ERGODIC OPTIMIZATION OF SUPER-CONTINUOUS FUNCTIONS IN THE SHIFT

ANTHONY QUAS AND JASON SIEFKEN

Abstract. Ergodic Optimization is the process of finding invariant probability measures that maximize the integral of a given function. It has been conjectured that "most" functions are optimized by measures supported on a periodic orbit, and it has been proved in several separable spaces that an open and dense subset of functions is optimized by measures supported on a periodic orbit. All known positive results have been for separable spaces. We give in this paper the first positive result for a non-separable space, the space of *super-continuous* functions on the full shift, where the set of functions optimized by periodic orbit measures contains an open dense subset.

1. Introduction

Given an expansive map $T : \Omega \to \Omega$ and a continuous function f, we say that a T-invariant probability measure μ *optimizes* f if

$$\int f \, \mathrm{d}\mu \geq \int f \, \mathrm{d}\nu$$

for all T-invariant probability measures ν. If y is a periodic point (i.e., $T^i y = y$ for some i), let μ_y be the unique T-invariant probability measure supported on $\mathcal{O}y$, the orbit of y. We call μ_y a *periodic orbit measure*. If μ_y optimizes f, we will also say that f is optimized by the periodic point y.

General Belief. *"Most" functions are optimized by measures supported on a periodic orbit.*

"Most" can take various meanings, but for our purposes, we consider "most" to be an open dense set or a residual set.

Conjecture 1. *In an expansive dynamical system, the set of Lipschitz functions optimized by periodic orbit measures contains an open set that is dense in the class of Lipschitz functions.*

Analogs to Conjecture 1 have been shown false in the general case of continuous functions [6], however they have been shown true in a handful of separable spaces. Further, various numerical experiments on many important dynamical systems support this conjecture (and hint towards some very interesting relationships between parameterized families of functions and the period of optimizing orbits) [4, 5, 8].

We present a non-separable space where the analog of Conjecture 1 holds true. Let $\Omega = \mathcal{A}^{\mathbb{N}}$ be the one-sided shift space on a finite alphabet. For a sequence $A_n \searrow 0$, define a metric $d_A(x, y) = A_n$ if x and y first differ in the nth place (i.e. $(x)_i = (y)_i$ for $0 \leq i < n$; $(x)_n \neq (y)_n$). Let $C_A(\Omega)$ denote the set of Lipschitz functions with respect to the d_A

Date: July 12, 2011.
2000 *Mathematics Subject Classification.* Primary: 37D20. Secondary: 37A05, 37D35, 37E10.
AQ was supported by NSERC; JS was supported by NSERC and the University of Victoria.

metric, equipped with the d_A-Lipschitz norm. If $\{A_n\}$ satisfies the additional property that $A_{n+1}/A_n \to 0$, we call $f \in C_A(\Omega)$ *super continuous.*

Theorem 2. *Suppose $A = \{A_n\}$ and $A_{n+1}/A_n \to 0$. For a periodic orbit measure μ_y supported on $\mathcal{O}y$, let $P_y = \{f \in C_A(\Omega) : \mu_y$ is the unique maximizing measure$\}$. Then, $\bigcup_{y\ periodic}(P_y)^\circ$ is dense in all of $C_A(\Omega)$ under the A-norm topology (where $(P_y)^\circ$ is the interior of P_y).*

We will briefly survey the most well-known positive results. A function f is a Walters function (introduced by Walters in [7]) if for every $\varepsilon > 0$ there exists a $\delta > 0$ so that for all $n \in \mathbb{N}$ and x and y,

$$\max_{0\leq i<n}\{d(T^i x, T^i y)\} \leq \delta \implies |S_n f(x) - S_n f(y)| < \varepsilon,$$

where $S_n f(w) = \sum_{i=0}^{n-1} f(T^i w)$. Bousch shows for Walters functions, the analog of Conjecture 1 holds [2].

Contreras, Lopes, and Thieullen showed in [3] that when using a Hölder norm external to a particular union of Hölder spaces, the analog of Conjecture 1 for Hölder spaces holds. Yuan and Hunt made significant progress towards proving Conjecture 1, though the full result has not yet been proved.

Presented are the already-established theorems for comparison. Note that although the theorems are stated in a variety of contexts (expanding maps of the circle, one-sided shifts etc.), the essence of the problem is present in the simple setting of the one-sided shift.

Theorem (Bousch [2]). *Let $T : X \to X$ be the one-sided shift map and let W denote the set of Walters functions on X. If $P \subset W$ is the set of Walters functions optimized by measures supported on periodic points, then P contains an open set dense in W with respect to the Walters norm.*

Theorem (Contreras-Lopes-Thieullen [3]). *Let T be a $C^{1+\alpha}$ expanding map of the circle. Let H_β be the set of β-Hölder functions on S^1 and let $\mathcal{F}_{\alpha+} = \bigcup_{\beta>\alpha} H_\beta$. Let $P_{\alpha+} \subset \mathcal{F}_{\alpha+}$ be the subset of functions uniquely optimized by measures supported on a periodic point. Then $P_{\alpha+}$ contains a set that is open and dense in $\mathcal{F}_{\alpha+}$ under the H_α topology (i.e., the α-Hölder norm).*

