)}{\mathrm{Vol}(g_0)} \rho^{1/2} \nu_g, \tag{5.2}$$

and hence the (5.1) implies that

$$\mathrm{Vol}(g) = \mathrm{Vol}(g_0).$$

Next, by Jensen's inequality, if $c(\rho)$ denotes the *conformality constant*

$$c(\rho) := \int_\Sigma \rho^{1/2}\, d\nu_g,$$

then

$$c(\rho) \le \left(\int_\Sigma \rho\, d\nu_g\right)^{1/2} = 1. \tag{5.3}$$

Definition 5.3. Let $f \in C^0(S^{g_0}\Sigma, \mathbb{R})$. Denote by

$$\langle f\rangle_{g_0} := \int_{S^{g_0}\Sigma} f\, d\mu_{g_0}$$

the *space average* of f.

Similarly given a closed curve γ of period T, let

$$f_{(\gamma)} := \frac{1}{T}\int_0^T f(\gamma(t))dt$$

denote the *time average* of f along γ.

The first part of the proof of Theorem 5.1 is to exploit Theorem 4.5; namely, the fact that if g_0 is a metric of constant negative curvature then the Liouville measure μ_{g_0} maximizes the metric entropy.

Lemma 5.4. *Let g_0 denote a metric of constant negative curvature on Σ. Then for any $\varepsilon > 0$, it holds that for all $f \in C^0(S^{g_0}\Sigma, \mathbb{R})$, the proportion of closed geodesics γ whose time average of f along γ is within ε of the space average of f tends to* 1 *as $T \to \infty$. In symbols:*

$$\frac{\left|\{\gamma \in \Lambda(g_0;T) \,:\, \left|f_{(\gamma)} - \langle f\rangle_{g_0}\right| < \varepsilon\}\right|}{n_{g_0}(T)} \to 1 \quad \text{as } T \to \infty.$$

Proof. Suppose not. Then there exists $f \in C^0(S^{g_0}\Sigma, \mathbb{R})$, a sequence T_n with $T_n \to \infty$ and $\varepsilon_0 > 0$ such that for all $n \in \mathbb{N}$, more than $\varepsilon_0 n_{g_0}(T_n)$ of the elements $\gamma \in \Lambda(T_n; g_0)$ satisfy $\left|f_{(\gamma)} - \langle f\rangle_{g_0}\right| > \varepsilon_0$. Let us suppose that actually $f_{(\gamma)} > \langle f\rangle_{g_0} + \varepsilon_0$ (the other case is similar).

Given $\gamma \in \Lambda(g_0)$ there exists a corresponding orbit $\dot\gamma$ of $\phi_t^{g_0} : S^{g_0}\Sigma \to S^{g_0}\Sigma$, and thus a δ-measure $m_{\dot\gamma} \in \mathcal{M}(\phi^{g_0})$. Let m_n denote the average of the δ-measures $m_{\dot\gamma}$ for $\gamma \in \Lambda(T_n; g_0)$. Then for all $n \in \mathbb{N}$, we have

$$\int_{S^{g_0}\Sigma} f\, dm_n > \int_{S^{g_0}\Sigma} f\, d\mu_{g_0} + \varepsilon_0. \tag{5.4}$$

But now suppose m is a weak limit of the measures m_n. Then it can be shown (see [2], Section 4) that

$$h_m(\phi^{g_0}) \ge \lim_{n\to\infty} \frac{\log \varepsilon_0 n_{g_0}(T_n)}{T_n},$$

and by Theorem 2.3 the right-hand side of the equation above is equal to $h_{\mathrm{top}}(g_0)$.This contradicts Theorem 4.5, since by (5.4), m cannot be the Liouville measure. □

Assume now the hypotheses of Theorem 5.1. The conformal equivalence theorem guarantees the existence of a function $\rho \in C^\infty(\Sigma, \mathbb{R})$ such that $g_0 := \rho g$ has constant negative curvature. We will apply Lemma 5.4 to the function

$$f = \rho^{-1/2}.$$

Here are two good reasons why this is a sensible choice for f. Firstly, note that the space average of f is given by

$$\begin{aligned}\langle f\rangle_{g_0} &= \int_{S^{g_0}\Sigma}\rho^{-1/2}\,d\mu_{g_0}\\ &= \int_{\Sigma}\rho^{-1/2}\,d\nu_g\\ &= \int_{\Sigma}\rho^{1/2}\,d\nu_g\\ &= c(\rho);\end{aligned}$$

the first equality holding since ρ is a function on Σ only, and the second since $\int_\Sigma \rho\,d\nu_g = 1$.

