RATE OF APPROXIMATION OF MINIMIZING MEASURES

XAVIER BRESSAUD AND ANTHONY QUAS

ABSTRACT. For T a continuous map from a compact metric space to itself and f a continuous function, we study the minimum of the integral of f with respect to the members of the family of invariant measures for T and in particular the rate at which this minimum is approached when the minimum is restricted to the family of invariant measures supported on periodic orbits of period at most N. We answer a question of Yuan and Hunt by demonstrating that the error of approximation decays faster than N^{-k} for all $k > 0$, and show that this is sharp by giving examples for which the approximation error does not decay faster than this.

1. INTRODUCTION

Let $T: X \to X$ be a continuous map from a compact metric space to itself and let $F: X \to \mathbb{R}$ be a continuous function of X. We let $M_{\text{inv}}(T)$ denote the collection of Borel probability measures that are invariant under T. It is well known that this set is non-empty and compact in the weak*-topology. An invariant measure μ will be called a *minimizing measure* for f if $\int f\,d\mu \le \int f\,d\nu$ for all $\nu \in M_{\text{inv}}(T)$.

For the purposes of this paper, we work in the case where X is a one-sided full shift and T is the shift map. For a point x in X, we use the notation x_s^t to denote the word $x_s x_{s+1} \ldots x_t$. We equip X with its standard metric: two points x and y are at a distance 2^{-n} if they first disagree in the nth position. The functions f that we consider will be Hölder continuous or Lipschitz. Various versions of the following conjecture and related questions appeared in a number of papers [7, 8, 10, 13]. The conjecture itself remains open.

Conjecture 1. *Let (X, T) be the one-sided full shift on a finite alphabet and let F denote the collection of Hölder continuous functions with a fixed exponent α or the collection of Lipschitz functions. Then the set*

Date: January 16, 2007.

The second author wishes to thank the Institut de Mathématiques de Luminy, where this research was carried out, for its hospitality. Research partially supported by NSERC.

of f in F for which there is a minimizing measure that is supported on a periodic orbit is a residual subset of F (i.e. it contains the intersection of a countable collection of dense open sets).

Bousch [2] established a version of the conjecture in which the functions are taken to satisfy the "Walters condition" and Bousch and Jenkinson [3] demonstrated that it failed in the case where F is taken to be the class of continuous functions. We remark that in the literature, it has been traditional to work with maximizing measures rather than minimizing measures. However the normal form given below due to Conze and Guivarc'h [9] and subsequently discovered independently by Bousch [1] and Contreras, Lopes and Thieullen [8] means that it is more convenient to work with minimizing measures.

It has been shown in considerable generality (i.e. for many transformations and classes of functions) by Jenkinson [11] that the set of f for which there is a unique minimizing measure is residual.

We note that although it is conjectured that for a residual set of f, the measure minimizing the integral is periodic, it is straightforward to give examples of functions f whose minimizing measure is not supported on a periodic orbit. To see this, let Y be any uniquely ergodic subshift of the full shift (e.g. the Morse system). Defining $f(x) = d(x, Y)$ gives a function whose unique minimizing measure is the invariant measure supported on Y. Further, by the Jewett-Krieger theorem, for any ergodic measure-preserving transformation (X, S) of finite entropy, there is a uniquely ergodic subshift (Y, T) that is measure-theoretically isomorphic to (X, S), so that there are many f for which the minimizing measure is not supported on a periodic orbit.

Although we do not resolve the conjecture above in this paper, it remains of interest to quantify the rate at which $\int f\,d\nu$ approaches $\int f\,d\mu$ (where μ is the minimizing measure) as ν ranges over periodic points of increasing periods. This question was raised explicitly by Yuan and Hunt [13]. More specifically, for a periodic point p, we denote by $P(p)$ its period and by μ_p the unique invariant measure supported on the orbit of p. We further denote $\mathrm{Per}_N = \{\mu_p \colon P(p) \leq N\}$.