Theorem (Yuan and Hunt [9]). *Let $T : M \to M$ be an Axiom A map or an expanding map from a manifold to itself and let C_{Lip} denote the class of Lipschitz continuous functions. For any $f \in C_{\mathrm{Lip}}$ optimized by a measure generated by an aperiodic point, there exists an arbitrarily small perturbation of f such that that measure is no longer an optimizing measure. Further, any $f \in C_{\mathrm{Lip}}$ optimized by a periodic orbit measure can be perturbed to be stably optimized by this periodic orbit measure.*

With the inclusion of this paper, the current state of the standing conjecture is somewhat curious. Notice that super-continuous functions are Lipschitz functions and Lipschitz functions are Walters functions. So, for both a larger and a smaller class than Lipschitz functions, analogs of Conjecture 1 have shown to be true, and yet proof of the Lipschitz case remains elusive.

1.1. **Notation & Definitions.** For some finite alphabet $\mathcal{A}$, let $\Omega = \mathcal{A}^{\mathbb{N}}$ be the space of one-sided infinite sequences on $\mathcal{A}$. For us $\mathbb{N}$ includes 0.

$T : \Omega \to \Omega$ is the usual shift operator, with T-invariant Borel probability measures on Ω denoted $\mathcal{M}$. We write $\mathcal{O}x$ for the orbit of x under T, and we say S is a *segment* of $\mathcal{O}x$ if it is an ordered list of the form $(T^i x, T^{i+1}x, \ldots, T^{i+p-1}x)$ for some i, p. Abusing notation, we may say $S \subset \mathcal{O}x$.

We use d to denote the standard metric on sequences. That is, $d(x, y) = 2^{-k}$ where $k = \inf\{i : (x)_i \neq (y)_i\}$ and $(z)_i$ is the ith symbol of z. We follow the convention that $2^{-\infty} = 0$.

Definition 3 (Shadowing). *For two points x, y, we say that x ε-*shadows *a segment $S = (T^m y, \ldots, T^{m+n-1}y) \subset \mathcal{O}y$ if*

$$d(T^i x, T^{i+m}y) \leq \varepsilon$$

for all $0 \leq i < n$.

Definition 4 (ε-close). *A point x is said to stay ε-close to a set Y for p steps if for all $0 \leq i < p$,*

$$d(T^i x, Y) \leq \varepsilon.$$

Notation 5 (Ergodic Average). *For a function f and a point x,*

$$\langle f \rangle (x) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x),$$

when the limit exists.

Notation 6. *If $x = a_0 a_1 a_2 \cdots$ is a point,*

$$(x)_i^j = a_i a_{i+1} \cdots a_{j-1} a_j$$

is the subword of x from position i to j.

2. Summable Variation

Definition 7 (Variation). *The* variation *of a function over level k cylinder sets is the maximum a function changes in a distance of 2^{-k}. That is, if f is a function*

$$\mathrm{var}_k(f) = \sup\{|f(x) - f(y)| : d(x, y) \leq 2^{-k}\}.$$

Note that in a shift space, we have additional structure because distances can only take values of the form 2^{-k}.

Definition 8 (Summable Variation). *The function f is of* summable variation *if*

$$\sum_{k=0}^{\infty} \mathrm{var}_k(f) < \infty.$$

Notation 9. $\mathrm{V}_k(f)$ *represents the tail sum of the variation of f over distances smaller than 2^{-k+1}. That is*

$$\mathrm{V}_k(f) = \sum_{j=k}^{\infty} \mathrm{var}_j(f).$$

Functions of summable variation form a much larger class than Lipschitz functions. However the general method used in this paper to show Theorem 2 is to perturb functions by a small multiple of some canonical "sharpest" function. Yuan and Hunt used this strategy when dealing with Lipschitz functions by perturbing by $-d(x, \mathcal{O}y)$ [9]. But, for functions of summable variation (with the natural norm of $\|f\| = V_0(f) + \|f\|_\infty$), there is no such "sharpest" function. Using the A-norms gives us these sharpest functions again.

We will frequently refer to A-metrics and A-norms as briefly introduced earlier.

Definition 10 (A-sequence)**.** *An A-*sequence*, $(A_n)_{n=0}^\infty$, is a decreasing sequence of positive numbers with $A_n \to 0$.*

If there exists $0 < \delta < 1$ such that $A_{n+1}/A_n < 1 - \delta$ for each n, then we say that (A_n) is lacunary*.*

Recall that the metric d_A is defined by $d_A(x, y) = A_n$ if $(x)_i = (y)_i$ for $0 \le i < n$ but $(x)_n \ne (y)_n$.

Definition 11 (A-norm)**.** *If (A_n) is an A-sequence, the Lipschitz constant of f is $\mathrm{Lip}_A(f) = \sup_k \mathrm{var}_k(f)/A_k$. The A-norm is defined by $\|f\|_A = \mathrm{Lip}_A(f) + \|f\|_\infty$.*

Of course if A is the sequence $(2^{-n})_{n=0}^\infty$, we recover the standard distance and Lipschitz norm. We write the set of Lipschitz functions with respect to d_A as $C_A(\Omega)$ or simply C_A.

