Ergodic Optimization in the Shift

by

Jason Siefken
H.B.Sc., Oregon State University, 2008

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Mathematics and Statistics

Ergodic Optimization in the Shift

by

Jason Siefken
H.B.Sc., Oregon State University, 2008

Supervisory Committee

Dr. Anthony Quas, Supervisor
(Department of Mathematics and Statistics)

Dr. Christopher Bose, Departmental Member
(Department of Mathematics and Statistics)

Dr. Frank Ruskey, Outside Member
(Department of Computer Science)

Supervisory Committee

Dr. Anthony Quas, Supervisor
(Department of Mathematics and Statistics)

Dr. Christopher Bose, Departmental Member
(Department of Mathematics and Statistics)

Dr. Frank Ruskey, Outside Member
(Department of Computer Science)

ABSTRACT

Ergodic optimization is the study of which ergodic measures maximize the integral of a particular function. For sufficiently regular functions, e.g. Lipschitz/Hölder continuous functions, it is conjectured that the set of functions optimized by measures supported on a periodic orbit is dense. Yuan and Hunt made great progress towards showing this for Lipschitz functions. This thesis presents clear proofs of Yuan and Hunt's theorems in the case of the Shift as well as introducing a subset of Lipschitz functions, the super-continuous functions, where the set of functions optimized by measures supported on a periodic orbit is open and dense.

Contents

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

Dr. Anthony Quas, for mentoring, support, encouragement, patience, and most of all, optimism.

The University of Victoria, for providing funding and a house of learning.

Paul Struass, Kenny Barrese, Max Brugger, Chee Sing Lee, Dr. Bob Burton, for keeping the graduate school spirit, along with my spirit, alive and in good condition.

DEDICATION

To Western Culture

—Without your bounty on reason, no mathematician would have a job.

蝸牛 そろそろ登れ
富士の山
一茶小林

Chapter 1

Introduction

Ergodic optimization is the study of measures that maximize the integral of a particular function. For example, if one wishes to integrate a bell curve tightly focused about zero, a measure that puts more weight around zero will produce a larger integral than measures that put weight on points far from zero, where the bell curve becomes exponentially small. If one is allowed to choose an arbitrary measure, then putting all the mass of the measure at the maximum of the function trivially achieves an optimum. But, this case is not very interesting or applicable, so ergodic optimization places some restrictions on the types of measures we optimize over, namely they must be invariant probability measures.

With this restriction applied, a curious phenomenon is observed in experiments and occasionally demonstrated by proof: most functions tend to be optimized by measures supported on a periodic orbit. These measures, which are also referred to as periodic orbit measures, are the simplest of all possible measures and form the basis for the question targeted in this thesis: *For what spaces can we say that the set of functions optimized by periodic orbit measures is "large."*

We take a "large" set to be one that contains an open and dense subset. Though it is known that in the fully general space of continuous functions, the set of functions optimized by periodic orbit measures does not contain an open, dense subset[5], there are a handful of results showing spaces where the set of functions optimized by periodic orbits is open and dense. Bousch has shown in [2] that the set of Walters functions satisfy this property, and Contreras, Lopes, and Thieullen show in [4] that a curious union of Hölder spaces satisfies this property, though the space they show it in is not a Banach space and the norm used is "outside" the space studied. Yuan and Hunt attempted to produce a similar result for Lipschitz functions in [13] but only managed to show that no Lipschitz function can be stably optimized by a measure that is not supported on a periodic orbit.

This thesis works to clearly prove and explain Yuan and Hunt's results about the optimization of Lipschitz functions as well as present a subspace of Lipschitz functions, the super-continuous functions, where the "most functions are optimized by a periodic orbit measure" conjecture holds true. This is all done in the context of the Shift on doubly-infinite sequences, which is the premier object of study in symbolic dynamics.

In order to talk about ergodic optimization in the Shift, we must first have a handle on what the Shift space is and what it means to be an ergodic measure on the Shift space. Further, it is important to understand how the Ergodic Decomposition Theorem shows that results about optimizing ergodic measures are general in the sense that results about invariant measures may be decomposed into results about ergodic measures.

1.1 Dynamical Systems

A *dynamical system* is simply a space X with an associated transformation $T : X \to X$. If T is continuous, then (X, T) is called a *continuous dynamical system*. This transformation is iterable, so for any point $x \in X$, there is a sequence of points $x, Tx, T^2x, \ldots$ that may or may not be distinct. Applying the transformation T is viewed as incrementing time. That is, x is referred to as the point x at time 0; $T^{50}x$ is referred to as the point x at time 50. Though in many dynamical systems it often makes sense to talk about x at time 1.4 (i.e. $T^{1.4}x$), this thesis only concerns itself with integral time intervals.

In order to discuss the dynamics of a dynamical system (X, T), it is useful to have some notion of distance between points in X. We will only be using distance functions that satisfy the conditions of a metric.

Definition 1.1. *A* metric *on a space X is defined to be a real-valued binary function d that satisfies*

1. $d(x, y) = d(y, x)$
2. $d(x, y) \geq 0$
3. $d(x, y) = 0$ *if and only if* $x = y$

4. $d(x,z) \le d(x,y) + d(y,z)$

for all $x, y, z \in X$*. The properties are referred to as: symmetry, positivity, definiteness, and the triangle inequality. A metric is called an* ultrametric *if it satisfies a stronger version of the triangle inequality, namely*

$$d(x,z) \le \max\{d(x,y), d(y,z)\}.$$

A metric gives us some way to define the distance between points, and henceforth we will assume all spaces come equipped with a metric and are therefore *metric spaces*. A metric on a space is particularly useful when it allows points to be approximated by a sequence of other points.

Definition 1.2. *If* (X,d) *is a metric space, a set* $D \subset X$ *is* dense *in* X *if for each point* $x \in X$*, there exists a sequence* $\{x_i\} \subset D$ *such that* $x_i \to x$ *in the metric* d*.*

Definition 1.3. *A metric space* (X,d) *is* separable *if there exists* $D \subset X$ *that is countable and dense.*

Separable spaces are often the most useful spaces since by making lists, the only infinity we can approximate is countable infinity. Separability says this is good enough. From only a countable number of choices, one can pick a sequence that gets arbitrarily close to any point.

We will soon define the particular dynamical system that this thesis concerns itself with, namely the Shift. However, we first need to define some vocabulary which is applicable to all dynamical systems.

Definition 1.4. *The* orbit *of a point* x*, denoted* $\mathcal{O}x$*, is the set of all places* x *goes under* T*. That is,*

$$\mathcal{O}x = \{T^i x\}$$

for i *ranging over* $\mathbb{Z}$ *if* T *is invertible, and* i *ranging over* $\mathbb{N}$ *otherwise.*

Points in X may be classified by the size of their orbits. If $|\mathcal{O}x|$ is infinite, then x is said to be aperiodic. If $|\mathcal{O}x|$ is finite, x is said to be pre-periodic.

Definition 1.5 (Pre-periodic)**.** *A point* x *is said to be* pre-periodic with period p *if* $T^{p+i}x = T^i x$ *for some* i*. If* $i = 0$*,* x *is also periodic. We call* p *the minimal period of* x *if* $p = |\mathcal{O}T^i x|$*.*

Definition 1.6 (Periodic)**.** *A point x is said to be* periodic with period p *if $T^p x = x$. We call p the minimal period of x if $p = |\mathcal{O}x|$.*

If T is invertible, all pre-periodic points are periodic points. If T is not invertible, the pre-periodic points are the points that eventually become periodic.

Definition 1.7. *A set A is called* invariant with respect to T *if*

$$T^{-1}A = A,$$

where $T^{-1}A = \{x : Tx \in A\}$ is the inverse image of A under T.

We consider the inverse image of A rather than the forward image of A to accommodate cases when T is not invertible.

Definition 1.8. *If (X, T) is a dynamical system, a closed, invariant subset $M \subset X$ is called* minimal *if $\mathcal{O}y$ is dense in M for all $y \in M$.*

Definition 1.9. *$S \subset \mathcal{O}x$ is called a* segment *of $\mathcal{O}x$ if*

$$S = \{T^i x, T^{i+1}x, \ldots, T^{i+n}x\}$$

for some i, n.

We may say that the point y stays ε-*close* to some segment S for p steps. This means that for all $0 \le i < p$, $d(T^i x, S) \le \varepsilon$.

Definition 1.10 (Shadowing)**.** *If x, y are points, we say that x* ε-shadows *$\mathcal{O}y$ for length p if there exists an m such that*

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

for all $0 \le i < p$.

We may use the term *shadowing* somewhat more liberally than this definition, saying things such as x ε-shadows $\mathcal{O}y$ when we really mean there is some j so $T^j x$ ε-shadows $\mathcal{O}y$. The important distinction between ε-shadowing and simply staying ε-close is that shadowing $\mathcal{O}y$ implies that you follow $\mathcal{O}y$ in order.

1.1.1 The Shift

The Shift is a specific dynamical system that is at once very general and provides enough simplifying assumptions to make proofs easier (for example, almost all associated constants are either 1 or 2). The Shift is defined on a particularly tractable space: the space of sequences.

Definition 1.11. *If $\mathcal{A}$ is some finite alphabet, then the set of* two-sided sequences*(doubly-infinite sequences), $\Omega = \mathcal{A}^{\mathbb{Z}}$ is defined to be the set of all elements of the form*

$$\cdots a_{-2}a_{-1}.a_0a_1a_2a_3\cdots,$$

where $a_i \in \mathcal{A}$.

A radix point is used as an anchor point to give some way to discuss why, for example, if $\mathcal{A} = \{0, 1, 2\}$, $\cdots 0000.1222\cdots$ is different from $\cdots 00001.222\cdots$.

Notice that the decimal representation of a real number is a subset of two-sided sequences on $\mathcal{A} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. However, the standard metric on $\mathbb{R}$ is not a definite metric on $\Omega = \mathcal{A}^{\mathbb{Z}}$ because $d(1.0\cdots, .99\cdots) = 0$, but $1.0\cdots$ and $.99\cdots$ are different as sequences (note that $a_{-1}.a_1a_2\cdots$ actually means $\cdots 00a_{-1}.a_1a_2\cdots$). Thus, a metric on Ω must be defined slightly differently.

To simplify discussion of sequence space, we will introduce some notation.

Notation 1.12. *If $a \in \Omega$, the space of two-sided sequences of some alphabet $\mathcal{A}$, then a_i refers to the ith symbol in the standard representation*

$$a = \cdots a_{-2}a_{-1}.a_0a_1a_2a_3\cdots.$$

Further, $(a)_i^j$ refers to the ordered list of symbols $a_i, \cdots, a_j$. That is,

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

Another useful concept when talking about sequence spaces is the notion of cylinder sets.

Definition 1.13 (Cylinder Set)**.** *For each finite subset $Z = \{z_0, z_1, \ldots, z_n\} \subset \mathbb{Z}$, a* cylinder set *fixed on Z is the collection of all points B such that for $b \in B$,*

$$b_{z_i} = a_i$$

for some fixed choice of symbols $\{a_i\}$.