For a given function f, let μ be the minimizing measure. We define $\Delta(N)$ as follows:

$$\Delta(N) = \min_{\nu \in \mathrm{Per}_N} \left(\int f\,d\nu - \int f\,d\mu \right).$$

Yuan and Hunt noted that in many examples that have been studied, $\Delta(N)$ decays exponentially in N. They obtained a bound of $\Delta(N) = O(N^{-k})$ for a fixed k determined by the dimension of the ambient space.

Collier and Morris [6] produce examples such that $\Delta(N)$ decays slower than exponentially (specifically they achieve $\Delta(N) = \Omega(\exp(-N^\epsilon)))$.

Our main result is the following

Theorem 2. *Let (X, T) denote the one-sided full shift on a finite alphabet. Let F denote the class of Hölder continuous functions with a fixed exponent α or the class of Lipschitz functions. Then for every $f \in F$, we have for each $k > 0$, $\Delta(N) = o(N^{-k})$.*

This bound is sharp in the following sense. Suppose that we are given an $h \geq 0$, an $M > e^h$, and a sequence (a_N) such that for each $k > 0$, $a_N = o(N^{-k})$. Then there exists a uniquely ergodic subshift Y of the full shift X on M symbols such that

- *defining $f(x) = d(x, Y)$, there is a subsequence N_i along which $\Delta(N_i) > a_{N_i}$.*
- *The entropy of Y is h.*

Corollary 3. *The conclusions of Theorem 2 hold also for the class of C^2 expanding maps and Axiom A diffeomorphisms restricted to hyperbolic sets.*

Let (X, T) be a one-sided full shift and let μ be an invariant measure. For a measure ν, let $c(\nu, \mu) = \max_{x \in \text{supp}(\nu)} \min_{y \in \text{supp}(\mu)} d(x, y)$. Notice that the Hausdorff distance between $\text{supp}(\mu)$ and $\text{supp}(\nu)$ is given by $\max(c(\nu, \mu), c(\mu, \nu))$. We define

$$\Delta'_\mu(N) = \min_{\nu \in \text{Per}_N} c(\nu, \mu).$$

The proof of our main theorem (Theorem 2) above also gives the following result

Theorem 4. *Let (X, T) be a one-sided full shift and let μ be any invariant measure. Then for all k, $\Delta'_\mu(N) = o(N^{-k})$.*

This is sharp in the following sense. Fix any sequence (a_N) such that for each k, $a_N = o(N^{-k})$. Then there exists an invariant measure μ such that for infinitely many N, $\Delta'_\mu(N) > a_N$.

The above shows that for any invariant measure μ and for any k, we can find periodic orbits of period less than N, all of whose points lie within N^{-k} of the support of μ. One might ask whether there is a periodic orbit such that the Hausdorff distance of its support from the support of μ is less than N^{-k}. One quickly sees that this is hopeless by considering the case where μ has positive entropy. If the entropy of μ is h, then the number of cylinders of length n intersecting the support is at least e^{hn}. In particular, for N fixed, let $n = 1 + \lfloor (\log N)/h \rfloor$. Since the number of cylinders of length n exceeds N, any periodic orbit of period N or less must fail to enter one of the n-cylinders. In particular, for any $\nu \in \mathrm{Per}_N$, we have $c(\mu,\nu) \geq 2^{-n} \geq CN^{-\log 2/h}$.

2. Proof of the Upper Bound

We remark in the proof of the upper bound of Theorem 2, that although the conclusion does not depend on the entropy of the minimizing measure, the proofs in the zero and positive entropy cases are complementary. Specifically one is trying to construct a periodic orbit in the ambient shift space with small period such that the orbit remains close to the support of the minimizing measure. We approximate the points staying close to the orbit by a shift of finite type whose states correspond to words of length n that intersect the support of the measure.

In the case where the entropy is 0, the approximation is possible essentially by the pigeonhole principle: there are a subexponential number of words of length n and hence there must be a recurrence in a subexponential time. This allows one to construct an approximating orbit staying within a distance 2^{-n} of the support with period at most $2^{n/k}$ giving $\Delta(N) = o(N^{-k})$ for any k.