Notice that since A satisfies $A_n \to 0$, $C_A(\Omega) \subset C(\Omega)$ is a subset of the continuous functions on Ω. Further, C_A is a non-separable Banach space as the functions $f_x(\cdot) = d(x, \cdot)$ for $x \in \Omega$ are an uncountable uniformly discrete set.

3. Preliminary lemmas

We will first establish several results that do not depend on super continuity.

Definition 12 (In Order for One Step)**.** *For points x, y, let $S = (T^j y, T^{j+1} y, \ldots, T^{j+k} y) \subset \mathcal{O}y$, and suppose that there is a unique closest point $y' \in S$ to x. That is,*

$$d(x, y') < d(x, S \backslash \{y'\}).$$

We say that x follows S in order for one step *if $Ty' \in S$ and Ty' is the unique closest point to Tx. That is $Ty' \in S$ and*

$$d(Tx, Ty') < d(Tx, S \backslash \{Ty'\}).$$

Definition 13 (In Order)**.** *For some point y, let $S = (T^j y, T^{j+1} y, \ldots, T^{j+k} y) \subset \mathcal{O}y$. For some point x, we say that x* follows S in order *for p steps if $x, Tx, \ldots, T^{p-1}x$ each follow S in order for one step.*

Following in order is very similar to the concept of shadowing except that the distance requirement in shadowing is replaced by a uniqueness requirement. The following In Order Lemma is due to Yuan and Hunt [9].

Lemma 14 (In Order Lemma)**.** *Let y be a periodic point of period p, and let*

$$\rho \le \min_{0 \le i < j < p} d(T^i y, T^j y)/4.$$

For any point x, if x stays ρ-close to $\mathcal{O}y$ for $k+1$ steps, then x follows $\mathcal{O}y$ in order for k steps. That is, there exists some i' such that for $0 \le j \le k$,

$$d(T^j x, T^{i'+j} y) \le \rho.$$

Proof. Let $\gamma = \min_{0\le i<j<p} d(T^i y, T^j y)$. We first derive a fact about the shift space due to its ultrametric properties. Suppose $y', y'' \in \mathcal{O}y$ and for some point x, $d(x,y'), d(x,y'') \le \gamma/2$. By the ultrametric triangle inequality we have

$$d(y',y'') \le \max(d(x,y'), d(x,y'')) \le \gamma/2. \tag{1}$$

Since γ was the smallest distance between points in $\mathcal{O}y$, equation (1) gives $y' = y''$. This shows that for any point x, if $d(x,\mathcal{O}y) \le \gamma/2$, then there is a unique closest point in $\mathcal{O}y$ to x.

Let x be a point that stays ρ-close to $\mathcal{O}y$ for $k+1$ steps. By definition, we have

$$d(x,\mathcal{O}y) \le \rho \le \gamma/4.$$

Since γ is the minimum distance between points in $\mathcal{O}y$, there is a unique i' such that

$$d(x, T^{i'} y) \le \rho.$$

We then have that

$$d(Tx, T^{i'+1} y) \le 2\rho \le \gamma/2,$$

and so $T^{i'+1}y$ is the unique closest point to Tx. Thus, x follows $\mathcal{O}y$ in order for one step. But, by assumption we have $d(Tx,\mathcal{O}y) \le \rho$, so $d(Tx,\mathcal{O}y) = d(Tx, T^{i'+1}y)$ gives us that Tx follows $\mathcal{O}y$ in order for one step and so x follows $\mathcal{O}y$ in order for two steps. Continuing by induction, we see that x follows $\mathcal{O}y$ in order for k steps; that is

$$d(T^j x, T^{i'+j} y) \le \rho \qquad \text{for} \quad 0 \le j \le k.$$

■

Lemma 15 (Shadowing Lemma). *For a point y, let $S = (T^i y, T^{i+1} y, \dots, T^{i+k-1} y)$ be a segment of $\mathcal{O}y$. For any $\rho < 1$, if a point x ρ-shadows S for k steps, the distance from $T^j x$ to S for $0 \le j < k$ is bounded by*

$$d(T^j x, T^{i+j} y) \le \rho 2^{-((k-1)-j)}.$$

Proof. Let $l = \inf\{w : 2^{-w} \le \rho\}$ and note $\rho < 1$ implies $l \ge 1$. Since x ρ-shadows S for k steps, we have $(T^j x)_0^{l-1} = (T^{i+j} y)_0^{l-1}$ for $0 \le j \le k-1$, and so $(x)_0^{k+l-2} = (T^i y)_0^{k+l-2}$, which gives the result. ■

Lemma 16 (Parallel Orbit Lemma). *For a function of summable variation f, if $T^m x$ 2^{-r}-shadows $\mathcal{O}y$ for k steps (i.e., there exists i so $d(T^{m+j}x, T^{i+j}y) \le 2^{-r}$ for $0 \le j < k$), then for $r > 0$,*

$$\sum_{j=0}^{k-1} \left| f(T^{m+j} x) - f(T^{i+j} y) \right| \le \mathrm{V}_r(f).$$

Proof. Suppose x, y are points such that $d(T^{m+j}x, T^{i+j}y) \le 2^{-r}$ where $r \ge 1$ for $0 \le j < k$. The Shadowing Lemma (Lemma 15) gives us that

$$d(T^{m+j}x, T^{i+j}y) \le 2^{-(r+(k-1)-j)}.$$

We then have

$$\sum_{j=0}^{k-1} |f(T^{m+j}x) - f(T^{i+j}y)| \le \sum_{j=r}^{r+k-1} \mathrm{var}_j(f) \le \mathrm{V}_r(f).$$