Secondly, the time averages of f are also useful. Indeed, let $l_g(\gamma)$ denotes the g-length of a curve, and similarly $l_{g_0}(\gamma)$ denotes the g_0-length. Then if $\gamma\in\Lambda(g_0)$ it follows from (5.2) that

$$l_g(\gamma) = l_{g_0}(\gamma)\cdot f_{(\gamma)}. \tag{5.5}$$

Now let us denote by $\Lambda_\varepsilon(g_0)$ the subset of $\Lambda(g_0)$ such that for all $\gamma\in\Lambda_\varepsilon(g_0)$, the time average of f along γ is within ε of the space average $\langle f\rangle_{g_0}$, and similarly let $\Lambda_\varepsilon(T;g_0) = \Lambda_\varepsilon(g_0)\cap\Lambda(T;g_0)$. By Lemma 5.4 we may choose $\varepsilon>0$ and $T>0$ such that

$$\#\Lambda_\varepsilon(T;g_0) \geq (1-\varepsilon)n_{g_0}(T). \tag{5.6}$$

Note also that for any $\gamma\in\Lambda_\varepsilon(T;g_0)$, (5.5) implies that

$$l_g(\gamma) < (c(\rho)+\varepsilon)T. \tag{5.7}$$

In order to proceed further we need to study the free homotopy classes of Σ. Recall that a classical theorem in Riemannian geometry (see for instance [1], Theorem IV.5.1) states that for any Riemannian manifold (M,g), every non-trivial free homotopy class $\sigma\in\Pi(M)$ contains a shortest curve, and every such curve is a closed geodesic. Moreover if g has negative curvature and M is closed, such a closed geodesic is unique.

Definition 5.5. Let $\Pi(\Sigma)$ denote the set of free homotopy classes of Σ. Let $s_g(T)$ denote the number of free homotopy classes σ such that there exists $\gamma\in\Lambda(T;g)$ representing σ. Then the theorem quoted above implies that

$$s_g(T)\leq n_g(T),$$

and for a metric g_0 of negative curvature,

$$s_{g_0}(T) = n_{g_0}(T).$$

Returning to the proof, fix $\gamma\in\Lambda_\varepsilon(T;g_0)$. Then γ represents a unique free homotopy class σ, and in σ there exists a closed g-geodesic γ'. Then as γ' is then a shortest g-curve belonging to σ, we have by (5.7),

$$l_g(\gamma')\leq l_g(\gamma) < (c(\rho)+\varepsilon))T.$$

Coupling this with (5.6), we have shown

$$s_g((c(\rho)+\varepsilon)T) \geq (1-\varepsilon)s_{g_0}(T). \tag{5.8}$$

The reason for introducing the function s_g is that we can compare it to the function N^g used in the definition of the topological entropy. Indeed, there exists a constant $\delta>0$ depending only on g such that if $\sigma\neq\sigma'\in\Pi(\Sigma)$ and $\gamma,\gamma'\in\Lambda(g)$

with γ representing σ and γ' representing σ' then $\dot\gamma_1$ and $\dot\gamma_2$ are δ-*separated*, that is, for any $s, t \in \mathbb{R}$ we have

$$d((\gamma_1(s), \dot\gamma_1(s)), (\gamma_2(t), \dot\gamma_2(t))) \geq \delta.$$

From this it follows that

$$N^g(T, \delta) \geq s_g(T). \tag{5.9}$$

We will now use (5.8) and (5.9) to complete the proof of Theorem 5.1. Observe that we have:

$$\begin{aligned} h_{\mathrm{top}}(g) &\geq \limsup_{T\to\infty} \frac{\log N^g((c(\rho)+\varepsilon)T, \delta)}{(c(\rho)+\varepsilon)T} \\ &\geq \limsup_{T\to\infty} \frac{\log\, s_g((c(\rho)+\varepsilon)T)}{(c(\rho)+\varepsilon)T} \\ &\geq \limsup_{T\to\infty} \frac{\log\,(1-\varepsilon)s_{g_0}(T)}{(c(\rho)+\varepsilon)T} \\ &= \frac{1}{c(\rho)+\varepsilon} \limsup_{T\to\infty} \frac{\log\,(1-\varepsilon)n_{g_0}(T)}{T} \\ &= \frac{1}{c(\rho)+\varepsilon} h_{\mathrm{top}}(g_0), \end{aligned}$$