On the other hand, in the case where the entropy bound is positive, a pigeonhole argument would only establish a bound of the form $\Delta(N) = o(N^{-1/h})$. This was essentially the manner in which the bound was given in Yuan and Hunt [13]. In the positive case, we make use of the entropy to ensure that there are many possible transitions in the shift of finite type and use these to guarantee the existence of a short cycle. It turns out that the zero entropy argument follows from the proof that we give in the positive entropy case, but we include the zero entropy argument for its simplicity.

In proving an upper bound, we will need the following lemma to prove the existence of periodic points with small period in certain shifts of finite type.

Lemma 5. *Suppose that Y is a shift of finite type (with forbidden words of length 2) with M symbols and entropy h. Then Y contains a periodic point of period at most $1 + Me^{1-h}$.*

Proof. Let $k+1$ be the period of the shortest periodic orbit in Y. Then we claim that a word of length k in the shift of finite type is determined by the set of symbols that it contains. First note that if there since there is no periodic orbit of length k or less, any allowed k-word must contain k distinct symbols. Next, suppose that u and v are two distinct words of length k containing the same symbols. Then since the words are distinct, there exists a consecutive pair of symbols, a and b say, in v that occur in the opposite order (not necessarily consecutively) in u. Then the infinite concatenation of the segment of u starting at b and ending at a gives a word in Y of period at most k.

It then follows that there are at most $\binom{M}{k}$ words of length k. We therefore have using basic properties of topological entropy,

$$\begin{aligned} e^{hk} &\leq \binom{M}{k} \leq \frac{M^k}{k!} \\ &\leq (Me/k)^k. \end{aligned}$$

Taking kth roots, we see that $k \leq Me^{1-h}$ as required. □

Remarks.

(1) In graph-theoretic language, this lemma states that for a directed graph on M vertices, the girth g satisfies $g \leq 1+Me/\lambda_1$, where λ_1 is the leading eigenvalue of the adjacency matrix.

We note the similarity with the Caccetta-Häggkvist conjecture [4], which states that in a directed graph with minimum out-degree δ_+, $g \leq \lceil M/\delta_+\rceil$. This remains open although there are a number of results [5, 12] giving bounds of the form $g \leq M/\delta_+ + C$ for a universal constant C.

(2) A class of simple examples demonstrates that the bound given in Lemma 5 for the shortest period is close to optimal in some ranges. Specifically, given M and h_0 with $\frac{1}{2}\log M \leq h_0 \leq (1-\beta)\log M$, there exists a shift of finite type Y with M symbols and entropy h satisfying $|h-h_0| \leq M^{-\beta}$ containing no periodic points of period less than $Me^{1-h} - C\log M$.

The examples consist of shifts of finite type of the following form. The alphabet is $\{1,\ldots,n\}$ and for $i<k$, the vertex i is connected only to $i+1$; the vertex k is connected to all vertices j with $j>k$; and for $i>k$, i is connected to all vertices j with $j>i$ as well as to 1.

We conjecture that the above examples are extremal in the sense that for given alphabet size M and minimum period k, these examples maximize the entropy.

We would like to thank Julien Cassaigne for bringing these examples to our attention.

Proof of Theorem 2 (Upper Bound). We first establish the upper bounds. We deal separately with the cases of positive entropy and zero entropy. We assume that the function f is Hölder continuous with Hölder exponent α so that $|f(x)-f(y)| = O(d(x,y)^\alpha)$.

We note that if g is cohomologous by a continuous transfer function to $f-C$ for some constant C (that is $g = f - C + r - r\circ T$ for a continuous r), then g and f share the same minimizing measures and further $\Delta_f(N) = \Delta_g(N)$ for all N. The normal form result [9, 8, 1] mentioned earlier states that there exists a C and a function g that is cohomologous to $f-C$ by a Hölder continuous transfer function with the same exponent α such that g is non-negative, and the set on which g takes the value 0 contains a non-empty subshift. Notice that for any measure $\nu \in M_{\mathrm{inv}}(T)$, $\int f\,d\nu - \int f\,d\mu = \int g\,d\nu - \int g\,d\mu$. We let $Y = \{x\colon g(x) = 0\}$ and observe that the set of minimizing measures for g (and hence f) is precisely the set of invariant measures supported on Y.