■

4. Mañé-Conze-Guivarc'h normal form and main result

Heuristically, let us consider the following: Suppose f is optimized by $\mu_{\max}$ and $\int f\mathrm{d}\mu_{\max} = 0$. We will define a function f^* to represent the "payoff of going backwards to infinity." Before we describe what f^* means, let us consider the payoff of going backwards a finite number of steps. For a point x, there is some point $a_1^1 x \in T^{-1}x$ such that $f(a_1^1 x) \geq f(b_1 x)$ for any symbol b_1. In other words, $a_1^1 x$ is a maximal one-step backwards extension of x. Continuing, there is some point $a_2^2 a_1^2 x \in T^{-2}x$ so that $f(a_2^2 a_1^2 x) + f(a_1^2) \geq f(b_2 b_1 x) + f(b_1 x)$ for any word $b_2 b_1$, making $a_2^2 a_1^2 x$ a maximal two-step backwards extension of x. It is important to note that the symbol a_1^2 need not be the same as the symbol a_1^1, and so it is in no way immediate that there should be some convergent way to pick an infinite maximal backwards extension of x.

However, ignoring these issues for the moment, one can imagine that n-step backwards extensions of x look more and more like generic points of $\mu_{\max}$ (if $\mu_{\max}$ is a periodic orbit measure, this should be especially plausible). We now informally define f^* as

$$f^*(x) = f(a_1^\infty x) + f(a_2^\infty a_1^\infty x) + f(a_3^\infty a_2^\infty a_1^\infty x) + \cdots ,$$

where $\cdots a_3^\infty a_2^\infty a_1^\infty x$ is an infinite maximal backwards extension of x. Since $\int f\mathrm{d}\mu_{\max} = 0$, it is reasonable to expect that if f^* converges, it is bounded above. Ignoring any issues of convergence, consider

$$f^* \circ T - f^*.$$

Suppose $x = x_0 x_1 \cdots$ is a point with maximal backwards extension $\cdots a_2 a_1 x_0 x_1 \cdots$. We immediately see $(f^* \circ T - f^*)(x) \geq f(x)$, since either the maximal backwards extension of $Tx = x_1 x_2 \cdots$ is $\cdots a_2 a_1 x_0 x_1 \cdots$, which would give us $(f^* \circ T - f^*)(x) = f(x)$, or there is an alternative backwards extension of Tx that yields a bigger payoff than $\cdots a_2 a_1 x_0 x_1 \cdots$ and so $(f^* \circ T - f^*)(x) > f(x)$.

Since $f^* \circ T - f^*$ is a co-boundary (a function of the form $h - h\circ T$) and so integrates to zero with respect to any invariant measure, the function $\hat{f} = f - (f^* \circ T - f^*)$ is co-homologous to f (and so $\int f\mathrm{d}\mu = \int \hat{f}\mathrm{d}\mu$ for all invariant measures μ), with the added property that $\hat{f} \leq 0$.

The Mañé-Conze-Guivarc'h procedure is a way of producing a well defined f^*. We use a method due to Bousch [1], which produces f^* as a fixed point of an operator that reflects the idea of a maximal backwards extension.

For $f \in C_A$, define the operator $\Phi_f : C_A \to C_A$ by

$$(\Phi_f g)(x) = \max_{y \in T^{-1}x} \{(f+g)(y)\}.$$

Proposition 17 (Bousch)**.** *Let (A_n) be a lacunary A-sequence. For a fixed function $f \in C_A$ with $\sup_{\mu\in\mathcal{M}} \int f\mathrm{d}\mu = 0$, the operator Φ_f as defined above has a fixed point.*

The proof follows standard lines with minor adaptations for the case of A-norms rather than Lipschitz norms. We briefly summarize the steps, referring the reader to Bousch [1] for more details.

Proof sketch. Let $A_{n+1}/A_n < 1 - \delta$ for all n (where $0 < \delta < 1$). We claim that Φ_f maps $C = \{g\colon \mathrm{Lip}_A(g) \leq \mathrm{Lip}_A(f)/\delta\}$ into itself. We do part of this step in detail since we need a fact from it later. Let $g \in C$ and let x and x' differ first in their $(n-1)$st coordinates. Using the notation ix to denote the sequence with its first symbol defined by $(ix)_0 = i$ and all remaining symbols defined by $(ix)_{k+1} = x_k$, we have

$$\begin{aligned}\Phi_f(g)(x) - \Phi_f(g)(x') &= \max_i (f(ix) + g(ix)) - \max_j (f(jx') + g(jx')) \\ &\le \max_i (f(ix) + g(ix) - f(ix') - g(ix')) \\ &\le \mathrm{var}_n(f) + \mathrm{var}_n(g)\end{aligned}$$

By symmetry we deduce

$$\mathrm{var}_{n-1}(\Phi_f(g)) \le \mathrm{var}_n(f) + \mathrm{var}_n(g). \tag{2}$$

Straightforward manipulation then shows that $\Phi_f(g) \in C$.