In intuitive terms, cylinder sets are just collections of points where a finite number of positions have fixed values. It is often useful to use $[\cdot]$ to denote cylinder sets.

Notation 1.14. *For* $\{a_i\}$ *symbols in some alphabet,* $A = [a_{-m}a_{-m+1}\cdots a_{-1}.a_0a_1\cdots a_n]$ *is the the cylinder set that satisfies*

$$A = \big\{x : (x)_{-m}^{n} = (a_{-m}, a_{-m+1}, \cdots, a_{-1}, a_0, a_1, \cdots, a_n)\big\}.$$

"$*$" may be used to represent a wild card, for example $[.1*1]$ *is the set of all sequences whose first and third digit to the right of the radix point is a 1. Further, if the radix point is omitted, it is assumed the cylinder set has been specified to start from the radix point. That is,* $[110] = [.110]$.

Cylinder sets are easy to write down and allow for easier computation than arbitrary sets. And, as we will see in the next section, it is often sufficient to prove results on cylinder sets that carry over to all measurable sets.

Definition 1.15 (Complexity)**.** *Let* C_n *be the collection of cylinder sets of length* n *and for a point* x*, let*

$$f_x(n) = \#\{C \in C_n : T^i x \in C \text{ for some } i\}.$$

The complexity *of* x *is defined as the function*

$$\sigma(x) = f_x.$$

In simple terms, $\sigma(x)$ counts the number of different subwords of length n that occur in x. For example, if x's binary digits were determined uniformly at random, $\sigma(x) = 2^n$. If x were a periodic point with period p, then $\sigma(x) = p$.

Throughout this thesis, we will use d to represent the *standard metric on sequences.*

Definition 1.16. *If* Ω *is the space of two-sided sequences of some alphabet* $\mathcal{A}$*, then for* $a, b \in \Omega$*, between*

$$d(a,b) = \sup_k\{2^{-k} : (a)_{-k}^{k} \neq (b)_{-k}^{k}\}.$$

That is, $d(a,b) = 2^{-k}$ where k is the number of places away from the zero symbol where the first disagreement between a and b occurs (and if no disagreement occurs, then the convention that $2^{-\infty} = 0$ maintains the validity of this interpretation).

Notice that in our previous example, with the standard metric, $d(1.0\cdots, .99\cdots) = 1$, so indeed these two points, which are identical as real numbers, are very far apart as sequences. It is worth noting that the furthest two sequences may be from each other is 1.

We are now ready to define the central dynamical system of this thesis: the Shift.

Definition 1.17. *The* Shift *is a dynamical system (Ω, T) where Ω is the set of two-sided sequences on some alphabet $\mathcal{A}$ and T is the transformation that moves the sequence one position to the left (equivalently, moves the radix point one position to the right). That is*

$$T(\cdots a_{-2}a_{-1}.a_0a_1a_2a_3\cdots) = \cdots a_{-2}a_{-1}a_0.a_1a_2a_3\cdots.$$

This dynamical system is called the Shift because T "shifts" the radix point one to the right.

Fact 1.18 (Expansivity)**.** *It is a direct consequence of the definition of the standard metric on sequences and the definition of T that for any points x, y,*

$$\frac{1}{2}d(x,y) \leq d(Tx,Ty) \leq 2d(x,y).$$

1.2 Invariant Measures

In order to integrate, in the Riemann case, one must first understand how to find the area of a rectangle (or rectangular prism). In the two dimensional Euclidean case, this is simply width times height. However, in general spaces (like the Shift space, for instance) it is unclear what "width" means. This question of "width" is solved by introducing measures and generalized rectangles. A generalized rectangle is the analog of a rectangle without the need for the notion of a line segment. That is, a generalized rectangle may be thought of as function that takes the value h on some set D and the value 0 otherwise. The "area" of this generalized rectangle will unsurprisingly be the measure of the set D times h, but for that, we need the concept of measure.

Definition 1.19. *A collection of sets $\mathscr{A}$ is called a* σ-algebra *if*

1. *$\mathscr{A}$ is nonempty*
2. *for all $A \in \mathscr{A}$, $A^C \in \mathscr{A}$*
3. *and, if $\{A_n\} \subset \mathscr{A}$ is a countable collection of sets, then $\bigcup_{i \in \mathbb{N}} A_i \in \mathscr{A}$.*

Definition 1.20. *Given some topological space X, the* Borel σ-algebra *of X is the smallest σ-algebra that contains all the open sets in X. A* Borel set *is a member of the Borel σ-algebra.*

Definition 1.21. *A* measure *is a function $\mu : \mathscr{B} \to [0, \infty]$ defined on some σ-algebra of subsets $\mathscr{B} \subset \wp(\Omega)$ (where $\wp(\Omega)$ denotes the power set of Ω) that satisfies*

1. *$\mu(\emptyset) = 0$,*
2. *and, if A_i is a countable collection of disjoint sets,*

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i).$$

A probability measure *is a measure μ with the added assumption that $\mu(\Omega) = 1$.*

Definition 1.22. *A* Borel measure *on X is a measure defined on the Borel subsets of X.*

Definition 1.23. *If μ is a measure, the set A is called* measurable with respect to μ *if A is in the σ-algebra on which μ is defined.*

Definition 1.24. *If μ and ν are two measures defined on the same σ-algebra, μ and ν are said to be* mutually singular *if there exists two disjoint sets A, B such that $A \cup B$ is the whole space and $\mu(R) = 0$ for all measurable subsets $R \subset A$ and $\nu(S) = 0$ for all measurable subsets $S \subset B$.*

Detailed explanations of these concepts may be found in [9]. For our purposes, $\mathscr{B}$ will always be the Borel subsets of Ω.

Once the concept of measure is established, it is a straightforward process to define the integral of real valued functions on Ω in terms of generalized rectangles/characteristic functions. A full description of this process may be found in any introductory analysis textbook such as [9].

To define a measure, one needs to assign a real number to every set in $\mathscr{B}$. Since $\mathscr{B}$ is most often uncountable, this can be quite intimidating. Fortunately, the Kolmogorov Extension Theorem will allow us to get away with only specifying μ on a much smaller object called a semi-algebra. We call such a function a pre-measure.

Definition 1.25. *If $\mathcal{X}$ is a collection of sets that contains the empty set, a function $\mu : \mathcal{X} \to [0, \infty]$ is called a* pre-measure *if*

1. $\mu(\emptyset) = 0$
2. *If $\{A_i\} \subset \mathcal{X}$ is a countable collection of disjoint sets and $\bigcup A_i \in \mathcal{X}$, then*

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i).$$

Notice that a pre-measure is essentially a measure, but instead of being defined on a σ-algebra, it is defined on an arbitrary collection of sets that contains the empty set.

Definition 1.26. *A* semi-algebra $\mathscr{A} \subset \wp(\Omega)$ *is a collection of sets that satisfies*

1. $\emptyset \in \mathscr{A}$,
2. $\mathscr{A}$ *is closed under finite intersections,*
3. *and if $A \in \mathscr{A}$, then A^C may be written as a finite union of disjoint elements of $\mathscr{A}$.*

Semi-algebras are often countable (and so usually smaller than a σ-algebra), and their usefulness comes from the following theorem.

Theorem 1.27 (Kolmogorov Extension Theorem)**.** *If μ is a probability pre-measure defined on a semi-algebra $\mathscr{A}$, then μ uniquely extends to a probability measure on $\mathscr{B}$, the σ-algebra generated by $\mathscr{A}$.*

A proof of the Kolmogorov Extension Theorem using Dynkin's π–λ lemma may be found in [3].

Although the concept of measures is a very general one, dynamics is concerned only with specific types of measures—namely those measures that have something to do with the transformation. The most interesting subset of these measures are the *invariant measures.*

Definition 1.28. *For a dynamical system* (Ω, T)*, a measure* μ *is said to be* invariant with respect to T *if for any measurable set* A*,*

$$\mu(T^{-1}A) = \mu(A),$$

where $T^{-1}A$*, the* inverse image *of* A*, is defined as* $T^{-1}A = \{x : Tx \in A\}$.

It may seem strange that we are considering inverse images in our definition of an invariant measure, but this is indeed the proper formulation. For an invertible transformation T, it makes no difference whether we demand that $\mu(T^{-1}A) = \mu(A)$ or $\mu(TA) = \mu(A)$, however for non-invertible transformations, insisting that $\mu(TA) = \mu(A)$ restricts the class of invariant measures too much (see Example 1.29).

1.2.1 Examples

We are now equipped to describe many different classes of invariant measures.

Example 1.29 (Bernoulli Measures)**.** Let $\Omega = \{0,1\}^{\mathbb{N}}$, or all one-sided sequences of zeros and ones. Let $T : \Omega \to \Omega$ be the shift by one. That is, given $(a_0, a_1, a_2, \ldots) \in \Omega$,

$$T(a_0, a_1, a_2, \ldots) = (a_1, a_2, a_3, \ldots).$$

The *Bernoulli measures* on Ω are defined by picking a probability vector $p = (p_0, p_1)$ with $p_0 + p_1 = 1$, then defining a pre-measure μ' on cylinder sets by

$$\mu'([a_0, a_1, \ldots, a_n]) = \prod_{i=0}^{n} p_{a_i},$$

with the familiar property that $\mu'(A \cup B) = \mu'(A) + \mu'(B)$ on disjoint cylinder sets A and B.

It is straight-forward to show that if A is a cylinder set, $\mu'(A) = \mu'(T^{-1}A)$.

Proof. Let $A = [a_0, a_1, \ldots, a_n]$ be an arbitrary cylinder set. $T^{-1}(A) = [0, a_0, a_1, \ldots, a_n] \cup [1, a_0, a_1, \ldots, a_n]$. $T^{-1}(A)$ is the union of two disjoint cylinder sets, so by definition,

$$\mu'(T^{-1}(A)) = \mu'([0, a_0, a_1, \ldots, a_n]) + \mu'([1, a_0, a_1, \ldots, a_n])$$

$$= p_0 p_{a_0} \cdots p_{a_n} + p_1 p_{a_0} \cdots p_{a_n} = (p_0 + p_1) p_{a_0} \cdots p_{a_n}.$$

But, p is a probability vector, so $p_0 + p_1 = 1$, leaving

$$\mu'(T^{-1}(A)) = p_{a_0} \cdots p_{a_n} = \mu'(A).$$

□

Since μ' is a pre-measure defined on the semi-algebra of cylinder sets, by Theorem 1.27, μ' uniquely extends to a measure μ on all Borel sets. We must now show that μ itself is invariant.

Let $\nu = \mu \circ T^{-1}$. It is clear that because inverse images preserve disjointness, ν is a measure. Further, for any cylinder set A, $\mu(A) = \nu(A)$, and so ν is an extension of μ'. Since Theorem 1.27 gives us that such extensions are unique, $\mu = \nu$ and so for an arbitrary Borel set B,

$$\mu(B) = \nu(B) = \mu(T^{-1}B),$$

and so μ is invariant.