We assume first that Y has topological entropy 0. It follows that given $k > 0$, there exists an n_0 such that for all $n \geq n_0$, Y has less than $\lfloor 2^{\alpha n/k}\rfloor$ legal words of length n. Now given any $N \geq 2^{\alpha n_0/k}$, let $n = \lfloor (k/\alpha)\log_2 N\rfloor$ so that Y has less than N words of length n. Now given any point y in Y, it follows using the pigeonhole principle that there exist $0 \leq s < t \leq N$ such that $y_s^{s+n-1} = y_t^{t+n-1}$. Letting p be the point consisting of infinite concatenations of y_s^{t-1}, we see that p has period at most N. Since each subword of p of length $n+1$ agrees with a subword of the same size in y and since $g(T^j y) = 0$ for all j, it follows that for each $i \geq 0$, $g(T^i p) \leq C2^{-\alpha n}$ so that $\int g\,d\mu_p \leq C2^{-n\alpha} = O(N^{-k})$. This shows that $\Delta(N) = O(N^{-k})$ as required.

In the case where the subshift Y has positive entropy h, we argue as follows. Let $k > 0$ be given. For $n > 0$, we define a 1-step shift of finite type $Z^{(n)}$, where the symbols are legal words of length n in Y (i.e. the words defining cylinder sets of length n that intersect Y). The transition from u to v in $Z^{(n)}$ is allowed if the word uv of length $2n$ is a legal word in Y. (Note that this is not the standard transition matrix for the n-block recoding of a shift of finite type). Since the subshift of finite type defined this way is a supersystem of T^n acting on Y, the entropy of the subshift exceeds nh. Since the topological entropy of Y is h, the number of symbols in $Z^{(n)}$ is given by $K_n e^{nh}$ where K_n grows at a subexponential rate. From Lemma 5, the shortest periodic orbit in $Z^{(n)}$ is of period at most $1 + K_n e^{nh} e^{1-hn} = 1 + eK_n$. Since each symbol in $Z^{(n)}$ corresponds to a word of length n in the original shift space, such an orbit corresponds to a periodic orbit in the original shift space of period at most $n(1 + eK_n)$. Denote this function by $H(n)$. Since $H(n)$ is subexponential, for all k, it is the case that $H(\lceil k\log_2 N/\alpha\rceil) \leq N$ for all large N.

Now given a large N, let $n = \lceil k \log_2 N/\alpha \rceil$. The above establishes that $Z^{(n)}$ contains a periodic orbit corresponding to a periodic orbit p in X of period at most N. We claim that any n-block in p is a legal n-block in Y. To see this, note that if the n-block is one of the symbols in $Z^{(n)}$, then this follows from the definition of the symbols, whereas if the n-block is a sub-block of the concatenation of two symbols in $Z^{(n)}$, the transition rules in $Z^{(n)}$ force the n-block to be legal in Y. It follows that $d(T^i p, Y) \leq 2^{-n}$ for each i, so that using the Hölder property of g, $\Delta(N) \leq \int g \, d\mu_p \leq 2^{-\alpha n}$. By the choice of n, this is bounded above by N^{-k} as required. □

3. Proof of the Lower Bound

The following idea and lemma is the basis for the lower bound in Theorem 2.

Let W be a collection of words, all of the same length ℓ. We denote by $\overline{W}$ the collection of all (one-sided) infinite concatenations of words in W and by X_W the subshift $\bigcup_{i=0}^{\ell-1} T^i \overline{W}$. A finite word v_0^{s-1} is said to be legal in X_W if there is an element x of X_W such that $x_0^{s-1} = v_0^{s-1}$.

A number $d > 0$ is called a *decoding length* for W if given any finite word y_0^{d-1} that is legal in X_W, there is a unique $t < \ell$ such that there exists a $z \in \overline{W}$ (not unique) such that $y_0^{d-1} = z_t^{t+d-1}$. It follows that if W has decoding length d, then knowing a section of a word x in X_W of length d is sufficient to determine the positions of the boundaries of the atomic words (those in W) in the infinite word x. Note that not every collection of words has a decoding length (for example $W = \{01, 10\}$ has no decoding length. In fact, a compactness argument shows that W has a decoding length if and only if $T^i \overline{W} \cap T^j \overline{W} = \emptyset$ for $0 \leq i < j < \ell$.