Taking a quotient of C by the relation $\sim$ where two functions g and g' are related if they differ by a constant, one obtains a compact (with respect to the quotient of the supremum norm topology) convex set $C/\sim$ on which Φ_f acts continuously. Hence, there is a fixed point. This fixed point corresponds to a function $h \in C$ such that $\Phi_f(h) = h + \beta$ for some constant β. One then shows that $\sup_{\mu\in\mathcal{M}} \int f\, d\mu = 0$ implies $\beta = 0$ ∎

Theorem 18. *Let (A_n) be a lacunary A-sequence. There exists a constant $\gamma_A > 1$, dependent only on the choice of A-sequence, such that for all $f \in C_A$ with $\sup_{\mu\in\mathcal{M}} \int f \mathrm{d}\mu = 0$, there exists a co-homologous function $\hat{f}$ with $\hat{f} \le 0$ and*

$$\|\hat{f}\|_A \le \gamma_A \|f\|_A \qquad \mathrm{V}_n \hat{f} \le \gamma_A \|f\|_A A_n.$$

Proof. Suppose $A_{n+1}/A_n < 1 - \delta$ for all n (for some $0 < \delta < 1$). By Proposition 17, we may find h, a fixed point of Φ_f with

$$\|h\|_A \le \frac{\mathrm{Lip}_A(f)}{\delta} + \|h\|_\infty \le (A_0 + 1)\frac{\|f\|_A}{\delta}.$$

However, from (2) we have

$$\mathrm{var}_{n-1}(h) = \mathrm{var}_{n-1}(\Phi_f h) \le \mathrm{var}_n(f) + \mathrm{var}_n(h).$$

This gives

$$\frac{\mathrm{var}_n(h\circ T)}{A_n} \le \frac{\mathrm{var}_{n-1}(h)}{A_n} \le \frac{\mathrm{var}_n(f) + \mathrm{var}_n(h)}{A_n},$$

and so $\|h\circ T\|_A \le \|f\|_A + \|h\|_A$. Let $\hat{f} = f + h - h\circ T$. $\hat{f}$ has the desired properties and

$$\|\hat{f}\|_A \le \|f\|_A + \|h\|_A + \|h\circ T\|_A \le 2\|f\|_A + 2\|h\|_A \le \frac{2(A_0+1+\delta)}{\delta}\|f\|_A.$$

Let us now focus on finding a constant such that $\mathrm{V}_n \hat{f} \le K\|f\|_A A_n$. From our bound on $\|\hat{f}\|_A$, we know $\mathrm{var}_k \hat{f} \le \frac{2(A_0+1+\delta)}{\delta}\|f\|_A A_k$. $A_{k+1}/A_k < 1-\delta$ for all k gives that $\sum_{k\ge n} A_k \le A_n/\delta$ and so

$$\mathrm{V}_n \hat{f} \le \frac{2(A_0+1+\delta)}{\delta^2}\|f\|_A A_n$$

Letting $\gamma_A = 2(A_0+1+\delta)/\delta^2$ completes the proof. ∎

It should be noted that Theorem 18 can trivially be applied to functions f where $\sup_{\mu\in\mathcal{M}} \int f \mathrm{d}\mu = \beta \ne 0$ by letting $\hat{f} = \widehat{f-\beta} + \beta$.

Corollary 19. *Theorem 18 holds with the weakened assumption that $\limsup A_{n+1}/A_n < 1$.*

Proof. Since $\limsup A_{n+1}/A_n < 1$, we can construct a sequence B_n such that $B_{n+1}/B_n < 1-\delta$ for some $0<\delta<1$ and $B_i = A_i$ for $i > N$ for some finite N. Since we only changed a finite number of terms of A to produce B, $\|\cdot\|_A$ and $\|\cdot\|_B$ are equivalent. Let M be such that $\|f\|_A \le M\|f\|_B$ for all $f \in C_A$ and $M' = \max A_n/B_n$. Letting $\gamma_A = MM'\gamma_B$ completes the proof. ■

Though not dependent on Theorem 18, it is convenient to note that γ_A from Theorem 18 also bounds $\mathrm{V}_n f$ in the expected way.

Fact 20. *If (A_n) is a lacunary A-sequence, then for $f \in C_A$*

$$\mathrm{V}_n f \le \gamma_A \|f\|_A A_n,$$

where γ_A is as in Theorem 18.

We now have machinery in place to give a quick proof of Proposition 21, which establishes a relationship between the number of points in the support of a periodic orbit measure and how close such measures come to optimizing a fixed function. This result was first established by Yuan and Hunt (without using the Mañé-Conze-Guivarc'h Lemma) in [9] for Lipschitz functions.

Proposition 21 (Yuan and Hunt)**.** *Let (A_n) be a lacunary A-sequence. Let $f \in C_A$ and x be an optimal orbit for f (i.e., a typical point of a maximizing measure). Let y be a point of period p, and $r > 0$. If a segment of $\mathcal{O}x$ 2^{-r}-shadows $\mathcal{O}y$ for one period (i.e., there exist m, m' such that $d(T^{i+m}x, T^{i+m'}y) \le 2^{-r}$ for $0 \le i < p$), then*

$$\langle f\rangle(x) - \gamma_A\|f\|_A A_r/p \le \langle f\rangle(y) \le \langle f\rangle(x),$$

where γ_A is as in Theorem 18.