If we take the specific case of $p_0 = p_1 = 1/2$ and consider $\{0,1\}^{\mathbb{N}}$ to be the binary expansion of points in the unit interval, μ corresponds to Lebesgue measure.

Taking different values of p_0, p_1, we get very different looking measures. In fact, for any distinct values p_0, p_1 and p'_0, p'_1, the resulting Bernoulli measures are mutually singular.

If we interpret sequences of 0's and 1's to be binary representations of points in the unit interval, Figure 1.1 shows the relative measure of cylinder sets of length ten with $p_0 = 0.4$ and $p_1 = 0.6$ (relative to the $0.5, 0.5$ Bernoulli measure). The $0.4, 0.6$ Bernoulli measure of a set may be approximated by integrating against the function in Figure 1.1 with respect to $0.5, 0.5$ Bernoulli measure. Shaded in gray is the area below 1 and also below the curve. It is interesting to note that if one continues plotting for finer and finer cylinder sets (length 11, 12, etc.), the area in gray will tend towards zero. However, the whole space still integrates to one, so the peaks of the function (at 1, 1/2, 3/4, etc.) will become infinitely tall to compensate. This makes it clear that the Bernoulli measures (except for $p_0 = p_1 = 1/2$) are singular with respect to Lebesgue measure.

Example 1.30 (Periodic Orbit Measures)**.** As in Example 1.29, let $\Omega = \{0,1\}^{\mathbb{N}}$ be the set of one-sided sequences and let T be the one-sided shift. The function δ_x is a

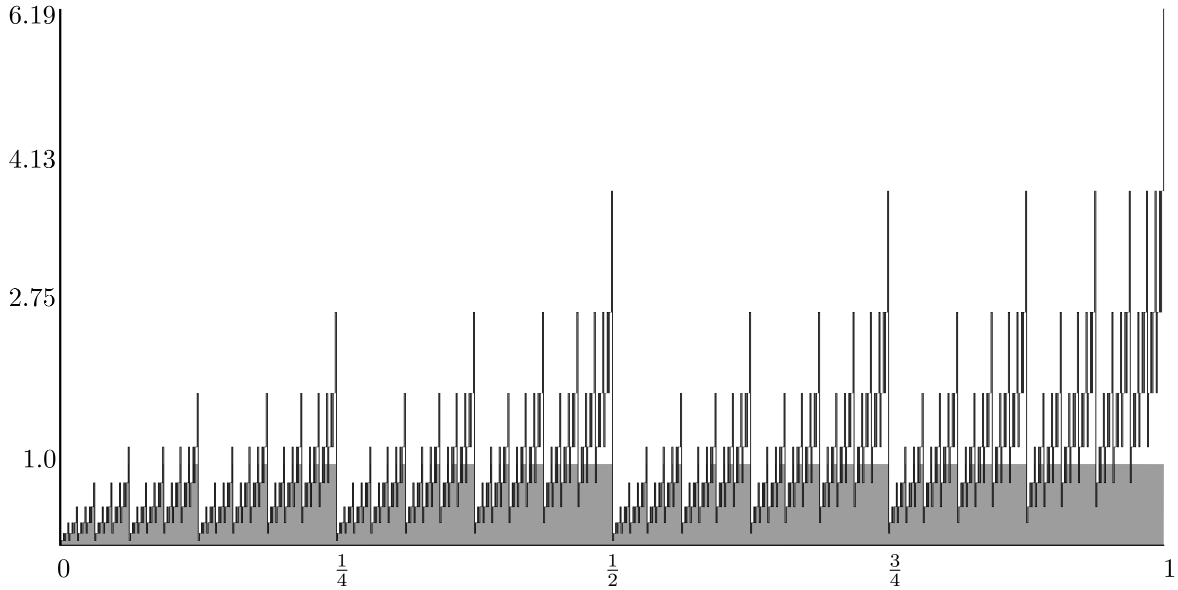

Figure 1.1: $p_0 = 0.4$, $p_1 = 0.6$ Bernoulli measure of cylinder sets of length 10 interpreted as subsets of $[0, 1]$.

function on sets that indicates whether x is a member of the set. That is

$$\delta_x A = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ else} \end{cases} .$$

Given an n-periodic element $a \in \Omega$ (i.e. $T^n a = a$), we may define an invariant measure supported on $\{T^i a\}$ by

$$\mu(A) = \frac{1}{n}\Big(\delta_a A + \delta_{Ta} A + \delta_{T^2 a} A + \cdots + \delta_{T^{n-1} a} A\Big).$$

It is easy to see that μ is invariant with respect to T.

Proof. Let $a = (a_1, a_2, \ldots, a_n, a_1, a_2, \ldots)$ be a periodic sequence and μ be a periodic orbit measure supported on $\{T^i a\}$. It is clear that

$$\mu(\{a\}) = \frac{1}{n}.$$

Computing $T^{-1}\{a\}$, we see $T^{-1}\{a\} = \{(1, a_1, a_2, \ldots), (0, a_1, a_2, \ldots)\}$. Because a is periodic, there exists a unique a_0 such that $(a_0, a_1, \ldots) \in \mathcal{O}a$. Thus, only one element

$a' \in T^{-1}\{a\}$ has the property that $a' = T^k a$ for some k. Therefore,

$$\mu(T^{-1}\{a\}) = \frac{1}{n} = \mu(\{a\}). \tag{1.1}$$

If $b \notin T^i\{a\}$, then $T^{-1}\{b\} \cap T^i\{a\} = \emptyset$, so

$$\mu(T^{-1}\{b\}) = 0 = \mu(\{b\}). \tag{1.2}$$

Any set $A \subset \Omega$ may be partitioned into $W = A \backslash \mathcal{O}a$ and $Y = A \cap \mathcal{O}x$. W, by construction, contains no points in the orbit of a, so by (1.2), $\mu(W) = \mu(T^{-1}W) = 0$. Y consists of $k \leq n$ distinct points in the orbit of a. For any two distinct points $p, q \in Y$, we know $T^{-1}\{p\} \cap T^{-1}\{q\} = \emptyset$. This, combined with (1.1) gives us $\mu(Y) = \mu(T^{-1}Y) = k/n$.

Because μ is additive, we get that

$$\mu(T^{-1}A) = \mu(T^{-1}W) + \mu(T^{-1}Y) = \mu(A).$$

□

Note that periodic orbit measures are defined in the same way in the case of the two-sided shift, but since the two-sided shift is invertible, invariance of periodic orbit measures becomes a trivial consequence.

Example 1.31 (Sturmian Measures)**.** As in Example 1.29, let $\Omega = \{0,1\}^{\mathbb{N}}$ be the set of one-sided sequences and T be the one-sided shift. A Basic Sturmian sequence on the alphabet $\{0,1\}$ may be defined for any rotation number $\gamma \in (0,1)$. Given a starting position $x \in [0,1)$, a Basic Sturmian sequence $\{s_n\}_{n=0}^{\infty}$ is defined as,

$$s_n(x) = \lfloor x + (n+1)\gamma \rfloor - \lfloor x + n\gamma \rfloor.$$

An equivalent conceptualization is to graph a line of slope γ and time-zero intercept of x on a regular grid. For every horizontal unit moved in the grid where a horizontal line is passed, put a 1. If no horizontal line is passed put a 0 (See Figure 1.2).

For a fixed irrational γ, every value of x produces a unique Basic Sturmian sequence. The set of Sturmian sequences is the closure of the set of Basic Sturmian sequences. A Sturmian sequence $\{s_i(x)\}$ has the lowest possible complexity for non-

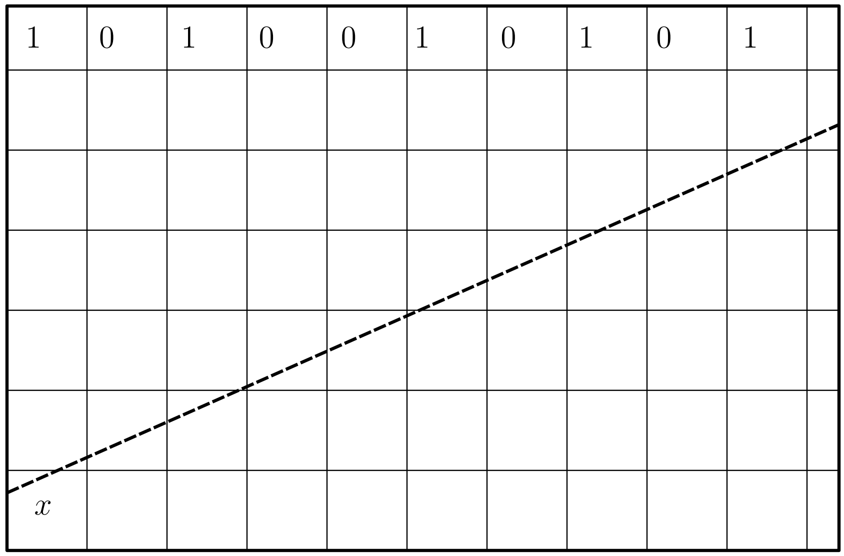

Figure 1.2: The Sturmian sequence produced by a line with slope γ and y-intercept x.

periodic sequences, with $\sigma(\{s_i(x)\}) = n + 1$ (see Definition 1.15) [5].

Sturmian sequences provide another way to define an invariant measure with respect to the shift transformation. A presentation of Sturmian measures in terms of the doubling map on $[0, 1)$ may be found in [6], but since it is nicer to work with sequence space, we may define Sturmian measures slightly differently. Let h be the map that takes a point to its Sturmian sequence in $\{0, 1\}^{\mathbb{N}}$. That is,

$$h(x) = (s_0(x), s_1(x), \ldots).$$

The Sturmian measure ς_γ corresponding to γ, may be defined as the push-forward of Lebesgue measure on $[0, 1)$ under h. That is $\varsigma_\gamma(A) = m(h^{-1}A)$, where m is Lebesgue measure.

Let $R_\gamma(x) = x + \gamma \mod 1$ be a rotation by γ on the unit circle. We now have the relation $T \circ h = h \circ R_\gamma$. Consider

$$m \circ h^{-1}(T^{-1}A) = m(R_\gamma^{-1} \circ h^{-1}A) = m(h^{-1}A),$$

and so $m \circ h^{-1}$ is T-invariant, but this is precisely the definition of σ_γ.

Example 1.32 (Combinations of Measures)**.** If μ_1,μ_2 are invariant measures, then

$$\mu = \alpha\mu_1 + \beta\mu_2$$

is an invariant measure for $\alpha, \beta \in \mathbb{R}^+$ (positiveness of α and β is needed only to ensure that μ satisfies the positiveness property of a measure). Further, if μ_1, μ_2 are invariant probability measures and $\alpha + \beta = 1$, then μ is an invariant probability measure.

In fact, if μ is a convex combination (infinite or not) of any number of invariant probability measures, then μ is an invariant probability measure.

1.3 Ergodicity

Ergodic measures are the building blocks of invariant measures. That is, every invariant measure has a decomposition into ergodic measures. This will be seen with the introduction of the Ergodic Decomposition Theorem (1.48), but first we must examine what ergodic measures are.