where the last inequality used Theorem 2.3. Since we could have chosen $\varepsilon > 0$ arbitrarily small, we conclude that

$$h_{\mathrm{top}}(g) \geq \frac{1}{c(\rho)} h_{\mathrm{top}}(g_0).$$

Now use Theorem 2.3 once again to obtain:

$$\begin{aligned} \liminf_{T\to\infty} \frac{\log\, n_g(T)}{T} &\geq h_{\mathrm{top}}(g) \\ &\geq \frac{1}{c(\rho)} h_{\mathrm{top}}(g_0) \\ &\geq \sqrt{-K}, \end{aligned}$$

where g_0 has constant curvature $K < 0$, and on the last line we used (5.3) and Theorem 3.1.

Finally, by the Gauss-Bonnet theorem,

$$\int_\Sigma K\, d\nu_{g_0} = 2\pi\chi(\Sigma),$$

and Theorem 5.1 follows.

We will conclude by saying a word about the case of equality. It is clear from (5.3) that we *cannot* have equality if g does not have constant curvature, however rather more work is required to show that all the other inequalities become equalities when g is of constant curvature.

Appendix A. The Geometry of the Tangent Bundle

In the Appendix we recall some elementary material on the geometry of the tangent bundle and the structure of the geodesic flow. No proofs are included - for these we refer the reader to Chapter 1 of [10]. Let (M, g) denote a closed orientable n-dimensional Riemannian metric. We let $\pi : TM \to M$ denote the projection map of the tangent bundle and $\pi^g : S^gM \to M$ denote its restriction to S^gM.

Definition A.1. The *vertical subbundle* $\mathcal{V} \subseteq TTM$ is given by the kernel of the projection π:

$$\mathcal{V}(x,v) := \ker\left\{d_{(x,v)}\pi : T_{(x,v)}TM \to T_xM\right\} \subseteq T_{(x,v)}TM.$$

Alternatively, one can think of $\mathcal{V}(x,v)$ as being the tangent space to the fibre $\mathcal{V}(x,v) = T_{(x,v)}T_xM$.

We aim to construct another subbundle called the *horizontal subbundle* such that TTM splits as a direct sum of the horizontal and vertical subbundles. Unlike the vertical subbundle, this subbundle will depend on the metric. Let ∇^g denote the Levi-Civita of (M,g).

Definition A.2. The *connection map* of ∇^g is a bundle morphism $K^g : TTM \to TM$ defined as follows. Fix $(x,v) \in TM$. Given $\xi \in T_{(x,v)}TM$, let $\Gamma : (-\varepsilon,\varepsilon) \to TM$ be a curve *adapted* to ξ, that is, $\Gamma(0) = (x,v)$ and $\dot{\Gamma}(0) = \xi$. Let $\gamma : (-\varepsilon,\varepsilon) \to M$ denote the curve $\gamma = \pi \circ \Gamma$. Then we can write $\Gamma(t) = (\gamma(t), V(t))$, where $V(t)$ is a vector field along γ. We then define $K^g_{(x,v)}(\xi) \in T_xM$ to be the vector

$$K^g_{(x,v)}(\xi) = \nabla^g_{\dot{\gamma}}V(0).$$

Now we can define the horizontal subbundle to be the kernel of K^g.

Definition A.3. The *horizontal subbundle* $\mathcal{H}^g \subseteq TTM$ is given by

$$\mathcal{H}^g(x,v) := \ker\left\{K^g_{(x,v)} : T_{(x,v)}TM \to T_xM\right\} \subseteq T_{(x,v)}TM.$$

The following result is well known. For a proof see for instance [10], p13.

Proposition A.4. *The maps*

$$d_{(x,v)}\pi|\mathcal{H}^g(x,v) : \mathcal{H}^g(x,v) \to T_xM$$

and

$$K^g_{(x,v)}|\mathcal{V}(x,v) : \mathcal{V}(x,v) \to T_xM$$

are linear isomorphisms. Moreover

$$TTM = \mathcal{H}^g \oplus \mathcal{V},$$

and hence if $f^g : TTM \to TM \times TM$ *denotes the map given fibrewise by*

$$f^g_{(x,v)}(\xi) = (d_{(x,v)}\pi(\xi), K^g_{(x,v)}(\xi))$$

then f^g *is a linear isomorphism.*

Given $\xi \in TTM$ we write $f^g(\xi) = (\xi_h, \xi_v)$ and call ξ_h and ξ_v the *horizontal* and *vertical components* of ξ respectively. We will use these components to turn TM into a Riemannian manifold $(TM, \hat{g})$.