Proof. Let y be a period p point with the property that a segment of $\mathcal{O}x$ 2^{-r}-shadows $\mathcal{O}y$ for p steps. By renaming some $T^j y$ as y, without loss of generality we may assume that a segment of $\mathcal{O}x$ 2^{-r}-shadows y. That is, there exists some m so that $d(T^{m+i}x, T^i y) \le 2^{-r}$ for $0 \le i < p$. Let $x' = T^m x$.

By Theorem 18, we may find $\hat f$ co-homologous to f with $\hat f(\mathcal{O}x) = \hat f(\mathcal{O}x') = \langle f\rangle(x)$. Since for $0 < i \le p$ we have

$$d(T^i x', T^i y) \le 2^{-(r+(p-1-i))},$$

we may apply the Parallel Orbit Lemma (16) to get

$$\left|\sum_{i=0}^{p-1}\left(\hat f(T^i x') - \hat f(T^i y)\right)\right| = \left|\left(\sum_{i=0}^{p-1}\hat f(T^i x')\right) - p\left\langle \hat f\right\rangle(y)\right| \le \mathrm{V}_r \hat f.$$

The proposition follows from the fact that $\hat f(T^i x') = \langle f\rangle(x)$ and that by Theorem 18 $\mathrm{V}_r\hat f \le \gamma_A\|f\|_A A_r$. ■

Using methods similar to those in Yuan and Hunt[9], one can show that Proposition 21 holds for any function f of summable variation, and one can produce a slightly stronger bound of $\langle f\rangle(x) - 4V_r f/p \le \langle f\rangle(y) \le \langle f\rangle(x)$.

We are now ready to prove Theorem 2 by using $d_A(\cdot, \mathcal{O}y)$ as a "sharpest" function that will penalize any measure that gives mass to $(\mathcal{O}y)^c$.

Theorem (Theorem 2)**.** *Let (A_n) be an A-sequence satisfying $A_{n+1}/A_n \to 0$. For a periodic orbit measure μ_y supported on $\mathcal{O}y$, let $P_y = \{f \in C_A(\Omega) : \mu_y$ is the unique maximizing measure$\}$. Then, $\bigcup_{y \text{ periodic}} (P_y)^\circ$ is dense in $C_A(\Omega)$ (where $(P_y)^\circ$ is the interior of P_y).*

Proof. We will show that for any function f, there exists an arbitrarily small perturbation, $\tilde{f}$, of f and a periodic orbit measure μ_y, such that all functions in an open neighbourhood of $\tilde{f}$ are uniquely optimized by μ_y.

Since $\liminf A_{n+1}/A_n = 0$, by Corollary 19, passing to an equivalent norm if necessary, we may assume $A_{n+1}/A_n \le 1/2$ for all n. Fix $f \in C_A$ and let $\mu_{\max}$ be an optimizing measure for f. Fix $x \in \operatorname{supp}(\mu_{\max})$. Without loss of generality, assume $\langle f \rangle (x) = 0$ and let $\hat{f}$ be co-homologous to f with $\hat{f} \le 0$.

Suppose we showed that an arbitrarily small perturbation $\hat{f} + g$ of $\hat{f}$ were such that the open ball of radius ε about $\hat{f}+g$ is uniquely optimized by a periodic orbit measure μ_y. Since $\hat{f}$ and f are co-homologous, this means that $f + g$ is uniquely optimized by μ_y and in fact the open ball of radius ε about $f + g$ is uniquely optimized by μ_y. Thus, it is sufficient to only consider small perturbations of $\hat{f}$.

Fix $0 < \varepsilon < 1$. For a fixed k (to be determined later), find a minimal recurrence in x of a block of k symbols. That is, find $i < j$ such that $d(T^i x, T^j x) \le 2^{-k}$ but for $i \le i' < j' < j$, we have $d(T^{i'} x, T^{j'} x) > 2^{-k}$. Notice that such a minimal recurrence exists for all k by the pigeonhole principle.

Let $p = j - i$ and let y be the point of period p satisfying $(y)_i^{j-1} = (x)_i^{j-1}$. Since $d(T^i x, T^j x) \le 2^{-k}$ we see that $(y)_i^{j+k-1} = (x)_i^{j+k-1}$. It follows that the orbit segment $(T^i x, \dots, T^{j-1} x)$ $2^{-(k+1)}$-shadows $T^i y$.

Let $2^{-l} = \min_{i \le i' < j' < j}\{d(T^{i'} y, T^{j'} y)\}$ be the minimum distance between points in $\mathcal{O}y$ and notice that by construction of y and the ultrametric property, $2^{-l} \ge 2^{-(k-1)}$.

Define the perturbation function g by $g(t) = -d_A(t, \mathcal{O}y)$, and let $\tilde{f} = \hat{f} - \varepsilon g$.

We will now show that provided k is sufficiently large, the measure supported on $\mathcal{O}y$ is the unique optimizing measure for functions lying in a $\|\cdot\|_A$-open ball about $\tilde{f}$.