Definition 1.33. *A probability measure μ is said to be* ergodic with respect to a transformation T *if T is measure-preserving with respect to μ and whenever A is an invariant set,*

$$\mu(A) = 0 \textit{ or } 1.$$

We say a process or system (Ω, T, μ) is ergodic if (Ω, T) is a dynamical system and μ is an ergodic measure with respect to T.

Being ergodic essentially means that all sets that do not have full measure (and do not have zero measure) get smeared around the space by T. Not only that, but for an ergodic system, almost all points get moved around in a way representative of the ergodic measure. This fact is not obvious, but is clarified by the Birkhoff Ergodic Theorem, a proof of which may be found in [11, p. 34]

Theorem 1.34 (Birkhoff Ergodic Theorem)**.** *If (Ω, T, μ) is an ergodic probability space and f has finite integral with respect to μ, then for almost all x (with respect to μ),*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) = \int_\Omega f \, \mathrm{d}\mu.$$

The Birkhoff Ergodic Theorem gives a powerful way to analyze ergodic measures and integration against ergodic measures by merely looking at the orbit of single points. In fact, the Birkhoff Ergodic Theorem says that the integral of a function is its average value along the orbit of almost any point. I.e., the spacial averages ($\int f \,\mathrm{d}\mu_x$) agree with the time averages ($\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f(T^i x)$).

There are also various strengthenings of ergodicity which can be used to get stronger forms of the Birkhoff Ergodic Theorem. One such is unique ergodicity.

Definition 1.35 (Uniquely Ergodic)**.** *A continuous dynamical system* (Ω, T) *is said to be* uniquely ergodic *if there is only one invariant measure.*

A uniquely ergodic (Ω, T, μ) satisfies that for any continuous f,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N-1} f(T^i x) = \int_\Omega f \,\mathrm{d}\mu$$

for all $x \in \Omega$ [11].

We will attack the problems of Ergodic Optimization by studying time averages along particular points (which will be easier than studying ergodic measures themselves). We will refer to the average value along a particular orbit so often, we require some notation.

Notation 1.36. *If* (Ω, T) *is a dynamical system and* $f : \Omega \to \mathbb{R}$ *is a function, then for any* $x \in \Omega$*,*

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

if the limit exists.

Phrased in this notation, the Birkhoff Ergodic Theorem states that if μ is an ergodic measure, then $\langle f \rangle (x) = \int f \,\mathrm{d}\mu$ for μ-almost all x.

Though the Birkhoff Ergodic Theorem applies to any μ-integrable function, we can derive many more useful results if we restrict ourselves to continuous functions. From now on, the following are standing assumptions:

1. T is a continuous transformation
2. Ω is compact
3. Functions are restricted to continuous real-valued functions of Ω

We have already seen several examples of ergodic measures (periodic orbit measures, Bernoulli measures, Sturmian measures [1]), so it might appear that invariant measures are ergodic. However, that is not always the case. Non-ergodic invariant measures are easy to come by and can be created simply by taking a non-trivial convex combination of ergodic measures (See Example 1.51). However, it turns out that this is a complete classification of non-ergodic invariant measures.

Before we formally state the Ergodic Decomposition Theorem, which loosely says that any invariant measure is a convex combination of ergodic measures, we must introduce some definitions. The trouble is that a particular invariant measure may be the combination of an uncountable number of ergodic measures. Therefore, before we state the Ergodic Decomposition Theorem, we must have a framework for describing such an uncountable combination.

Notation 1.37. *For a set Ω, $\mathcal{M}_\Omega$ denotes the set of all Borel probability measures on Ω.*

Notation 1.38. *If (Ω, T) is a dynamical system, $\mathcal{M}_T$ denotes the set of all T-invariant probability measures.*

Notice that $\mathcal{M}_T \subset \mathcal{M}_\Omega$

Our end goal is to describe an invariant measure as an integral against ergodic measures in $\mathcal{M}_T$. We will motivate the Ergodic Decomposition Theorem by investigating the topological properties of $\mathcal{M}_T$.

Definition 1.39. *If $\mathcal{M}_T$ is the set of all invariant measures on a dynamical system (Ω, T), then the* weak* *topology on $\mathcal{M}_T$ is the coarsest topology that for any continuous f, the map $\mu \mapsto \int f \,\mathrm{d}\mu$ is continuous for all $\mu \in \mathcal{M}_T$.*

Equivalently, the weak* topology on $\mathcal{M}_T$ is the topology such that

$$\mu_i \to \mu \text{ if and only if } \int f \,\mathrm{d}\mu_i \to \int f \,\mathrm{d}\mu$$

for all continuous functions f.

Interestingly, Walters [11, p. 148] gives an explicit metric on $\mathcal{M}_T$ that generates the weak* topology: for $\mu, \nu \in \mathcal{M}_T$,

$$d(\mu, \nu) = \sum_{n=0}^{\infty} \frac{|\int f_n \,\mathrm{d}\mu - \int f_n \,\mathrm{d}\nu|}{2^n \|f_n\|}$$

where $\{f_i\}$ is a fixed countable dense subset of the set of continuous functions on Ω.

We will now invoke several classical results about the structure of $\mathcal{M}_T$ equipped with the weak* topology.

Definition 1.40. *X is a* convex *subset of a real vector space if for $a, b \in X$,*

$$\alpha a + (1-\alpha)b \in X$$

for all $\alpha \in [0, 1]$.

Definition 1.41. *The* extreme points *of a convex set X are those points $x \in X$ such that if $\alpha \in (0, 1)$ and $a, b \in X$,then*

$$x = \alpha a + (1-\alpha)b,$$

implies $a = b = x$.

Fact 1.42. *If $\{X_i\}$ is an arbitrary collection of convex sets, then $\bigcap X_i$ is a convex set.*

Definition 1.43. *The* convex hull *of a set X,* hull(X)*, is the intersection of all convex sets containing X.*

By Fact 1.42, hull(X) is convex.

Theorem 1.44 (Krein-Milman)**.** *If X is a compact convex subset of a locally convex real vector space, then if E is the set of extreme points of X,*

$$X = \overline{\text{hull}(E)}.$$

That is, X is the closure of the convex hull of its extreme points.

Theorem 1.45 (Riesz Representation Theorem)**.** *Let C be the set of all continuous functions on Ω. The dual space C^* of C (i.e., the space of all bounded linear functionals on C) is isomorphic to the set of signed measures on Ω.*

For a detailed discussion of notions of convexity, local convexity, vector spaces, signed measures, and the Riesz Representation Theorem, see [9].

Theorem 1.46. *If Ω is compact, then $\mathcal{M}_T$ is a compact, convex set in the weak* topology and $\mathcal{M}_T$ is locally convex.*

Note that compactness of $\mathcal{M}_T$ follows from the fact that $\mathcal{M}_T$ is a closed subset of $\mathcal{M}_\Omega$ and when Ω is compact, $\mathcal{M}_\Omega$ is compact (This follows from an application of the Riesz Representation Theorem and Alaoglu's Theorem, noting that Borel probability measures are a closed subset of the closed unit ball in the dual space of continuous functions on Ω). We now see that Theorem 1.44 and 1.46 imply that any invariant measure is the limit of convex combinations of extreme points of $\mathcal{M}_T$. We are now almost prepared for the Ergodic Decomposition Theorem.

Theorem 1.47. *If (Ω, T) is a dynamical system with Ω compact and T continuous, then μ is an extreme point of $\mathcal{M}_T$ if and only if μ is ergodic.*

Derivations of the Theorem 1.47 may be found in [11, p. 153].

It is worth pointing out now that (Ω, T) being uniquely ergodic implies that the invariant measure μ is also ergodic (since there is only one invariant measure, it must be an extreme pont).

If $\mathcal{M}_T$ were finite-dimensional, it would be easy to see that since every point in $\mathcal{M}_T$ is a finite convex combination of its extreme points, the Ergodic Decomposition Theorem would imply that every invariant measure has a decomposition into a finite convex combination of ergodic measures. However, $\mathcal{M}_T$ is rarely finite dimensional, and so it may be impossible to write a particular measure as a finite convex combination of ergodic measures. However, Theorems 1.44 and 1.46 do say that any invariant measure may be written as a limit of convex combinations of extreme points of $\mathcal{M}_T$. In fact, by the Ergodic Decomposition Theorem, any invariant measure may be written as some infinite convex combination.

Theorem 1.48 (Ergodic Decomposition Theorem)**.** *Let $\mathcal{E} \subset \mathcal{M}_T$ denote the set of all ergodic measures with respect to T. If (Ω, T) is a dynamical system with Ω compact and T continuous, then for every $\mu \in \mathcal{M}_T$, there exists a unique probability measure ρ on $\mathcal{E}$ such that*

$$\mu = \int_{\mathcal{E}} \nu \mathrm{d}\rho(\nu).$$

From the Ergodic Decomposition Theorem, we finally see that ρ represents the "weights" for the convex combination of ergodic measures that form μ. A derivation of the Ergodic Decomposition Theorem may be found in [8].

This is now a good time to introduce empirical measures—that is, measures generated by a point.

Definition 1.49. *Given a point x, construct a sequence of measures*

$$\mu_{x,N} = \frac{1}{N}\Big(\delta_x + \delta_{Tx} + \delta_{T^2x} + \cdots + \delta_{T^{N-1}x}\Big).$$

If $\mu_{x,N_i} \to \mu$ is a convergent subsequence (in the weak* topology) of $\mu_{x,N}$, we say that μ is an empirical measure generated by x. We reserve the notation μ_x for the case where $\mu_{x,N} \to \mu_x$ converges. I.e., there is a unique empirical measure generated by x.

In the next chapter we will deal with empirical measures in more detail. However, it should be noted that an empirical measure exists for any x. Since we know that $\mathcal{M}_\Omega$ is compact for compact Ω and $\mu_{x,N} \in \mathcal{M}_\Omega$, we know that the sequence of measures $\mu_{x,N}$ (as in Definition 1.49) has a convergent subsequence and therefore there exists at least one empirical measure. But $\mathcal{M}_\Omega$ is only the set of Borel probability measures. We need to check any empirical measure is also an invariant measure. To show this, we will use a convenient fact about invariant measures: a measure μ is invariant if and only if $\int f \,\mathrm{d}\mu = \int f \circ T \,\mathrm{d}\mu$ for all continuous f [11]. Suppose now that μ is a limit point of $\mu_{x,N}$. That is, there is some subsequence so $\mu_{x,N_i} \to \mu$. It should be clear that $\mu_{Tx,N_i} \to \mu$ as well (these two sequences differ at most by $\delta_x/N_i + \delta_{T^{N_i}x}/N_i$ and $\delta_x/N_i + \delta_{T^{N_i}x}/N_i \to 0$). However, $\int f \circ T \,\mathrm{d}\mu_{x,N_i} = \int f \,\mathrm{d}\mu_{Tx,N_i}$, and so taking limits we see μ is invariant.