If $d > 0$ is a decoding length for W, then given an infinite word y and any subword y_s^{s+d-1} that is legal in X_W, it follows there is a unique $t < \ell$ such that there is a $z \in \overline{W}$ such that $y_s^{s+d-1} = z_{s+t}^{s+t+d-1}$. In this case, t is called the *offset* of the word y_s^{s+d-1}. (Note that t depends not only on the block but also on its position in y.)

Lemma 6. *Let W be a collection of words of length ℓ with decoding length d and let y be a periodic point with period $m < \ell$. Then there exists a subword $y_t \dots y_{t+d}$ of length $d+1$ that is not legal in X_W.*

Proof. Assume for a contradiction that every subword of y of length $d+1$ is legal in X_W. As observed above, each subword of y has a unique offset of some size between 0 and $\ell - 1$. Further, considering a subword of size $d+1$, y_s^{s+d}, by assumption there is a $z \in \overline{W}$ and a $t < \ell$ such that $y_s^{s+d} = z_{s+t}^{s+t+d}$. It follows that the words y_s^{s+d-1} and y_{s+1}^{s+d} have the same offset t. Now by induction since all of the words of length d in y have a unique offset and all adjacent words have the same offset, it follows that all of the words in y have the same offset t. In particular, the two equal words y_0^{d-1} and y_m^{m+d-1} separated by one period of y have the same offset t.

On the other hand, let $z \in \overline{W}$ be such that $y_0^{d-1} = z_t^{t+d-1}$. If $t \geq m$, then since we also have $y_m^{m+d-1} = z_t^{t+d-1}$, we see that the offset of y_m^{m+d-1} is $t - m$ which is a contradiction. If $t < m$, then construct a new word $z' = wz \in \overline{W}$ by concatenating on the left an element of W. Then ${z'}_{\ell+t}^{\ell+t+d-1} = z_t^{t+d-1} = y_0^{d-1} = y_m^{m+d-1}$ so that y_m^{m+d-1} has offset $\ell - m + t$, which is less than ℓ and distinct from t, again contradicting the uniqueness of the offset □

Proof of Theorem 2 (Lower Bound). Let $h \geq 0$, $M > e^h$ and a sequence (a_n) be given with $a_n = o(n^{-k})$ for each k. We are aiming to build a minimal uniquely ergodic subshift X on M symbols with topological entropy h and a sequence (ℓ_i) such that defining the function $f(x) = d(x, X)$ gives $\Delta(\ell_i - 1) > a_{\ell_i - 1}$ for all i. We note that such a function f is Lipschitz and in particular lies in the intersection of the Hölder classes with exponents $\alpha \leq 1$.

To construct X, we will recursively construct a sequence of sets of words of increasing lengths. We will call the jth collection of words W_j. The collection W_j will consist of N_j words of length ℓ_j. We shall let X_j denote the one-sided subshift obtained by taking all shifts of arbitrary infinite concatenations of words in W_j. Since each word in W_j will be a concatenation of words in W_{j-1}, it follows that $X_j \subseteq X_{j-1}$. The desired subshift X will then be the intersection of the X_j.

Throughout the recursion, we will have $d_j \leq \ell_j$. The entropy of X_j is given by $h_j = (\log N_j)/\ell_j$. We shall arrange that $h + 5^{-j} \leq h_j \leq h + 2(5^{-j})$ for all j so that the intersection has the required entropy.

To start the recursion, we have an alphabet consisting of M symbols. We choose a k_0 such that $M > e^{h+3(5^{-k_0})}$. The first set will then be a set W_{k_0} constructed as follows using a standard marker argument from symbolic dynamics.