Let $Q = \{\tilde{f} + h : \|h\|_A < \varepsilon\sigma\}$ with $\sigma < 1$ to be determined later. Fix $\hat{f} - \varepsilon g + h \in Q$ and let q be its normalization, $q = \hat{f} - \varepsilon g + h + \beta$ where $\beta = -\sup_{\mu \in \mathcal{M}} \int (\hat{f} - \varepsilon g + h) \mathrm{d}\mu$.

Let γ_A be as in Theorem 18. Recall that $\gamma_A > 1$. We then have $\mathrm{V}_n \hat{f} \le \gamma_A \|f\|_A A_n$. Further, since $\varepsilon, \sigma < 1$, Fact 20 gives us $\mathrm{V}_n(\varepsilon g), \mathrm{V}_n h \le \gamma_A A_n$. Let $L = \gamma_A^2(\|f\|_A + 2)$. Since $\mathrm{V}_n(\hat{f} - \varepsilon g + h) = \mathrm{V}_n q$ we have

$$\mathrm{V}_n \hat{f}, \mathrm{V}_n \tilde{f}, \mathrm{V}_n q \le L A_n \qquad \text{and} \qquad \gamma_A \mathrm{V}_n f \le L A_n,$$

with the second inequality following from Fact 20. Further, L only depends on A and $\|f\|_A$.

Since x $2^{-(k+1)}$-shadows $\mathcal{O}y$ for p steps, we can get a good bound for β. By construction

$$\langle q \rangle (y) = \langle f \rangle (y) - \varepsilon \langle g \rangle (y) + \langle h \rangle (y) + \beta \le 0,$$

and so

$$\beta \le - \langle f \rangle (y) + \varepsilon \langle g \rangle (y) - \langle h \rangle (y) = - \langle f \rangle (y) - \langle h \rangle (y).$$

Proposition 21 gives us $\langle f \rangle (x) - \gamma_A \mathrm{V}_{k+1}(f)/p = -\gamma_A \mathrm{V}_{k+1}(f)/p \le \langle f \rangle (y)$ so that $-\langle f \rangle (y) \le L A_{k+1}/p$. Combining this with the fact that $\|h\|_\infty \le \|h\|_A < \varepsilon\sigma$ gives $\beta < L A_{k+1}/p + \varepsilon\sigma$. Since $q = \hat{f} - \varepsilon g + h + \beta$ and the first two terms are non-positive, we see that

$$\begin{aligned} h(\omega) + \beta &< \frac{L A_{k+1}}{p} + 2\varepsilon\sigma \text{ for all } \omega \in \Omega; \text{ and} \\ q(\omega) &< \frac{L A_{k+1}}{p} + 2\varepsilon\sigma \text{ for all } \omega \in \Omega. \end{aligned} \tag{3}$$

Let $q^{(n)}$ be the co-cycle $q^{(n)}(z) = q(T^{n-1}z) + q(T^{n-2}z) + \cdots + q(z)$, and note that if $n > m$, $q^{(n)}(z) - q^{(m)}(z) = q^{(n-m)}(T^m z)$.

We know by Proposition 17 that there exists q^*, a fixed point of Φ_q. Let $z \in \Omega$ be arbitrary. We know there exists some symbol a_1 such that $q^*(z) = q(a_1 z) + q^*(a_1 z)$. Iterating this process, we may find an infinite sequence of preimages (a_i) such that for any $n > 0$,

$$\begin{aligned} q^*(z) &= q(a_1 z) + q(a_2 a_1 z) + \cdots + q(a_n \cdots a_1 z) + q^*(a_n \cdots a_1 z) \\ &= q^{(n)}(a_n \cdots a_1 z) + q^*(a_n \cdots a_1 z). \end{aligned} \tag{4}$$

Fix any such preimage infinite sequence (a_i). We will now identify a (possibly finite) sequence of times, (t_n), by the following recursive procedure: For a time t, define $\omega_t = a_t a_{t-1} \cdots a_1 z$. Let t_0 be the smallest number (if it exists) such that $d(\omega_{t_0}, \mathcal{O}y) > 2^{-(k+1)}$. Given t_n, let $t_{n+1} > t_n$ be the next smallest number (again, if it exists) so that $d(\omega_{t_{n+1}}, \mathcal{O}y) > 2^{-(k+1)}$. Our goal is to show that the length of the sequence is finite. From this it follows that the preimages ω_t accumulate to $\mathcal{O}y$. It will then follow that the periodic orbit measure supported on $\mathcal{O}y$ is the unique maximizing measure.

Since $2^{-l} \geq 2^{-(k-1)}$ (and so $2^{-l}/4 \geq 2^{-(k+1)}$), for times strictly between t_n and t_{n-1}, the In Order Lemma (Lemma 14) gives that we $2^{-(k+1)}$-shadow $\mathcal{O}y$.