A priori, there is no reason to assume that a point x will produce a unique empirical measure. However, the Birkhoff Ergodic Theorem states that if you fix an f and an ergodic measure μ, then for μ-almost all x, if μ_x is the empirical measure generated by x,

$$\int f \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu_x,$$

and, in fact, $\mu_x = \mu$.

It is also worth noting that by construction,

$$\langle f \rangle (x) = \int f \,\mathrm{d}\mu_x,$$

where μ_x is the empirical measure generated by x.

1.3.1 Examples

Example 1.50 (Periodic Orbit Measures)**.** Consider (Ω, T) as defined in Example 1.30. For a periodic point x, let μ_x be the periodic orbit measure supported on x.

That is,

$$\mu_x(A) = \frac{1}{|\mathcal{O}x|}\Big(\delta_x A + \delta_{Tx} A + \delta_{T^2x} A + \cdots + \delta_{T^{|\mathcal{O}x|-1}x} A\Big).$$

Notice that any set with positive measure must contain a point in $\mathcal{O}x$. Suppose that A contained some points in $\mathcal{O}x$ but not all. Since A does not contain all of $\mathcal{O}x$, there must exist $T^i x \notin A$. Yet, since A also contains at least one point in $\mathcal{O}x$, there must be a transition where $T^i x \notin A$ but $T^{i+1}x \in A$.

We then see that since $T^i x \in T^{-1}(T^{i+1}x)$ and $T^i x \notin A$, $T^{-1}A \neq A$. Thus, if A contains some but not all points of $\mathcal{O}x$, then A is not an invariant set. Thus, if B is an invariant set, it contains all or none of $\mathcal{O}x$. In the first case, $\mu_x(B) = 1$ and in the second, $\mu_x(B) = 0$. Therefore, μ_x is ergodic.

Example 1.51 (Combination of Periodic Measures)**.** From Example 1.50 we see that periodic orbit measures are ergodic. Let (Ω, T, μ_x) be as in Example 1.50.

Let $a, b \in \Omega$ be periodic such that $\mathcal{O}a$ and $\mathcal{O}b$ are disjoint (in the case of a shift space, the existence of such points is obvious). Consider

$$\mu = \frac{1}{2}\mu_a + \frac{1}{2}\mu_b.$$

It is clear that μ is invariant since for any set A,

$$\mu(T^{-1}A) = \frac{1}{2}\mu_a(T^{-1}A) + \frac{1}{2}\mu_b(T^{-1}A) = \frac{1}{2}\mu_a(A) + \frac{1}{2}\mu_b(A) = \mu(A).$$

However, μ is not ergodic. We will construct an invariant set that has measure $1/2$. Notice that $\mu(\mathcal{O}a) = 1/2$ and so since μ is invariant, $\mu(T^{-1}\mathcal{O}a) = 1/2$. Since $T^{-1}\mathcal{O}a \cap \mathcal{O}b = \emptyset$ (a direct consequence of a, b being periodic and $\mathcal{O}a \cap \mathcal{O}b = \emptyset$), we have that $A = \big(T^{-1}\mathcal{O}a\big) \setminus \mathcal{O}a$ satisfies $\mu(A) = 0$. Further, notice that $\mathcal{O}a \cup T^{-1}A = T^{-2}\mathcal{O}a$. Thus we have for

$$A' = \bigcup_i T^{-i}A,$$

$A' \cup \mathcal{O}a$ is an invariant set (we've already included all its inverse images). But, since $\mu(A) = 0$ and μ is invariant, we may take a countable limit to get that $\mu(A') = 0$. Thus by the disjoint additivity of measures,

$$\mu(A' \cup \mathcal{O}a) = \mu(A') + \mu(\mathcal{O}a) = 0 + 1/2 = 1/2,$$

and so we have found an invariant set with measure neither 1 nor 0.

Example 1.52 (Uniquely Ergodic System)**.** Let α be an irrational number, $\Omega = [0, 1)$, and $T(x) = x + \alpha \mod 1$. Then, (Ω, T) is uniquely ergodic with the unique ergodic measure being Lebesgue measure. This system also happens to be minimal.

Example 1.53 (Uniquely Ergodic Subsystems)**.** Let $\Omega = \{0, 1\}^{\mathbb{Z}}$ and let T be the Shift. For any periodic point y, $(\mathcal{O}y, T)$ is uniquely ergodic with $\mu(A) = |A|/|\mathcal{O}y|$ being the unique ergodic measure.

Chapter 2

Ergodic Optimization

Now that we are familiar with ergodic measures, the building blocks of invariant measures, and the fact that for any given dynamical system, there are often uncountably many distinct ergodic measures, the question arises of how the integral of a function changes when integrating against different ergodic measures. In ergodic optimization we are interested in which ergodic measures maximize the values of integrals of particular classes of functions.

This class of questions can be considered a limiting case of the thermodynamical notion of finding equilibrium measures: measures that maximize the value of $\int f \, \mathrm{d}\mu + h(\mu)$ where $h(\mu)$ is the entropy of μ. If one considers measures that optimize $\int nf \, \mathrm{d}\mu + h(\mu)$ across all n, a quick renormalization gives $\lim_{n\to\infty} \frac{1}{n}\Big(\int nf \, \mathrm{d}\mu + h(\mu)\Big) = \int f \, \mathrm{d}\mu$ and so this problem is precisely that of Ergodic Optimization.

2.1 Measures from Points

To further simplify matters, because the measures we are studying are ergodic, we may restrict our study to points, their orbit, and the measures they generate. Let us consider an ergodic probability measure μ. By the Birkhoff Ergodic Theorem, we have that for a fixed f, for μ-almost all x,

$$\int f \, \mathrm{d}\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x).$$

Recall that we have restricted ourselves to the space of continuous functions and Ω is compact. Therefore, we may find a countable set of continuous functions $\{f_i\}$, dense in the set of all continuous functions with respect to the supremum norm (for example, polynomials with rational coefficients when $\Omega = [0,1]$) [9]. For each f_i, we have a set X_i of points that satisfy the Birkhoff Ergodic Theorem and $\mu(X_i) = 1$. Thus, since μ is a probability measure, if $X_\mu = \bigcap X_i$, then $\mu(X_\mu) = 1$.

The defining quality of X_μ is now for any point $x \in X_\mu$ and for any continuous function f,

$$\int f \, \mathrm{d}\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x).$$

Fix a particular $x \in X_\mu$. Let us now recall empirical measures. Consider the sequence of probability measures

$$\mu_{x,N} = \frac{1}{N}\Big(\delta_x + \delta_{Tx} + \delta_{T^2x} + \cdots + \delta_{T^{N-1}x}\Big).$$

Notice that $\mu_{x,N}$ is indeed a probability measure for each N, though it is not necessarily invariant (unless x is periodic with N a multiple of the period). We have constructed $\mu_{x,N}$ such that

$$\frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) = \int f \, \mathrm{d}\mu_{x,N}.$$

Since $\int f \, \mathrm{d}\mu_{x,N} \to \int f \, \mathrm{d}\mu$ for all continuous f, $\mu_{x,N} \to \mu$ by the definition of weak* convergence (see Example 2.5 for cases where $\mu_{x,N}$ is not convergent). Thus, x uniquely generates μ.

It is worth noting that the Riesz Representation Theorem also gives us that if T is continuous and x satisfies,

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

exists for all continuous f, then there exists a probability measure μ such that

$$\int f \, \mathrm{d}\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x)$$

for all continuous f.

It is often useful to have a way of describing the limiting behavior of x. This is called the ω-limit of x.

Definition 2.1 (ω-limit). *For a point x, the ω-*limit *of x, denoted by $\omega(x)$ is defined to be*

$$\omega(x) = \bigcap_{n=0}^{\infty} \overline{\bigcup_{i=n}^{\infty} \{T^i x\}},$$

where $\overline{\bigcup_{i=n}^{\infty} \{T^i x\}}$ denotes the set closure.

The ω-limit of x is essentially the lim sup of the closure of the orbit of x, and so it captures the "limit behavior" of x. It is clear that if x and y both generate some invariant measure μ, x need not equal y.

It should now be evident that the problem of studying an ergodic measure μ is translatable to studying the points that generate μ. There are many such points, so we would like to further restrict the points we study to that of Shift spaces and so-called *measure recurrent* points (definition 2.2).

Consider one-sided sequences $\{0,1\}^{\mathbb{N}}$ and consider the measure μ supported on the point $111\cdots$. It is clear that μ is generated by $111\cdots$. However, μ is also generated by $0001111\cdots$, or even $0\cdots 01111\cdots$ where there are $10^{100!}$ zeros before the first one. It becomes clear that the initial behavior of our generating point does not matter, but to make life easier, it would be nice to study points that do not contain irrelevant symbols in their initial positions.

Definition 2.2 (Measure Recurrent). *We say that the point x is* measure recurrent *if for each string of symbols $S = a_1 a_2 \cdots a_n$ in x, S occurs with positive frequency in x. That is*

$$\lim_{N\to\infty} \frac{\#\textit{of times } S \textit{ appears in the first } N \textit{digits of } x}{N} = \alpha_S$$

exists for all S, and $\alpha_S > 0$.

Note that the strength of the definition of measure recurrent comes from the fact that we assert that each sequence of symbols has a limiting probability. This means that if x is measure recurrent, x uniquely generates an ergodic measure μ_x. This gives us an equivalent characterization of measure recurrence.

Lemma 2.3. *The point x is measure recurrent if and only if x uniquely generates a measure μ_x with the property that if $S = a_1 a_2 \cdots a_n$ is a string of symbols that occurs in x,*

$$\mu_x([S]) > 0,$$

where $[S]$ *is the cylinder set* $[a_1a_2\cdots a_n]$.

Proof. Let $L_x(S) = \lim_{N\to\infty} \frac{\#\text{of times } S \text{ appears in the first } N \text{digits of } x}{N}$. If x is measure recurrent, by virtue of the fact that $L_x(S)$ exists for all S that occur in x and $L_x(S) = 0$ for all S that do not occur in x, x generates a unique measure μ_x. Further, recalling the construction of an empirical measure, we see $\mu_x([S]) = \alpha_S > 0$ for all S that occur in x.

If x uniquely generates a measure μ_x, we know $L_x(S)$ exists for all S. Further, if $\mu_x([S]) > 0$ for all S that occur in x, $L_x(S) > 0$ and so x is measure recurrent. □

Theorem 2.4. *For any ergodic measure* μ*, there exists a measure recurrent point* x *such that* x *generates* μ.

