THE CIRCLE PROBLEM ON SURFACES OF VARIABLE NEGATIVE CURVATURE

Mark Pollicott and Richard Sharp

Warwick University and Oxford University

Abstract. In this note we study the problem of orbit counting for certain groups of isometries of simply connected surfaces with possibly variable negative curvature. We show that if $N(t)$ denotes the orbit counting function for a convex co-compact group of isometries then for some constants $C, h > 0$, $N(t) \sim Ce^{ht}$, as $t \to +\infty$.

0. Introduction

This paper addresses the so-called circle problem on surfaces of variable negative curvature. Let X be a complete simply connected Riemannian manifold and let Γ be a discrete group of isometries of X. Pick a point $p \in X$ and define a counting function $N(t) = Card(B(p,t) \cap \Gamma p)$, where $B(p,t)$ is the ball of radius t centred at p. The circle problem is to determine how $N(t)$ behaves as $t \to +\infty$. The classical circle problem is where X is taken to be $\mathbb{R}^2$ and Γ is taken to be $\mathbb{Z}^2$ with the usual action and in this case $N(t)$ is easily seen to be asymptotic to πt^2 (and the subtler problem is to obtain the best possible estimates on the error term $N(t) - \pi t^2$).

In this paper we concentrate on the case where X is negatively curved. If X has all its sectional curvatures pinched between two negative constants and Γ is co-compact, i.e., the quotient $M = X/\Gamma$ is a compact manifold, then Margulis showed that $N(t) \sim Ce^{\nu t}$, as $t \to +\infty$, for some positive constant C and where $\nu > 0$ is the exponential growth rate of the volume of balls in X [4]. (We are using the notation $F(t) \sim G(t)$ to denote that $\lim_{t\to+\infty} F(t)/G(t) = 1$.)

As our main result we restrict to the case where X is a surface but consider more general groups of isometries Γ. We have the following theorem.

Theorem 1. *Let X be a surface with its curvature bounded between two negative constants. Let Γ be a convex co-compact group of isometries of X. Then*

$$N(t) \sim Ce^{ht}, \text{ as } t \to +\infty,$$

where C is a positive constant and $h > 0$ is the topological entropy of the geodesic flow over $M = X/\Gamma$.

Here we are using the term "convex co-compact" in a more general sense than that familiar for groups of isometries of hyperbolic space $\mathbb{H}^{n+1}$, $n \geq 1$ (where all the sectional curvatures are identically 1). Recall that the orbit of any point $p \in \mathbb{H}^{n+1}$ under Γ accumulates on S^n, regarded as the sphere at infinity for $\mathbb{H}^{n+1}$, and the set of accumulation points Λ is called the limit set of Γ. Let $C(\Lambda)$ denote the intersection of the (hyperbolic) convex hull of Λ with $\mathbb{H}^{n+1}$, then Γ is said to be convex co-compact if the quotient $C(\Lambda)/\Gamma$ is compact. In the more general setting of variable negative curvature the natural extension of this notion is to require that the geodesic flow on the unit-tangent bundle SM of $M = X/\Gamma$ has a compact non-wandering set Ω. In the case of surfaces that we consider here, this is equivalent to requiring that M consists of a compact core with a finite number of infinite volume ends homeomorphic to $\mathbb{R} \times S^1$.

To put our result into context, Patterson has shown that in the case where $X = \mathbb{H}^{n+1}$, $N(t) \sim Ce^{\delta t}$, as $t \to +\infty$, for some constant $C > 0$ and where $\delta > 0$ is the Hausdorff dimension of Λ [6]. Earlier, Lax and Phillips had obtained a more refined result with rather precise error terms under the restriction that $\delta > n/2$ [3]. Subsequently, Lalley [1] recovered Patterson's result in the case $n = 1$ using methods based on symbolic dynamics [10] and "thermodynamic" ergodic theory [9].

Since the case where M is compact is covered by the result of Margulis, we shall suppose that M has at least one infinite volume end. This introduces the desirable (but not essential) simplication that then Γ is a free group. It is worthwhile to note that if the base point $x \in X$ is chosen so that it lies in the lift to X of the projection of the non-wandering set Ω from SM to M, then Theorem 1 follows from the proof of Margulis, as remarked by Yue [11]. Here, however, we impose no such restriction.