Definition A.5. We define the *Sasaki metric* $\hat{g}$ on TM by

$$\hat{g}_{(x,v)}(\xi,\eta) := g_x(\xi_h,\eta_h) + g_x(\xi_v,\eta_v) \quad \text{for } \xi,\eta \in T_{(x,v)}TM.$$

Under this metric the two subbundles $\mathcal{H}^g$ and $\mathcal{V}$ become orthogonal, and the projection $\pi : TM \to M$ becomes a Riemannian submersion with totally geodesic fibres.

Definition A.6. Define a 2-form $\omega_g \in \Omega^2(TM)$ by

$$\omega_g|_{(x,v)}(\xi,\eta) := g_x(\xi_h,\eta_v) - g_x(\xi_v,\eta_h) \quad \text{for } \xi,\eta \in T_{(x,v)}TM.$$

Define a 1-form $\alpha_g \in \Omega^1(TM)$ by setting

$$\alpha_g|_{(x,v)}(\xi) := g_x(\xi_h, v) \quad \text{for } \xi \in T_{(x,v)}TM.$$

In order to see the importance of these two differential forms, let us first make a few more definitions. We begin with an exercise.

Exercise A.7. Given $\xi \in T_{(x,v)}TM$, show that $\xi \in T_{(x,v)}S^gM$ if and only if $g_x(\xi_v, v) = 0$.

Let X^g denote the infinitesimal generator of the geodesic flow ϕ_t^g. We call X^g the *geodesic vector field.* Then X^g determines a 1-dimensional distribution $\mathbb{R}X^g \subseteq TS^gM$. Let $W^g(x,v) \subseteq T_{(x,v)}S^gM$ denote the orthogonal complement of $\mathbb{R}X^g$ with respect to the Sasaki metric, that is,

$$W^g(x,v) := \left\{\xi \in T_{(x,v)}S^gM \,:\, \hat{g}_{(x,v)}(\xi, X^g(x,v)) = 0\right\}.$$

The following exercise tells us that we can alternatively write

$$W^g(x,v) = \left\{\xi \in T_{(x,v)}S^gM \,:\, g_x(\xi_h, v) = 0\right\} = \ker\left\{\alpha_g|_{(x,v)}\right\}.$$

Exercise A.8. Show that

$$X^g(x,v)_h = v \quad \text{and} \quad X^g(x,v)_v = 0.$$

Conclude that α_g is the *dual* 1*-form* to X^g under the Sasaki metric $\hat{g}$, that is,

$$\alpha_g|_{(x,v)}(\xi) = \hat{g}_{(x,v)}(X^g(x,v), \xi) \quad \text{for } \xi \in T_{(x,v)}TM.$$

Now define two $(n-1)$-subbundles of TS^gM H^g and V by setting

$$H^g := \mathcal{H}^g \cap W^g, \quad V = \mathcal{V} \cap TS^gM.$$

Then we have

$$W^g = H^g \oplus V.$$

Exercise A.9. Use the *Gauss lemma* from Riemannian geometry to show that W^g is invariant under the differential of the geodesic flow, that is,

$$d_{(x,v)}\phi_t^g[W^g(x,v)] = W^g(\phi_t^g(x,v)).$$

We can finally state the two main theorems of this Appendix.

Theorem A.10. *The* 2*-form* ω_g *is the minues the exterior derivative of the* 1*-form* α_g*, that is,*

$$\omega_g = -d\alpha_g.$$

Moreover these forms are invariant under the geodesic flow: $(\phi_t^g)^*\alpha_g = \alpha_g$ *and* $(\phi_t^g)^*\omega_g = \omega_g$.