Note that since $M > 1$, the alphabet contains at least two symbols, 0 and 1 say. We will let U consist of all words of the form $10\ldots01w$, where the initial block $10\ldots01$ is of length a (to be chosen), the words w are of a fixed length b (also chosen below) and have the property that they do not contain any subword that is equal to the initial block $10\ldots01$ of length a. We claim that there exist arbitrarily large a and b such that

$$|U| \geq e^{(h+5^{-k_0})(a+b)}. \tag{1}$$

Assuming this for now, we define W_{k_0} to be an arbitrary subset of U consisting of $\lceil e^{(h+5^{-k_0})(a+b)}\rceil$ words, set $\ell_{k_0} = d_{k_0} = a+b$. Since we assumed that a and b may be arbitrarily large, we may take a and b such that

$$\frac{h+2}{|U|} < 5^{-(k_0+1)}. \tag{2}$$

To check that U (and hence W_{k_0}) has decoding length $a+b$, we suppose that $x = T^i z$ for $z \in \overline{U}$ and $i < a+b$; and $x' = T^{i'} z'$ for $z' \in \overline{U}$ and $i' < a+b$. Suppose further that $x_0^{a+b-1} = {x'}_0^{a+b-1}$. We need to show that $i = i'$. To see this, note that if x_j^{j+a-1} is the block $10\ldots01$ of length a where $0 < j < j+a \le a+b$ then by construction, we have $i = 0$ if $j = 0$ or $i = (a+b)-j$ otherwise. The same follows for x' so that we have $i = i'$ as required. On the other hand, if the block $10\ldots01$ does not occur in x_0^{a+b-1}, then by construction of U, the block occurs in x starting at the last occurrence of 1 in x_0^{a+b-1}. The same applies to x' and again we recover $i = i'$.

To show that it is possible to choose arbitrarily large a and b, such that (1) holds, we argue as follows: first choose a such that the subshift obtained by removing the fixed block of this length a has entropy greater than $h + 2(5^{-k_0})$. The length b is then chosen so that inserting copies of the fixed block at intervals of length b does not decrease the entropy by more than 5^{-k_0} (i.e. $(a/(a+b))(h+2(5^{-k_0})) < 5^{-k_0}$).

To perform the recursive step, suppose that W_{j-1} has already been determined. Choose m_j so that the following conditions hold:

$$\text{For all } n > m_j, \text{ we have } a_n \le n^{-2(5^j)-1}/2; \tag{3}$$

$$(\log m_j)/m_j < 5^{-j}; \text{ and} \tag{4}$$

$$m_j > (h+2)5^{j+1}. \tag{5}$$

We will now define an intermediate collection V_j of words consisting of concatenations of words in W_{j-1}. Let w_{j-1} be a distinguished word in W_{j-1}. We let V_j be the collection of all words $u = \xi_1\xi_2\ldots\xi_{N_{j-1}m_j}$ satisfying:

- $\xi_i \in W_{j-1}$ for each i;
- $\xi_i = w_{j-1}$ for $i = 1,\ldots,m_j$;
- Each word in W_{j-1} appears exactly m_j times in u

We can check that V_j consists of $(m_j(N_{j-1}-1))!/(m_j!)^{N_{j-1}-1}$ words of length $g_j = m_j N_{j-1}\ell_{j-1}$ and we can verify that

$$
\begin{aligned}
|V_j| &\geq N_{j-1}^{m_j(N_{j-1}-1)} e^{-m_j} m_j^{-N_{j-1}/2} e^{-N_{j-1}} \\
&= e^{h_{j-1}\ell_{j-1}m_j(N_{j-1}-1)} e^{-m_j} m_j^{-N_{j-1}/2} e^{-N_{j-1}} \\
&\geq \exp\left(m_j N_{j-1}\ell_{j-1}\left(h_{j-1} - \frac{h_{j-1}}{N_{j-1}} - \frac{1}{\ell_{j-1}N_{j-1}} - \frac{\log m_j}{2m_j\ell_{j-1}} - \frac{1}{m_j\ell_{j-1}}\right)\right) \\
&\geq \exp\left(g_j \left(h + 4(5^{-j}) - \frac{\log m_j}{m_j}\right)\right) \\
&\geq \exp\left(g_j (h + 3(5^{-j}))\right),
\end{aligned}
$$

where for the third inequality, we used the observation that $h_{j-1}/N_{j-1} + 1/(N_{j-1}\ell_{j-1}) \leq 5^{-j}$ (this is true when $j = k_0 + 1$ by (2) and follows from (5) for $j > k_0 + 1$) and $h_{j-1} \geq h + 5^{1-j}$; and for the last inequality, we used condition (4).