Our method of proof is closely akin to that employed by Lalley in the constant curvature case and is based on coding Γ (as an abstact group) in terms of a subshift of finite type. The relevant geometric information is recovered in terms of a certain real valued function and a crucial step is to establish that it is Hölder continuous; this follows from [8]. We are then in a position deduce Theorem 1 from an analysis of the associated Poincaré series via a class of transfer operators.

1. Subshifts of finite type

In this section we shall introduce a subshift of finite type $\sigma : \Sigma \to \Sigma$ associated to Γ regarded as an abstract group and show how to encode some of the original geometric information in terms of a Hölder continuous function $r : \Sigma \to \mathbb{R}$.

If M has genus g and has $p+1$ infinite volume ends then $\Gamma \cong \pi_1 M$ is free on $k = 2g+p$ generators, $a_1, \ldots, a_k$, say. Let $S = \{a_1, \ldots, a_k\} \cup \{a_1^{-1}, \ldots, a_k^{-1}\}$. Define a $0-1$ square matrix A with rows and columns indexed by $S \cup \{0\}$ by

(1) $A(x, y) = 1$, $x, y \in S$, $x \neq y^{-1}$;
(2) $A(x, x^{-1}) = 0$, $x \in S$;
(3) $A(x, 0) = 1$, $x \in S \cup \{0\}$; and
(4) $A(0, x) = 0$, $x \in S$.

Write $\Sigma = \{x = (x_n) \in \prod_{n \geq 0} S \cup \{0\} : A(x_n, x_{n+1}) = 1 \ \forall n \geq 0\}$ and define the subshift of finite type $\sigma : \Sigma \to \Sigma$ by $(\sigma x)_n = x_{n+1}$.

Define a metric ρ on Σ by

$$\rho(x, y) = \sum_{n=0}^{+\infty} \frac{1 - \delta_{x_n, y_n}}{2^n},$$

where $\delta_{i,j}$ is the usual Kroenecker symbol. With respect to this metric, Σ is compact and $\sigma : \Sigma \to \Sigma$ is a continuous transformation.

Observe that we may write Σ as the disjoint union of two σ-invariant subsets $\Sigma = \Sigma_0 \cup \Sigma_1$, where Σ_0 denotes the set of those sequences which end in an infinite string of zeros and Σ_1 denotes those sequences which do not contain a zero. Clearly, Σ_0 is a dense subset of Σ and there is an obvious one-to-one correspondence between Σ_0 and Γ.

Notice that Σ_1 is the shift space associated to the submatrix B obtained fron A by deleting the row and column indexed by 0. It is easy to check that B is aperiodic, i.e., that there exists $N > 0$ such that B^N has all its entries positive. It is well known that this is equivalent to $\sigma : \Sigma_1 \to \Sigma_1$ being topologically mixing, i.e., that for every pair of non-empty open sets $U, V \subset \Sigma_1$, there exists $N > 0$ such that $\sigma^n(U) \cap V \neq \emptyset$ for all $n \geq N$.

We now wish to translate some of the original geometric information into this new setting. Define a function $r : \Sigma_0 \to \mathbb{R}$ by $r(0, 0, \dots) = 0$ and

$$r(g_0, g_1, \dots, g_{n-1}, 0, 0, \dots) = d(p, g_0 g_1 \dots g_{n-1} p) - d(p, g_1 \dots g_{n-1} p).$$

If we write $r^n = r + r \circ \sigma + \dots r \circ \sigma^{n-1}$, then we have that

$$r^n(g_0, g_1, \dots, g_{n-1}, 0, 0, \dots) = d(p, g_0 g_1 \dots g_{n-1} p). \tag{1.1}$$

The following proposition will be crucial to our subsequent analysis.

Proposition 1. *The function $r : \Sigma_0 \to \mathbb{R}$ is Hölder continuous and extends to a Hölder continuous function $r \in C^\alpha(\Sigma, \mathbb{R})$, for some exponent $\alpha > 0$.*

Proof. This is Proposition 3 of [8].