Suppose $t_n - t_{n-1} > 1$ and let $y' \in \mathcal{O}y$ be the point that is $2^{-(k+1)}$-shadowed by ω_{t_n} for $t_n - t_{n-1} - 1$ steps (that is $d(T^i \omega_{t_n}, T^i y) \leq 2^{-(k+1)}$ for $0 < i < t_n - t_{n-1}$). Summing along this segment, the Parallel Orbit Lemma (Lemma 16) gives us

$$\sum_{0<i<t_n - t_{n-1}} \left[q(T^i \omega_{t_n}) - q(T^i y') \right] \leq \mathrm{V}_{k+1}(q) \leq L A_{k+1}.$$

so that

$$\sum_{0<i<t_n - t_{n-1}} q(T^i \omega_{t_n}) \leq L A_{k+1} + \sum_{0<i<t_n - t_{n-1}} q(T^i y')$$

Grouping $\sum_{0<i<t_n - t_{n-1}} q(T^i y')$ in blocks of length p together with at most $p-1$ singleton terms and using (3), we see

$$\sum_{0<i<t_n - t_{n-1}} q(T^i \omega_{t_n}) \leq L A_{k+1} + mp \left\langle q \right\rangle (y) + (p-1) \left(L A_{k+1}/p + 2\varepsilon\sigma \right),$$

where m is the integer part of $(t_n - t_{n-1} - 1)/p$. Since $\langle q \rangle (y) \leq 0$, we simplify to get

$$\sum_{0<i<t_n - t_{n-1}} q(T^i \omega_{t_n}) \leq 2 L A_{k+1} + 2(p-1)\varepsilon\sigma. \tag{5}$$

Notice that this equation holds also (trivially) if $t_n = t_{n-1} + 1$. We now evaluate $q(\omega_{t_n})$:

$$q(\omega_{t_n}) = \hat{f}(\omega_{t_n}) - \varepsilon g(\omega_{t_n}) + h(\omega_{t_n}) + \beta.$$

By construction we have $d(\omega_{t_n}, \mathcal{O}y) \geq 2^{-k}$ so that $g(\omega_{t_n}) \geq A_k$. Using (3) again and the fact that $\hat{f} \leq 0$ we have

$$q(\omega_{t_n}) \leq -\varepsilon A_k + \frac{L A_{k+1}}{p} + 2\varepsilon\sigma. \tag{6}$$

Combining equations (5) and (6) we get

$$q^{(t_n - t_{n-1})}(\omega_{t_n}) \leq -\varepsilon A_k + 3 L A_{k+1} + 2p\varepsilon\sigma,$$

and so for $\sigma \leq A_k/(4p)$ we have

$$q^{(t_n - t_{n-1})}(\omega_{t_n}) \leq -\frac{\varepsilon}{2} A_k + 3LA_{k+1}.$$

Since L only depends on (A_n) and $\|f\|_A$, our assumption that $A_{k+1}/A_k \to 0$ ensures that there exists a k such that $\alpha = \frac{\varepsilon}{2}A_k - 3LA_{k+1} > 0$. Fix this k and fix $\sigma = A_k/(4p)$. Let $(x)_i^{j-1}$ be the minimal recurrence segment identified in the proof and y be the corresponding periodic orbit. This fixes the open ball Q whose centre is at a distance ε from $\hat{f}$.

We have shown that for any function in Q, its normalized version q satisfies $q^{(t_i - t_{i-1})}(\omega_{t_i}) < -\alpha$. Expanding using (4) now gives

$$q^*(\omega_{t_0}) - q^*(\omega_{t_n}) = q^{(t_n - t_0)}(\omega_{t_n}) = \sum_{i=1}^{n} q^{(t_i - t_{i-1})}(\omega_{t_i}) \leq -n\alpha.$$

But q^* is a bounded function and so the number of terms in the sequence (t_n) is finite.

Since z was chosen arbitrarily, this is sufficient to show the periodic orbit measure supported on $\mathcal{O}y$ uniquely optimizes q. If not, then there would be points z and preimage sequences (a_i) satisfying (4) that do not eventually follow $\mathcal{O}y$, and so (t_n) would be infinite. ■

Theorem 2 proves both (a) that a function optimized by an aperiodic point can be perturbed to be optimized by a periodic point and (b) that a function optimized by periodic point can be perturbed to lie in an open set of functions optimized by the same periodic point. Following the methods of Yuan and Hunt in [9], one can prove (b) in the general context of A-norm spaces (dropping the assumption that $A_{n+1}/A_n \to 0$ entirely).

Acknowledgments. We would like to thank the referee for a careful reading and very useful suggestions.

References

[1] T. Bousch. Le poisson n'a pas d'arêtes. *Ann. Inst. Henri Poincaré, Probabilités et Statistiques*, 36:197–224, 2000.

[2] T. Bousch. La condition de Walters. *Ann. scientifiques de l'Ecole Normale Supérieure*, 34:287–311, 2001.

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

[4] B. R. Hunt and E. Ott. Optimal periodic orbits of chaotic systems. *Phys. Rev. Lett.*, 76(13):2254–2257, Mar 1996.

[5] B. R. Hunt and E. Ott. Optimal periodic orbits of chaotic systems occur at low period. *Phys. Rev. E*, 54(1):328–337, Jul 1996.

[6] O. Jenkinson. Ergodic optimization. *Discrete and Continuous Dynamical Systems*, pages 197–224, 2006.

[7] P. Walters. Invariant measures and equilibrium states for some mappings which expand distances. *Transactions of the AMS*, 236:127–153, 1985.

[8] T.-H. Yang, B. R. Hunt, and E. Ott. Optimal periodic orbits of continuous time chaotic systems. *Phys. Rev. E*, 62(2):1950–1959, Aug 2000.

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

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

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

Department of Mathematics and Statistics, University of Victoria, Victoria BC, Canada V8W 3R4