Now from V_j, we extract $A_j = \lceil \exp(5^{-j} g_j) \rceil$ disjoint subcollections $U_{j,0} \ldots U_{j,A_j-1}$, each consisting of $B_j = \lceil \exp(g_j(h + 5^{-j})) \rceil$ words in V_j. As a further condition, we impose the following

$$
\text{(6)} \qquad \begin{array}{c} \text{Each of the } N_{j-1}^2 \text{ possible pairs of words in } W_{j-1} \text{ occur} \\ \text{as consecutive words in some element of } \bigcup_i U_{j,i}. \end{array}
$$

This is easy to arrange since $|V_j|$ is exponentially larger than $|W_{j-1}|$. The jth collection of words W_j then consists of all possible concatenations of one word from $U_{j,0}$ followed by a word from $U_{j,1}$ and so on up to a word from U_{j,A_j-1}.

For the new alphabet, we have $\ell_j = A_j m_j N_{j-1}\ell_{j-1}$, $N_j = B_j^{A_j}$ so that

$$
h_j = (\log N_j)/\ell_j = (\log B_j)/g_j \in [h + 5^{-j}, h + 2(5^{-j})].
$$

We claim that $d_j = 2g_j = 2m_j N_{j-1}\ell_{j-1}$ is a decoding length for W_j. To see this, let $x \in X_j$. By definition, we have $x = T^i z$ for some $z \in \overline{W}_j$ and $i < \ell_j$. Suppose also that $x_0^{d_j-1} = {z'}_{i'}^{i'+d_j-1}$ for some $z' \in \overline{W}_j$ and $i' < \ell_j$. We need to show that $i = i'$.

Write $i = eg_j - f$, where $f < g_j$. Notice that $T^f x = T^{eg_j} z \in \overline{V}_j$. It follows that $x_f^{d_j-1} = z_{i+f}^{i+d_j-1}$ and in particular, $x_f^{f+g_j-1} = z_{eg_j}^{(e+1)g_j-1}$. Since $z \in \overline{W}_j$, this block belongs to $U_{j,e}$.

We also have however $x_f^{f+g_j-1} = {z'}_{i'+f}^{i'+f+g_j-1}$. Since $x \in \overline{W}_{j-1}$, it follows from the fact that $g_j > d_{j-1}$ that $i' \equiv i \pmod{\ell_{j-1}}$ as if not we could use z and z' to produce a point in W_{j-1} with two different offsets. In particular, it follows that $T^{i'+f}z' \in \overline{W}_{j-1}$.

Now we see that $z'^{\,i'+f+g_j-1}_{i'+f}$ consists of m_j consecutive w_{j-1}'s followed by the remainder of a word in $U_{j,e}$. Since $z' \in \overline{W}_j$, this uniquely identifies the position of $T^{i'+f}z'$ within the sequence of W_j blocks. It follows that $i' + f = eg_j = i + f$ so that $i = i'$, hence showing that d_j is a decoding length for W_j.

This completes the recursive construction of the X_j and hence by taking an intersection, the construction of X is complete. The inequalities for the entropy of the X_j demonstrate that X has the required entropy. Let $f(x) = d(x, X)$ so that f is a non-negative Lipschitz function that vanishes on X (and nowhere else).

We now check that $\Delta(\ell_j - 1) > a_{\ell_j - 1}$. To see this, let y be a periodic point in $W_0^{\mathbb{N}}$ of period at most $\ell_j - 1$. Since W_j consists of words of length ℓ_j and has decoding length d_j, we see from Lemma 6 that y must contain a subword $y_s^{s+d_j}$ of length $d_j + 1$ that is not legal in X_j and hence not legal in the subsystem X. It follows that $f(T^s(y)) \geq 2^{-d_j-1}$ so that the average value of f along the orbit of y is at least $2^{-d_j-1}/\ell_j$. From the definition of d_j, we see that $2^{-d_j} = A_j^{-2(5^j)} \geq \ell_j^{-2(5^j)}$ so that $\int f\, d\mu_y \geq (1/2)\ell_j^{-2(5^j)-1} \geq a_{\ell_j-1}$ using condition (3).