Theorem A.11. *The* 2*-form* ω_g *makes* TM *into a* $2n$*-dimensional symplectic manifold. The subbundles* $\mathcal{H}^g$ *and* $\mathcal{V}$ *are Lagrangian subbundles of* TTM*. The* 1*-form* α_g *makes* S^gM *into a* $(2n-1)$*-dimensional contact manifold. Moreover if* $\bar{\omega}_g$ *denotes the restriction to* W^g *of* ω_g *then* $(W^g, \bar{\omega}_g)$ *is a* $(2n-2)$*-dimensional symplectic manifold. The subbundles* H^g *and* V *are Lagrangian subbundles of* W^g*.*

As a consequence of the above two results, $\alpha_g \wedge d\alpha_g^{n-1}$ is a volume form on S^gM and thus defines a probability measure μ_g.

Definition A.12. Let μ_g denote the probabilitiy measure on S^gM defined by

$$\int_{S^gM} f\, d\mu_g := \frac{\int_{S^gM} f \cdot \alpha \wedge d\alpha_g^{n-1}}{\int_{S^gM} \alpha \wedge d\alpha_g^{n-1}} \quad \text{for } f \in C^0(S^gM, \mathbb{R}).$$

This probability measure μ_g is called the *Liouville measure.* Since $(\phi_t^g)^*\alpha_g = \alpha_g$, the Liouville measure is invariant under the geodesic flow, that is,

$$\mu_g(B) = \mu_g(\phi_t^g[B])$$

for any Borel measurable set $B \subseteq S^gM$.

As mentioned briefly in the introduction, there is another way to obtain the forms ω_g and α_g. We conclude with this. Let $\Phi_g : TM \to T^*M$ denote the isomorphism given by the Riemannian metric, that is, given $v \in T_xM$ we define $\Phi_g(v) = p \in T_x^*M$ by

$$p(w) = g_x(v, w) \quad \text{for } w \in T_xM.$$

Definition A.13. Let $\tau : T^*M \to M$ denote the projection map of the cotangent bundle. Define a 1-form $\lambda \in \Omega^1(T^*M)$ called the *Liouville* 1*-form* by

$$\lambda_{(x,p)}(\xi) = p(d_{(x,p)}\tau(\xi)) \quad \text{for } \xi \in T_{(x,p)}T^*M,$$

and define a 2-form $\omega_0 \in \Omega^2(T^*M)$ by $\omega_0 = -d\lambda$.

Proposition A.14. *The* 2*-form* ω_0 *makes* T^*M *into a* $2n$*-dimensional symplectic manifold. Moreover if* $S^{g*}M$ *denotes the unit cotangent bundle given by*

$$S_x^{g*}M = \left\{p \in T_x^*M \,:\, g_x(\Phi_g^{-1}(p), \Phi_g^{-1}(p)) = 1\right\}$$

then the Liouville 1*-form* λ *makes* $S^{g*}M$ *into a* $(2n-1)$*-dimensional contact manifold.*

This symplectic structure on T^*M is related to the structure on TM via Φ_g.

Theorem A.15. *The* 2*-forms* ω_g *and* ω_0 *are related by*

$$\omega_g = \Phi_g^*\omega_0.$$

The 1*-forms* α_g *and* λ *are related by*

$$\alpha_g = \Phi_g^*\lambda.$$

REFERENCES

[1] I. Chavel. *Riemannian Geometry: A Modern Introduction*, volume 98 of *Cam. Studies. in Adv. Math.* Cambridge University Press, 2nd edition, 2006.
[2] A. Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Publ. Math. IHES*, 51:137–173, 1980.
[3] A. Katok. Entropy and closed geodesics. *Ergod. Th. & Dynam. Sys.*, 2:339–367, 1982.
[4] A. Katok. Four applications of conformal equivalence to geometry and dynamics. *Ergod. Th. & Dynam. Sys.*, 8*:139–152, 1988.
[5] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*, volume 54 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, 1995.
[6] A. Manning. Topological entropy for geodesic flows. *Ann. Math.*, 110:567–573, 1979.
[7] G. Margulis. *On some problems in the theory of U-systems.* PhD thesis, Moscow State University, 1970.
[8] S. Newhouse. Continuity properties of entropy. *Ann. Math.*, 129:215–235, 1989.
[9] W. Parry and M. Pollicott. An Analogue of the Prime Number Theorem for Closed Orbits of Axiom A Flows. *Ann. of Math.*, 118(3):573–591, 1983.
[10] G. P. Paternain. *Geodesic Flows*, volume 180. Birkhäuser Verlag, 1999.
[11] N. Wiener. *The Fourier Integral and Certain of its Applications.* Cambridge University Press, 1967.