2. Periodic orbits and closed geodesics

Before we begin our analysis of Poincaré series, it will prove convenient to understand the connection between the subshift of finite type $\sigma : \Sigma \to \Sigma$, the function $r : \Sigma \to \mathbb{R}$ and closed geodesics on M.

First, observe that there is a natural correspondence between periodic orbits for $\sigma : \Sigma_1 \to \Sigma_1$ and *cyclically* reduced words in the generators and their inverses, i.e., expressions of the form $g_0 g_1 \dots g_{n-1}$, with each $g_i \in S$ and $g_{i+1} \neq g_i^{-1}$ (mod n). In a free group there is a unique cyclically reduced word in every conjugacy class, so we have a natural bijection between periodic orbits of $\sigma : \Sigma_1 \to \Sigma_1$ and (non-trivial) conjugacy classes in Γ. Furthermore, since M has strictly negative curvature, there is a unique closed geodesic on M in each such conjugacy class. We have now proved the first part of the following result, which relates the lengths of closed geodesics to the function $r : \Sigma \to \mathbb{R}$.

Proposition 2. *There is a bijection between periodic orbits $x = (g_0, g_1, \dots, g_{n-1}, g_0, g_1, \dots, g_{n-1}, \dots)$ for the shift $\sigma : \Sigma_1 \to \Sigma_1$ and closed geodesics γ on M, where γ is the unique closed geodesic in the conjugacy class containing $g_0 g_1 \dots g_{n-1}$. Furthermore, $r^n(x) = l(\gamma)$, the length of γ.*

Proof. We only need to prove the final statement. Write $g = g_0 g_1 \dots g_{n-1}$ and also, by an abuse of notation, let g^m, $m \geq 1$, denote the sequence obtained by concatenating $g_0, g_1, \dots, g_{n-1}$ m times and write $x = (g^\infty)$.

Since r is Hölder continuous, we have the following estimate that

$$|r^{mn}(g^\infty) - r^{mn}(g^m, 0, 0, \dots)| \le \sum_{k=1}^{mn} |r|_\alpha 2^{-k\alpha} = |r|_\alpha \frac{1 - 2^{-(mn+1)\alpha}}{1 - 2^{-\alpha}} \le \frac{|r|_\alpha}{1 - 2^{-\alpha}}.$$

Here, $|r|_\alpha$ denotes the best choice of Hölder constant for $r \in C^\alpha(\Sigma, \mathbb{R})$. Noting that $r^{mn}(g^\infty) = mr^n(g^\infty)$, the above estimate gives us that

$$r^n(g^\infty) = \lim_{m \to +\infty} \frac{1}{m} r^{mn}(g^m, 0, 0, \dots) = \lim_{m \to +\infty} \frac{1}{m} d(p, g^m p).$$

Thus it remains to show that this last quantity is equal to $l(\gamma)$. We shall do this by proving inequalities in both directions.

Choose a point $q \in X$ such that $\pi(q) \in \gamma$, where $\pi : X \to M$ is the projection. It is easy to see that $\lim_{m \to +\infty} \frac{1}{m} d(q, g^m q)$ only depends on the conjugacy class containing g; we shall denote this limit by $D(g)$. An easy application of the triangle inequality gives that $D(g) \le d(q, hq)$, where h is any group element which is conjugate to g. However, $l(\gamma)$ is the infimum of $d(q, hq)$ over all such elements h, so we have that $D(g) \le l(\gamma)$.

Now observe that

$$d(p, g^m p) \le d(p, q) + d(q, g^m q) + d(g^m p, g^m q) = 2d(p, q) + d(q, g^m q).$$

Thus we obtain

$$\lim_{m \to +\infty} \frac{1}{m} d(p, g^m p) \le D(g) \le l(\gamma)$$

giving the inquality in one direction.

On the other hand, the geodesic arc from $\pi(p)$ to $\pi(p)$ on M lying in the homotopy class g^m is, in particular, a closed curve which is freely homotopic to γ^m. As such its length, $d(p, g^m p)$ is at least $l(\gamma^m) = ml(\gamma)$. Thus $\lim_{m \to +\infty} \frac{1}{m} d(p, g^m p) \ge l(\gamma)$, giving the other inequality, as required.