Proof. Suppose μ is ergodic. Let W with $\mu(W) = 1$ be the set of points that generate μ (which is measure one by the Birkhoff Ergodic Theorem). Since μ is invariant, for any cylinder set C with $\mu(C) = 0$, we know for all $n \in \mathbb{Z}$, $\mu(\bigcup T^{-n}C) = 0$. Thus, since there are a countable number of cylinder sets, the set

$$A = \bigcup_{\substack{n \in \mathbb{Z} \\ \mu(C)=0}} T^{-n}C$$

satisfies $\mu(A) = 0$. This gives that $W \backslash A$ is non-empty and consists precisely of the measure recurrent points that generate μ. □

Notice that if x is measure-recurrent and μ_x is the unique measure generated by x, the support of μ_x is $\omega(x)$. However, for a general x (perhaps non-measure recurrent) and a cylinder set $S = [a_1a_2\cdots a_n]$, if μ is an empirical measure generated by x, the conditions $S \cap \mathcal{O}x \neq \emptyset$, $S \cap \omega(x) \neq \emptyset$, and $S \cap \text{supp}(\mu_x) \neq \emptyset$ are all different. In the first case, we have that the string of symbols $a_1a_2\cdots a_n$ occurs somewhere in x. The second case states that the string of symbols $a_1a_2\cdots a_n$ occurs infinitely many times in x, and in the last case, we have that the string of symbols $a_1a_2\cdots a_n$ occurs with positive frequency, which implies it occurs infinitely many times (see Example 2.6 for an illustration of these discrepancies). If x is measure recurrent, all three conditions are equivalent.

2.1.1 Examples

Example 2.5 (Multiple Empirical Measures)**.** Consider again $\Omega = \{0,1\}^{\mathbb{N}}$ and T the one-sided Shift. Let x be the point

$$x = (0,1,0,0,1,1,0,0,0,0,1,1,1,1,0,0,0,0,0,0,0,0,\ldots),$$

where every block of zeros is followed by an equal-length block of ones and every block of ones is followed by a block of zeros that is twice as long.

If we define a sequence of measures

$$\mu_N = \frac{1}{N}\Big(\delta_x + \delta_{Tx} + \delta_{T^2x} + \cdots + \delta_{T^{N-1}x}\Big)$$

as before, we know that there exist convergent subsequences. However, μ_N itself is not convergent.

To see this, notice that for any fixed k and $\varepsilon > 0$, there are infinitely many N such that, $|\mu_N([0^k]) - 1/2| < \varepsilon$ and $|\mu_N([1^k]) - 1/2| < \varepsilon$, where $[a^k]$ is the cylinder set starting with k a's (this is achieved for all large enough N of the form $N = \sum 2^i$). Thus, because μ_N are probability measures, there is a subsequence μ_{N_i} such that

$$\mu_{N_i} \to \frac{1}{2}\big(\delta_{(0,0,\ldots)} + \delta_{(1,1,\ldots)}\big).$$

However, since $\sum_{i=0}^{j} 2^i = 2^{j+1} - 1$, there are an infinite number of N such that the first N digits of x contain roughly twice as many zeros as ones. Thus, for a fixed k and $\varepsilon > 0$, there are infinitely many N such that $|\mu_N([0^k]) - 2/3| < \varepsilon$ and $|\mu_N([1^k]) - 1/3| < \varepsilon$, and so there is a subsequence μ_{N_i} such that

$$\mu_{N_i} \to \frac{2}{3}\delta_{(0,0,\ldots)} + \frac{1}{3}\delta_{(1,1,\ldots)}.$$

In fact, we may find subsequence μ_{N_i} of μ_N such that

$$\mu_{N_i} \to \alpha\delta_{(0,0,\ldots)} + (1-\alpha)\delta_{(1,1,\ldots)}$$

for any $\alpha \in [1/2, 2/3]$. This is a complete characterization of the convergent subsequences of μ_N, which can be seen by noticing that the only strings that occur with positive frequency are those of the form 0^k or 1^k (strings of the form 00111 for example happen with geometrically decreasing frequency and so do not exist in the limit), thus any convergent subsequence will result in a combination of $\delta_{(0,0,\ldots)}$ and $\delta_{(1,1,\ldots)}$. Since we have already calculated the extreme case of $\alpha = 2/3$, we have that $\alpha \leq 2/3$. Noticing that x gives the restriction $\alpha \geq 1-\alpha$ gives the complete characterization.

Because μ_N contains subsequences that converge to different measures, μ_N is not convergent. Thus there are multiple empirical measures derived from x.

Example 2.6 (Non-measure Recurrence)**.** Let Ω, T, and μ_N be defined as in Example 2.5 and let

$$x = (0,1,1,1,0,0,1,1,1,0,0,0,0,1,1,1,0,0,0,0,0,0,0,1,1,1,\ldots)$$

where each block of ones of length three is separated by a block of zeros whose length is increasing exponentially. In this case, we have that $\omega(x)$ contains the points $(1,1,1,0,0,\ldots)$, $(0,1,1,1,0,0,\ldots)$, $(0,0,1,1,1,0,0,0,\ldots)$, etc.. However, $\mu_N \to \delta_{(0,0,\ldots)}$ uniquely, $\text{supp}(\mu_x) = (0,0,\ldots)$ where μ_x is the empirical measure generated by x, and so $\text{supp}(\mu_x)$ and $\omega(x)$ are very different things.

Example 2.7 (Non-ergodic Empirical Measure)**.** Let Ω, T, and μ_N be defined as in Example 2.5 and let

$$x = (0,1,0,0,1,1,0,0,0,1,1,1,0,0,0,0,1,1,1,1,\ldots)$$

where every block of zeros is followed by an equal-length block of ones and every block of ones is followed by a block of zeros that is one longer. It is left to the reader to verify that since at position n of the sequence x, the number of zeros and ones up to position n is roughly bounded by $n/2 \pm \sqrt{n}$, there is indeed a limiting frequency of zeros and ones and so x uniquely generates the empirical measure

$$\mu_x = \frac{1}{2}\big(\delta_{(0,0,\ldots)} + \delta_{(1,1,\ldots)}\big).$$

(This differs from Example 2.5 in that μ_x is unique, which is partly a result of the fact that ones and zeros each have limiting probability 1/2, where in Example 2.5, the limit probability was not convergent).

Since μ_x is a non-trivial linear combination of two ergodic measures, μ_x is not ergodic.

2.2 Maximizing Measures

For a continuous function f, consider the question of the existence of an ergodic measure μ such that

$$\int f \,\mathrm{d}\mu = \max_{\rho \in \mathcal{M}_T} \int f \,\mathrm{d}\rho$$

where $\mathcal{M}_T$ is the set of all invariant probability measures.

We have already done all the work to show the existence of μ. By Theorem 1.46, we have that $\mathcal{M}_T$ is compact in the weak* topology, which is the smallest topology that makes integration continuous, so $\{\int f \,\mathrm{d}\rho : \rho \in \mathcal{M}_T\}$ is the continuous image of a compact set and therefore compact. Thus, $\max_{\rho \in \mathcal{M}_T} \int f \,\mathrm{d}\rho$ is attained by some measure ν. By the Ergodic Decomposition Theorem, ν may be written as a convex combination of ergodic measures, and thus there must be an ergodic measure μ that attains the maximum. In fact, almost every ergodic measure that makes up μ is maximizing.

Since μ is ergodic, from the results in the previous section, we know there is a point x that generates μ (in fact, a μ-measure one set of points that generate μ). We now say that x optimizes f.

Recall the notation

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

When we write $\langle f \rangle (y)$, it is implicitly assumed that $\langle f \rangle (y)$ is well defined.

We can now properly define what it means for f to be optimized.

Definition 2.8. *We say the ergodic probability measure μ optimizes the continuous function f if*

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

for all ergodic probability measures ν.

We say μ uniquely optimizes *f if*

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

for all ergodic probability measures $\nu \neq \mu$.

Since we have shown that any ergodic measure μ may be assumed to be the empirical measure generated by the measure recurrent point x, we have an equivalent notion of optimization given by the point x.

Definition 2.9. *We say that x (or $\mathcal{O}x$)* optimizes f *if for all y,*

$$\langle f\rangle(x) \geq \langle f\rangle(y).$$

It is immediate that x optimizes f if and only if μ_x optimizes f: since $\langle f\rangle(x) = \int f\,\mathrm{d}\mu_x$ and $\langle f\rangle(y) = \int f\,\mathrm{d}\mu_y$, it is a trivial consequence that if μ_x optimizes f, then $\langle f\rangle(x) \geq \langle f\rangle(y)$. The reverse implication is also trivial.

Uniquely optimizing is harder to quantify in terms of individual points. Since $\langle f\rangle(x) = \langle f\rangle(Tx)$, we clearly cannot say that x uniquely optimizes f. For this reason, it is preferable to only refer to a function being uniquely optimized by a particular measure and not by points that generate the optimizing measure.

Theorem 2.10. *If μ is a maximizing measure for a continuous function f, then*

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N-1} f(T^i x) \leq \int f\,\mathrm{d}\mu$$

for all x.

Proof. From our discussion of empirical measures, we know that $\limsup_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N-1} f(T^i x) = \int f\,\mathrm{d}\nu$ for some empirical measure ν. Since μ is a maximizing measure $\int f\,\mathrm{d}\nu \leq \int f\,\mathrm{d}\mu$. □

Theorem 2.10 allows us to bound the average along the orbit points that are not measure recurrent.

2.2.1 Examples

Example 2.11 (Constant Function)**.** The constant function $f(x) = 0$ is optimized by all invariant measures.

Example 2.12 (Uniquely Optimized)**.** If x is a measure recurrent point such that $(T, \omega(x))$ is uniquely ergodic, then the function

$$g(t) = -d(t, \mathcal{O}x)$$

is uniquely optimized by μ_x. Note that for all periodic points x, $(T, \mathcal{O}x)$ is uniquely ergodic, but there do exist aperiodic points so that $(T, \omega(x))$ is uniquely ergodic,

Example 2.13 (C^∞ Family)**.** The family of functions

$$\rho(t) = \alpha \cos(2\pi t) + \beta \sin(2\pi t)$$

for $\alpha, \beta \neq 0$ are all optimized by Sturmian measures (both generated by periodic Sturmian sequences and aperiodic ones). This result is due to Bousch [1].

2.3 Known Results

Many of the best known results of ergodic optimization are summarized in an excellent survey paper by Jenkinson [5], so this section will only highlight a handful of the most relevant known results.

Established results about ergodic optimization take place in a handful of related spaces: the doubling map, the one-sided Shift, and the two-sided Shift. The *doubling map* on the unit interval is defined by the transformation $Tx = 2x \mod 1$. It is very similar to the one-sided Shift, and one can often go back and forth between the doubling map and the one-sided Shift by interpreting a point x either as its binary representation, or interpreting a sequence of zeros and ones as a real number (in binary). The two-sided Shift has obvious relations to the doubling map and the one-sided Shift, but differs in the fact that for the two-sided Shift, T is invertible.

A *co-boundary* is any function that can be written in the form $g \circ T - g$ for some g. Two functions are called *co-homologous* if they differ by a co-boundary. Some of the most useful theorems in the field pertain to existence of special types of co-homologous functions. It should be noted that co-boundaries have the useful property that they integrate to zero with respect to any invariant measure (for the simple reason that, if μ is T-invariant, $\int g \,\mathrm{d}\mu = \int g \circ T \,\mathrm{d}\mu$ for any integrable g). Thus, any two functions that differ by a co-boundary (i.e., any two co-homologous functions) have the same integral with respect to any invariant measure. Therefore, any results about optimizing measures of a particular function immediately carry over to all co-homologous functions.