To see that X is minimal, let $x \in X$ and consider a block $x_0 \ldots x_\ell$. Let $\ell_{j-1} > \ell$. Then since $x \in X_{j-1}$, the block is contained in the concatenation of at most two words from W_{j-1}. By condition (6), there is a word in W_j containing the given block. This word appears in any word in X of length $2\ell_{j+1}$ so that each orbit is dense in X.

Finally for unique ergodicity, observe that any word in W_{j-1} occurs between $(n-1)m_j$ times and $(n+1)m_j$ times in any block of length $n\ell_j$ so that the rate of occurrence converges uniformly to m_j/ℓ_j. □

Proof of Corollary 3. If T is either a C^2 expanding map or an Axiom A diffeomorphism restricted to a hyperbolic set, then T is Hölder continuously semiconjugate to a shift space. Since distances in the original space are bounded above by a power of the distance in the shift space, the upper bounds obtained above on the shift space are sufficient to give an upper bound in the original space.

For the lower bounds, by restricting to a suitable subshift, one can ensure that the shift is conjugate to a totally disconnected set in the original space. Working in the same manner as above on this subshift, one obtains the lower bounds in the original space. □

We would like to thank Oliver Jenkinson for his careful reading of the manuscript and many helpful comments, and Julien Cassaigne for useful conversations. The anonymous referee also did an excellent job providing a number of suggestions and corrections that resulted in an improvement to the paper.

References

[1] T. Bousch. Le poisson n'a pas d'arêtes. *Ann. Inst. H. Poincaré Probab. Statist.*, 36:489–508, 2000.

[2] T. Bousch. La condition de Walters. *Ann. Sci. École Norm. Sup. (4)*, 34:287–311, 2001.

[3] T. Bousch and O. Jenkinson. Cohomology classes of dynamically non-negative C^k functions. *Invent. Math.*, 148:207–217, 2002.

[4] L. Caccetta and R. Häggkvist. On minimal digraphs with given girth. In *Proceedings of the Ninth Southeastern Conference on Combinatorics, Graph Theory, and Computing (Florida Atlantic Univ., Boca Raton, Fla., 1978)*, Congress. Numer., XXI, pages 181–187, Winnipeg, Man., 1978. Utilitas Math.

[5] V. Chvátal and E. Szemerédi. Short cycles in directed graphs. *J. Combin. Theory Ser. B*, 35:323–327, 1983.

[6] D. Collier and I. Morris. 2005. preprint.

[7] G. Contreras, J. Delgado, and R. Iturriaga. Lagrangian flows: the dynamics of globally minimizing orbits. II. *Bol. Soc. Brasil. Mat. (N.S.)*, 28:155–196, 1997.

[8] G. Contreras, A. O. Lopes, and P. Thieullen. Lyapunov minimizing measures for expanding maps of the circle. *Ergodic Theory Dynam. Systems*, 21:1379–1409, 2001.

[9] J.-P. Conze and Y. Guivarc'h. Croissance des sommes ergodiques. manuscript, 1993.

[10] B. Hunt and E. Ott. Optimal periodic orbits of chaotic systems occur at low period. *Phys. Rev. E*, 54:328–337, 1996.

[11] O. Jenkinson. Ergodic optimization. *Discrete and Continuous Dynamical Systems*, 15:197–224, 2006.

[12] J. Shen. On the Caccetta-Häggkvist conjecture. *Graphs Combin.*, 18:645–654, 2002.

[13] G. Yuan and B. R. Hunt. Optimal orbits of hyperbolic systems. *Nonlinearity*, 12:1207–1224, 1999.

E-mail address: `bressaud(a)iml.univ-mrs.fr`

Institut de Mathématiques de Luminy, Campus de Luminy, Université de la Meditarannée Aix-Marseille II, 13288 Marseille, FRANCE

E-mail address: `aquas(a)uvic.ca`

Department of Mathematics and Statistics, University of Victoria, Victoria, CANADA V8W 3P4