In particular, Proposition 2 implies that $r : \Sigma_1 \to \mathbb{R}$ is cohomologous to a positive function, i.e., there exists $u \in C^0(\Sigma_1, \mathbb{R})$ such that $r + u \circ \sigma - u$ is positive.

3. Poincaré series and transfer operators

Our approach to the asymptotic formula in Theorem 1 will be via a certain function of a complex variable called the Poincaré series. We shall obtain information about the analytic domain of this function and then use a Tauberian theorem to translate this into information about the behaviour of $N(t)$.

We define the Poincaré series $\eta(s)$ to be the infinite series

$$\eta(s) = \sum_{g \in \Gamma - \{e\}} e^{-sd(p, gp)}.$$

This expression converges absolutely provided that $Re(s)$ is sufficiently large and so defines an analytic function in a half-plane. In fact, the abscissa of convergence is $s = h$, for some $h > 0$. We aim to show that $\eta(s)$ has an extension as an analytic function to a neighbourhood of the line $Re(s) = h$, apart from a simple pole at $s = h$. By the main result of [8], $\eta(s)$ has an extension as a meromorphic function to a half-plane $Re(s) > h - \delta$, for some $\delta > 0$. Thus we need only show that $\eta(s)$ has no poles on the line $Re(s) = h$, apart from a simple pole at $s = h$.

Observe that $\eta(s)$ has an expression in terms of the function $r : \Sigma \to \mathbb{R}$. For $Re(s) > h$ we may write, using (1.1),

$$\eta(s) = \sum_{n=1}^{+\infty} \sum_{y \in \sigma^{-n}\dot{0}} e^{-sr^n(y)},$$

where we have written $\dot{0} = (0, 0, \dots)$.

In order to obtain results about the analytic domain of $\eta(s)$, we shall make use of the well studied class of Ruelle transfer operators, which we now define. For $f \in C^\alpha(\Sigma, \mathbb{C})$, define the transfer operator $L_f : C^\alpha(\Sigma, \mathbb{C}) \to C^\alpha(\Sigma, \mathbb{C})$ by

$$L_f g(x) = \sum_{\substack{\sigma y = x \\ y \neq \dot{0}}} g(y) e^{f(y)}.$$

Remark. Note that our definition of the transfer operator is slightly unusual in that we require $y \neq \dot{0}$, however it agrees with the usual definition for all $x \neq \dot{0}$. This reason we do this is so that $L_f : C^\alpha(\Sigma, \mathbb{C}) \to C^\alpha(\Sigma, \mathbb{C})$ and the resriction $L_f : C^\alpha(\Sigma_1, \mathbb{C}) \to C^\alpha(\Sigma_1, \mathbb{C})$ have the same spectrum.

In our present setting, we shall be interested in the special choice of $f = -sr$.

These operators are usually studied in the case of mixing subshifts, where their spectral properties are well understood, and so the relevant results only apply directly to $L_f : C^\alpha(\Sigma_1, \mathbb{C}) \to C^\alpha(\Sigma_1, \mathbb{C})$. We summarize what we require in the following propostion; proofs may be found in, for example, [5].

Proposition 3.

(1) *If $f \in C^\alpha(\Sigma_1, \mathbb{R})$ the L_f has a simple positive eigenvalue $e^{P(f)}$ with an associated positive eigenfunction such that the rest of the spectrum of L_f is contained in a disc of radius strictly less than $e^{P(f)}$. If $f \in C^\alpha(\Sigma_1, \mathbb{C})$ then the spectral radius of L_f is less than or equal to $e^{P(Re(f))}$. The spectral radius of L_f is strictly less than $e^{P(Re(f))}$ unless it is possible to write $Im(f) = u \circ \sigma - \sigma + \psi + a$, where $u \in C^0(\Sigma_1, \mathbb{R})$, $\psi \in C^0(\Sigma_1, 2\pi\mathbb{Z})$ and $a \in \mathbb{R}$. If such an identity does hold then L_f has a simple eigenvalue $e^{P(Re(f))+ia}$ and the rest of the spectrum of L_f is contained in a disc of radius strictly less than $e^{P(Re(f))}$.*