In particular, it is very convenient to work with functions that are non-positive and attain the value zero only on the support of an optimizing measure. And, there are several spaces where every function is co-homologous to a function with these properties. Bousch presents proofs for the existence of such co-homologous functions for the space of Hölder functions and functions satisfying the Walter's condition in the case of the doubling map [1, 2]. It is worth pointing out that the existence of these special co-homologous functions can be generalized from the space of Hölder functions with the doubling map to the space of Hölder functions in the two-sided Shift.

The most common way to derive such co-homologous functions is to first produce a lemma that guarantees the existence of a co-boundary that lies above your function. Let $M(f) = \max_\mu \int f \, \mathrm{d}\mu$ be the maximal integral of f. Suppose we can find a suitable function g (one with the desired properties of our space, e.g. Hölder) such that

$$f \leq M(f) + g \circ T - g.$$

We may then define a function

$$h = f - g + g \circ T \leq M(f).$$

We then know that if for some measure μ

$$\int h \, \mathrm{d}\mu = M(f),$$

then $h(x) = M(f)$ almost everywhere on $\mathrm{supp}(\mu)$. This gives us the condition that some measure ν is f-maximizing if and only if $h(x) = M(f)$ for ν-almost all x.

Using these h-type co-homologous functions, slightly simpler proofs of many of the theorems in this thesis regarding Lipschitz functions can be found. However, the existence of h-type functions for the larger class of functions of summable variation and functions of finite A-norm (a norm introduced in the next chapter) has not been established, so this thesis does not refer to h-type co-homologous functions.

Another tool for studying optimizing measures is the subordination principle: f satisfies the subordination principle if for every maximizing measure μ, all invariant measures ν with $\mathrm{supp}(\nu) \subset \mathrm{supp}(\mu)$ also maximize (see [2]). Large classes of functions, such as Hölder continuous functions, can be shown to satisfy the subordination condition. If there exist h-type co-homologous functions, the subordination principle immediately follows. If $f(x) = 0$ for $x \in \mathrm{supp}(\mu_{\max})$, then the optimum integral of f is zero. It is then clear that for any invariant measure ν with $\mathrm{supp}(\nu) \subset \mathrm{supp}(\mu)$, $\int f \, \mathrm{d}\nu = 0$ and so ν is also an optimizing measure. It is worth noting that although co-homologous functions are the most common method for demonstrating the subordination principle, Morris has shown the subordination principle can be proved without the need for h-like co-homologous functions [7].

One of the main goals of ergodic optimization is describing typical properties of measures that maximize functions in some particular class of function. Specifically, one is interested in proving results of the form: there is a large subset A of functions whose optimizing measures share nice properties (like having finite support, etc.). Large can take a variety of meanings, but most would consider that a set that is open and dense or even a set that is a countable intersection of open, dense sets (or a residual set) is "large." In terms of the thermodynamics equilibrium state problem of maximizing $\int f \, d\mu + h(\mu)$, the $h(\mu)$ often pushes μ towards fully supported measures. However, in ergodic optimization, it seems to be the case that optimizing measures have low complexity (and therefore low entropy).

The lowest complexity measures are those supported on a periodic point. Bousch showed in [1] that under the doubling map, $f(x) = \sin(2\pi(x - \alpha)$ is optimized by a Sturmian measure for all $\alpha \in [0, 1)$. Further, $\{\alpha \in [0, 1) : \sin(2\pi(x - \alpha))$ is optimized by a periodic Sturmian measure$\}$ is a Lebesgue measure one set whose complement has Hausdorff dimension zero. For α where $\sin(2\pi(x - \alpha))$ is not optimized by a periodic Sturmian measure, it is optimized by a Sturmian measure derived from an irrational slope and so is generated by a sequence with lowest possible complexity for a non-periodic sequence. In fact, it is unknown if there exists an analytic function that is optimized by a measure of positive entropy[5].

It would appear that functions optimized by periodic orbit measures constitute a large set. For the general class of continuous functions with T the doubling map or the shift (one-sided or two-sided), this is known not to be the case. It can be shown that the set of functions optimized by measures of full support is a residual set[5]. Thus, the set of continuous functions optimized by periodic orbit measures is small. However, if we restrict the class of functions to the Lipschitz/Hölder case, the results are much more promising.

Contreras, Lopes, and Thieullen showed in [4] that if you restrict yourself to the subspace of Hölder functions with Hölder exponent strictly larger than α, then the set of functions uniquely optimized by periodic orbit measures is open and dense. However, this result somewhat cheats in that the space of functions they consider is not a Banach space. The norm under which this set is open and dense is given by the α-Hölder norm and as such might be considered "outside" the space of functions being considered.

Yuan and Hunt made significant progress towards showing that there was an open, dense subset of Lipschitz functions optimized by periodic orbit measures in [13]. Propositions 3.8, 3.10, 3.11, 3.12, 3.19, 3.20, 3.24 and Corollary 3.9 of this thesis were first proved by Yuan and Hunt in the general case of a hyperbolic dynamical system. However, some of the propositions mentioned were presented as remarks without proof in [13] and others have been made slightly stronger than the original propositions. This thesis presents re-proofs of these results in the simplifying environment of the two-sided Shift, borrowing important ideas from proofs in [13], but significantly altering the logical flow of said proofs in an attempt to clarify the key concepts.

While not fully proving that Lipschitz functions optimized by periodic orbit measures contain an open, dense set, Yuan and Hunt showed that functions optimized by aperiodic measures were unstable in the sense that arbitrarily small perturbations would cause these functions to no longer be optimized by the original aperiodic measure.

Another class of functions that shows up often in ergodic optimization is functions that satisfy the Walters condition, introduced by Walters in [12]. We say f satisfies the Walters condition if for every ε 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 $S_n f(w) = \sum_{i=0}^{n-1} f(T^i w)$. The Walters functions are the set of f satisfying the Walters condition. Walters functions form a Banach space when equipped with an appropriate norm, and Bousch has shown that the set of Walters functions optimized by a periodic orbit measure is dense in the set of all Walters functions.

This thesis extends the results of Yuan and Hunt to the case of functions of summable variation and the Banach subspaces thereof generated by A-norms (a generalization of the Lipschitz norm). Many of the results of [13] carry over to A-norm generated spaces that are more general than Hölder continuous functions (that is, they contain the set of Hölder continuous functions). Further, we present a subclass of Lipschitz continuous functions (defined in terms of an A-norm), deemed *super-continuous* functions, where the set of functions optimized by periodic orbit measures contains an open, dense subset.

Presented for comparison are a list of the established theorems about the set of functions optimized by periodic orbit measures and the new result that this thesis provides.

Theorem (Bousch [2])**.** *Let $T : X \to X$ be the doubling map on S^1 and let W denote the set of Walters functions on X. If $P \subset W$ is the set of Walters functions optimized by a measure supported on a periodic point, then P contains an open set dense in W with respect to the Walters norm.*

Theorem (Contreras-Lopes-Thieullen [4])**.** *Let T be the doubling map on S^1. Let H_α be the set of α-holder 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 [13])**.** *Let T be the Shift (one-sided or two-sided) and let L denote the class of Lipschitz continuous functions. For any $f \in L$ optimized by a measure generated by an aperiodic point, there exists an arbitrarily small perturbation of f such that that measure is no longer the optimizing measure.*

This thesis presents the following addition.

Theorem (3.39)**.** *Let T be the two-sided Shift and let S be the Banach space generated by the A-norm $\|\cdot\|_A$ where A_n satisfies $A_n/A_{n+1} \to \infty$. Let $P \subset S$ be the set of functions uniquely optimized by a measure supported on a periodic orbit. Then, P contains a set that is open and dense in S under $\|\cdot\|_A$.*

Note that some of these theorems were proved in a slightly more general context than stated here. Those theorems that cannot be easily extended to apply to invertible spaces, specifically the two-sided Shift, are stated in terms of the doubling map. It should be noted that most results about the one-sided Shift carry over to the doubling map, and many results about the two-sided Shift carry over to the one-sided Shift. However, when it comes to issues of dense and open subsets, the usual method of applying a "forgetful" map (one that deletes all symbols to the left of the radix point) to transform the two-sided Shift into a one-sided Shift does not necessarily preserve openness or denseness.

Chapter 3

Stability and Instability

We are now prepared to analyze measures and the family of functions they optimize. It will turn out that for measures μ_x generated by x where x is periodic, there is an open set of functions optimized by μ_x. However, if x is aperiodic, the set of functions optimized by μ_x does not contain an open set. We will therefore say that functions optimized by measures supported on periodic points are somehow stable, while those that are not are unstable.

3.1 Lipschitz Functions

The first (and simplest) class of functions we will analyze are the Lipschitz continuous functions.

Definition 3.1 (Lipschitz). *A function f is said to be* Lipschitz continuous *with Lipschitz constant L (or just Lipschitz with Lipschitz constant L) if*

$$|f(x) - f(y)| \leq Ld(x, y)$$

for all x, y.

The space of all Lipschitz functions is also a Banach space with the norm

$$\|f\|_{\text{Lip}} = L_f + \sup_{x \in \Omega}\{|f(x)|\},$$

where L_f is the Lipschitz constant corresponding to f.

Most of our results about optimizing measures of Lipschitz functions rely on breaking up the orbit of points into places where points become very near to each other and those where they are reasonably far apart. Lipschitz continuity gives us that if two segments of an orbit are close enough to each other, the difference between averages along those segments is very small.

Definition 3.2 (In Order for One Step)**.** *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, suppose that there is a unique closest point $y' \in S$ to x. I.e.,*

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

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 3.3 (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 x in order for one step.*

Note that because of the uniqueness requirement for following S in order, if $y' \in S$ is the unique closest point to x, x following S in order for p steps implies $T^i y' \in S$ is the unique closest point to $T^i x$ for $0 \leq i < p$. It is also worth pointing out that following in order is very similar to shadowing, except for following in order, there is a uniqueness requirement for which point you are closest to at each step.

Definition 3.4 (Out of Order)**.** *We say x follows a segment S* out of order for one step *if x does not follow S in order for one step.*

Lemma 3.5 (In Order Lemma)**.** *Let y be a periodic point, and let*

$$\gamma = \min_{i \neq j}\{d(T^i y, T^j y)\} \qquad \rho \leq \frac{\gamma}{4}.$$

For any point x, if x stays ρ-close to $\mathcal{O}y$ for k steps, then x follows $\mathcal{O}y$ in order for k steps. I.e., there exists some i' such that for $0 \leq j < k$,

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

Proof. First, let us derive a fact about the Shift space due to its ultrametric properties. Suppose $y', y'' \in \mathcal{O}y$ and for some point x, $d(x, y'), d(x, y'') \leq \gamma/2$. By the ultrametric triangle inequality we have

$$d(y', y'') \leq \max\{d(x, y'), d(x, y'')\} = \gamma/2.$$

And $d(y', y'') \leq \gamma/2$ implies $d(y', y'') < \gamma$. Thus, since γ was the smallest distance between points in $\mathcal{O}y$, $y' = y''$. This shows that for any point x, if $d(x, \mathcal{O}y) \leq \gamma/2$, then there is a unique closest point in $\mathcal{O}y$ to x.