(2) *If $f \in C^\alpha(\Sigma_1, \mathbb{R})$ is cohomologous to a positive function then, for $t \in \mathbb{R}$, the function $t \to P(-tf)$ is real analytic and strictly decreasing. Furthermore, $P(-tf) \to -\infty$ $(+\infty)$, as $t \to +\infty$ $(-\infty)$, so, in particular, $P(-tr)$ has a unique zero.*

As remarked above, these results do not apply directly to $L_{-sr} : C^\alpha(\Sigma, \mathbb{C}) \to C^\alpha(\Sigma, \mathbb{R})$. However, as remarked above, the two spectra in fact coincide (cf. the discussion in Section 6 of [1]).

We are now in a position to write the Poincaré series in terms of iterates of L_{-sr}. By direct calculation, we have that

$$\eta(s) = \sum_{n=1}^{\infty} (L_{-sr}^n 1)(\dot{0}). \tag{3.1}$$

We know that r is cohomologous to a positive function. The abscissa of convergence h is the unique real number such that $P(-hr) = 0$. This formula characterizes the topological entropy of the suspended flow over Σ_1 with roof function cohomologous to r. In view of Lemma 2, this is equal to the topological entropy of the geodesic flow on the unit-tangent bundle over M. We aim to show that $\eta(s)$ has a pole at $s = h + it$ if and only if $t = 0$. First observe that if the spectral radius of $L_{-(h+it)r}$ is strictly less than 1 then the series (3.1) converges absolutely. If, on the other hand, the spectral radius of $L_{-(h+it)r}$ is equal to 1 then the restriction of r to Σ_1 satisfies $-tr = u \circ \sigma - u + \psi + a$, for some $u \in C^0(\Sigma_1, \mathbb{R})$, $\psi \in C^0(\Sigma_1, 2\pi\mathbb{Z})$ and $a \in \mathbb{R}$ and $L_{-(h+it)r}$ has e^{ia} as a simple isolated eigenvalue. By standard analytic eigenvalue perturbation theory for linear operators, this eigenvalue persists as a eigenvalue $e^{P(-sr)}$ for L_{-sr}, strictly maximal in modulus, provided that s is in a sufficiently small neighbourhood U of $h+it$. Thus, for $s \in U$ with $Re(s) > h$, we may write

$$(L_{-sr}^n 1)(\dot{0}) = e^{nP(-sr)}(\mathbb{P}_s 1)(\dot{0}) + (T_n(s)1)(\dot{0})$$

where $\mathbb{P}_s$ is the eigenprojection associated to the simple eigenvalue $e^{P(-sr)}$ and where $\limsup_{n\to+\infty} ||T_n(s)||^{\frac{1}{n}} < 1$.

Thus $\eta(s)$ has an extension to U as a meromorphic function given by

$$\eta(s) = \frac{(\mathbb{P}_s 1)(\dot{0})}{1 - e^{P(-sr)}} + A(s)$$

where $A(s) = \sum_{n=1}^{+\infty} (T_n(s)1)(\dot{0})$ is analytic. Observe that $s = h + it$ is a pole if and only if $e^{P(-(h+it)r)} = 1$, i.e., if and only if

$$-tr = u \circ \sigma - u + \psi \tag{3.2}$$

with $u \in C^0(\Sigma_1, \mathbb{R})$ and $\psi \in C^0(\Sigma_1, 2\pi\mathbb{Z})$. If (3.2) holds then we have that $-tr^n(x) \in 2\pi\mathbb{Z}$ whenever $\sigma^n x = x$. In view of Lemma 2, if $t \neq 0$, this implies that the set $\{l(\gamma) : \gamma \text{ is a closed geodesic on } M\}$ is contained in a discrete subgroup of $\mathbb{R}$. However, this is in contradiction to the well-known fact that the geodesic flow over M is weak-mixing [11]. Thus we conclude that $t = 0$, as required.

Finally, we need to show that $s = h$ is a simple pole for $\eta(s)$. The residue $s = h$ is given by

$$\mathbb{P}_h(1) \lim_{s\to h} \frac{s-h}{1 - e^{P(-sr)}} = \frac{-\mathbb{P}_h(1)}{P'(-hr)}$$

and since $P(-tr)$ is strictly decreasing, this is positive.

We are now ready to complete the proof of Theorem 1. Note that for $Re(s) > h$, we may write $\eta(s)$ as a Stieltjes integral with respect to $N(t)$:

$$\eta(s) = \int_0^{+\infty} e^{-st} dN(t).$$

Since we have that $\eta(s)$ has an extension as an analytic function to a neighbourhood of the line $Re(s) = h$, apart from a simple pole at $s = h$, the Ikehara Tauberian theorem [2] tells us that $N(t) \sim Ce^{ht}$, as $t \to +\infty$, where $C > 0$ is the residue of $\eta(s)$ at $s = h$.

4. Some generalizations

In this final section we discuss some generalizations of Theorem 1. First we give a higher dimensional version of the result. Observe that the only point at which we use the fact that M is a surface is to ensure that $\Gamma \cong \pi_1 M$ is a free group. Using the coding ideas of Series [10] the method of proof also works, although the correspondence between periodic orbits for the subshift and conjugacy classes is less obvious, if M is a compact surface of genus $g \geq 2$ in which case we have

$$\pi_1 M \cong < a_1, \dots, a_g, b_1, \dots, b_g : a_1 b_1 a_1^{-1} b_1^{-1} \dots a_g b_g a_g^{-1} b_g^{-1} = 1 > .$$

We immediately obtain the following result.

Theorem 2. *Let X be a complete simply connected Riemannian manifold with all its sectional curvatures bounded between two negative constants. Let Γ be a convex co-compact group of isometries of X such that Γ is isomorphic either to a free group or to the fundamental group of a compact surface of genus at least 2. Then, for some $C > 0$,*

$$N(t) \sim Ce^{ht}, \text{ as } t \to +\infty,$$

where $h > 0$ is the topological entropy of the geodesic flow over X/Γ.

As a second generalization, we consider normal subgroups $\Gamma_0 \lhd \Gamma$ such that $\Gamma/\Gamma_0 \cong \mathbb{Z}^\nu$, $\nu \geq 1$, and the associated orbit counting function $N_0(t) = Card(B(p,t) \cap \Gamma_0 p)$. In [7] we considered this problem in the case where Γ is co-compact. The arguments employed there may be applied in this case to deduce that

$$N_0(t) \sim C' \frac{e^{ht}}{t^{\nu/2}}, \text{ as } t \to +\infty,$$

for some $C' > 0$, but we shall not give more details here.

References

1. S. Lalley, *Renewal theorems in symbolic dynamics with applications to geodesic flows, noneuclidean tesselations and their fractal limits*, Acta Math. **163** (1990), 1-55.
2. S. Lang, *Algebraic Number Theory*, Addison-Wesley, New York, 1970.
3. P. Lax and R. Phillips, *The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces*, J. Funct. Anal. **46** (1982), 280-350.

4. G. Margulis, *On some applications of ergodic theory to the study of manifolds of negative curvature*, Funct. Anal. Appl. **3** (1969), 89-90.
5. W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque **187-188** (1990), 1-268.
6. S. Patterson, *On a lattice point problem in hyperbolic space and related questions in spectral theory*, Arkiv Mat. **26** (1988), 167-172.
7. M. Pollicott and R. Sharp, *Orbit counting for some discrete groups acting on simply connected manifolds with negative curvature*, Invent. math. **117** (1994), 275-302.
8. M. Pollicott and R. Sharp, *Poincaré series and comparison theorems for variable negative curvature*, Preprint (1994).
9. D. Ruelle, *Thermodynamic Formalism*, Addison-Wesley, New York, 1978.
10. C. Series, *Symbolic dynamics for geodesic flows*, Acta Math. **146** (1981), 103-128.
11. C. B. Yue, *The ergodic theory of discrete isometry groups on manifolds of variable negative curvature*, M.S.R.I. Preprint (1992).

Mark Pollicott, Mathematics Institute, University of Warwick, Coventry CV4 7AL, U.K.

Richard Sharp, Mathematical Institute, 24-29 St. Giles, Oxford OX1 3LB, U.K.