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

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

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

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

We then have that

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

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 $d(Tx, \mathcal{O}y) \leq \rho$, 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; specifically

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

for $0 \leq j < k$.

□

Lemma 3.6 (In Order Lemma part 2)**.** *For a point z, let $S = \{T^i z, T^{i+1}z, \ldots, T^{i+k}z\}$ be a segment of $\mathcal{O}z$. For any $\rho < 1$, if a point x ρ-shadows S, the distance between S and $T^j x$ for $0 \leq j \leq k$ is geometrically bounded by*

$$d(T^{i+j}z, T^j x) \leq \rho 2^{-\min\{j,k-j\}}$$

Proof. Since $\rho < 1$ and x ρ-shadows S for $k+1$ steps, $(x)_0^k = (T^i z)_0^k$. Let $l = \inf\{w : 2^{-w} \leq \rho\}$. Since all distances in the shift space are powers of two, $d(x, T^i z), d(T^k x, T^{i+k}z) \leq$

$2^{-l} \leq \rho$. Thus, we get $(x)_{-l}^{k+l} = (T^i z)_{-l}^{k+l}$. This gives us a stronger bound on $d(T^j x, T^{i+j} z)$, namely

$$d(T^{i+j}z, T^{i'+j}y) \leq \rho 2^{-\min\{j,k-j\}},$$

which is valid for $0 \leq j \leq k$. □

Lemma 3.7 (Parallel Orbit Lemma). *Let f be a Lipschitz function with Lipschitz constant L. For a point z, let $S = \{T^i z, T^{i+1} z, \ldots, T^{i+k} z\}$ be a segment of $\mathcal{O}z$. For any $\rho < 1$, if a point x ρ-shadows S, the average along S is close to the average along $\mathcal{O}x$ in the following sense:*

$$\sum_{j=0}^{k} \left| f(T^{i+j}z) - f(T^j x) \right| < 4L\rho.$$

Proof. By the In Order Lemma part 2, we have that

$$d(T^{i+j}z, T^j x) \leq \rho 2^{-\min\{j,k-j\}}.$$

Thus, by summing a geometric progression we conclude

$$\sum_{j=0}^{k} |f(T^{i+j}z) - f(T^j x)| \leq \sum_{j=0}^{k} L d(T^{i+j}z, T^j x) \leq \sum_{j=0}^{k} L\rho 2^{-\min\{j,k-j\}} < 4L\rho.$$

□

The coming proofs will make substantial use of the technique of cutting and splicing. Let x be a point and suppose that for some r, s we have $d(T^r x, T^s x) < 2^{-k}$ is very small. Since x comes very close to itself, we can imagine "cutting out" the portion of $\mathcal{O}x$ between $T^r x$ and $T^s x$. To make this formal, let $x' = (x)_{-\infty}^{r-1}.(x)_s^{\infty}$ be the concatenation of the symbols $(x)_{-\infty}^{r-1}$ and $(x)_s^{\infty}$ with the radix point occurring directly before $(x)_s^{\infty}$.

The new point x' for follows $\mathcal{O}x$ very closely in the sense that

$$d(T^i x, T^{i-r} x') < 2^{-k} \qquad \text{for } i \leq r \tag{3.1}$$

and

$$d(T^i x, T^{i-s} x') < 2^{-k} \qquad \text{for } i \geq s. \tag{3.2}$$

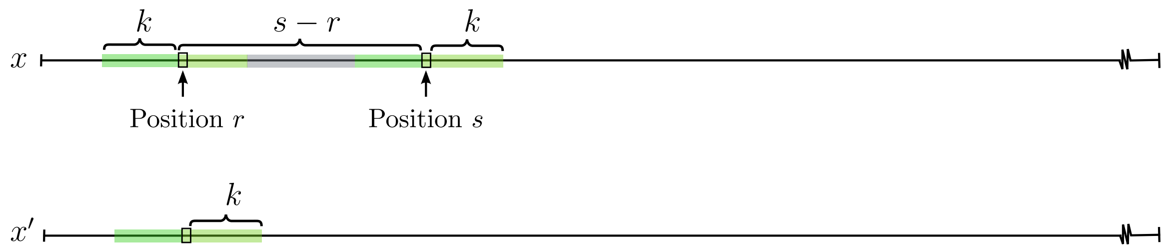

Figure 3.1: The cutting and splicing procedure.

Having only performed one cut-and-splice, we already see that the superscripts of $T^{i-s}x'\,T^{i-r}x'$ are becoming somewhat cumbersome. For this reason, we introduce simplifying notation. Divide the orbit of x into three segments: $S_0 = \{\ldots, T^{r-2}x, T^{r-1}x\}$, $P = \{T^r x, \ldots T^{s-1}x\}$, and $S_1 = \{T^s x, T^{s+1}x, \ldots\}$. Abusing the distinction between points and symbols, we might now say x' is the "concatenation" of S_0 and S_1.

Let $S_0' = \{\ldots, T^{r-2}x', T^{r-1}x'\}$ and $S_1' = \{T^r x', T^{r+1}x', \ldots\}$. We may use the notation $i \in S_1$ to mean not literally a point in S_1, but an exponent such that $T^i x \in S_1$. In situations where there is a clear relationship between segments, like there is between S_0 and S_0' or between S_1 and S_1', we may use i' to mean the exponents such that $T^{i'}x'$ is the point corresponding to $T^i x$.

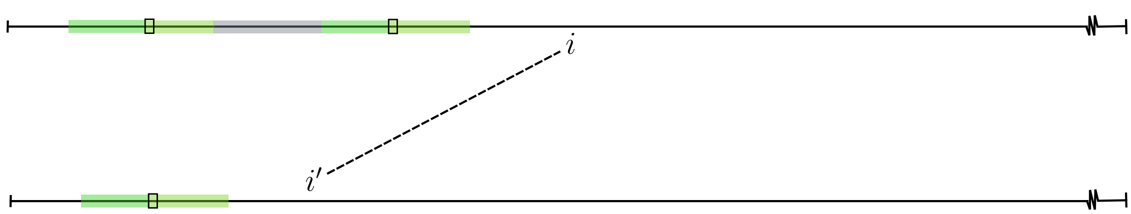

Figure 3.2: The correspondence between i and i'.

Using this notation, equations (3.1) and (3.2) may be rewritten as

$$d(T^i x, T^{i'}x') < 2^{-k} \qquad \text{for } i \in S_j$$

where $j \in \{0, 1\}$.

Extending this notation further, when it will not cause ambiguity, we may simply refer to $i \in S_j$ and the corresponding i' even when S_j' is not explicitly specified.

For the detail-oriented reader, we shall also give the explicit correspondence between i and i'. Let x be a point and divide $\mathcal{O}x$ into segments $S_j = \{T^{s_j}x, T^{s_j+1}x, \ldots, T^{s_j+d_j-1}x\}$ and $W_j = \{T^{w_j}x, T^{w_j+1}x, \ldots, T^{w_j+h_j-1}x\}$ for $j \in \mathbb{Z}$, where each S_j segment is preceded by a W_j, these segments completely cover $\mathcal{O}x$, and $T^{-1}x \in S_{-1}$ and $x \in W_0$. That is,

$$x = \cdots (x)_{w_{-1}}^{w_{-1}+h_{-1}-1}(x)_{s_{-1}}^{s_{-1}+d_{-1}-1}.(x)_{w_0}^{w_0+h_0-1}(x)_{s_0}^{s_0+d_0-1}\cdots.$$

Define the point

$$x' = \cdots (x)_{s_{-2}}^{s_{-2}+d_{-2}-1}(x)_{s_{-1}}^{s_{-1}+d_{-1}-1}.(x)_{s_0}^{s_0+d_0-1}(x)_{s_1}^{s_1+d_1-1}\cdots$$

as the concatenation of the S_j's. Then, for $i \in S_j$, we know that i takes the form

$$i = s_j + n \qquad 0 \le n < d_j.$$

Let

$$l = \begin{cases} s_j - \sum_{0 \le k \le j} h_k & \text{if } j \ge 0 \\ s_j + \sum_{j < k < 0} h_k & \text{if } j < 0 \end{cases}.$$

We may then write

$$i' = l + n.$$

We now have the needed notation to prove Proposition 3.8, which itself 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.

Proposition 3.8. *Let f be a Lipschitz function with constant L, x be a measure recurrent point with $\mathcal{O}x$ a measure-recurrent optimal orbit for f, y a point of period p, and $\delta < 1$. If $\mathcal{O}x$ δ-shadows $\mathcal{O}y$ for p steps (i.e., there exists an m such that $d(T^{i+m}x, T^iy) \le \delta$ for $0 \le i < p$), then*

$$\langle f\rangle(x) - 8L\delta/p \le \langle f\rangle(y) \le \langle f\rangle(x).$$

Proof. Let L be the Lipschitz constant of f. Since $\mathcal{O}x$ δ-shadows $\mathcal{O}y$ for a period p, without loss of generality assume there exists some m so that $d(T^{m+i}x, T^iy) \le \delta$ for $0 \le i < p$. Since x is measure recurrent, this happens with positive frequency. Let N_y be the number of disjoint occurrences of the symbols $(y)_0^{p-1}$ in $(x)_0^{N-1}$ as determined by the following greedy algorithm: Find the smallest $0 \leq k_0 < N - p$ such that $(x)_{k_0}^{k_0+p-1} = (y)_0^{p-1}$. Next, find the smallest k_1 with $k_0 + p \leq k_2 < N - p$ such that $(x)_{k_1}^{k_1+p-1} = (y)_0^{p-1}$. Continue this process until you reach the first i where k_i does not exist. Let $N_y = i$. Because x is measure recurrent, we know $N_y/N \to \eta > 0$.

Recall that

$$\frac{1}{M}\sum_{i=0}^{M-1} f(T^i x) \to \langle f \rangle (x).$$

For any M sufficiently large, we will construct a point x' from $(x)_0^{M-1}$ by greedily deleting (guided by the aforementioned algorithm) each segment of length p that δ-shadows $\{y, Ty, \ldots, T^{p-1}y\}$. We will enumerate the deleted sections by $P_1, P_2, \ldots, P_{\mathcal{M}}$ (Note the relation $\mathcal{M}/M \to \eta$ as $M \to \infty$) and the undeleted sections by $S_1, S_2, \ldots$. By construction, P_j has the property that

$$d(T^i x, T^{i'} y) \leq \delta$$

for $i \in P_j$. Partition the sequence x' into $S_1', S_2', \ldots$ where S_1' corresponds to S_1, etc.. If done properly, we have constructed the S_j' so that,

$$d(T^i x, T^{i'} x') \leq \delta$$

for $i \in S_j$. Let M' denote the length of x'. That is $M' = M - \mathcal{M}p \approx (1 - \eta p)M$.

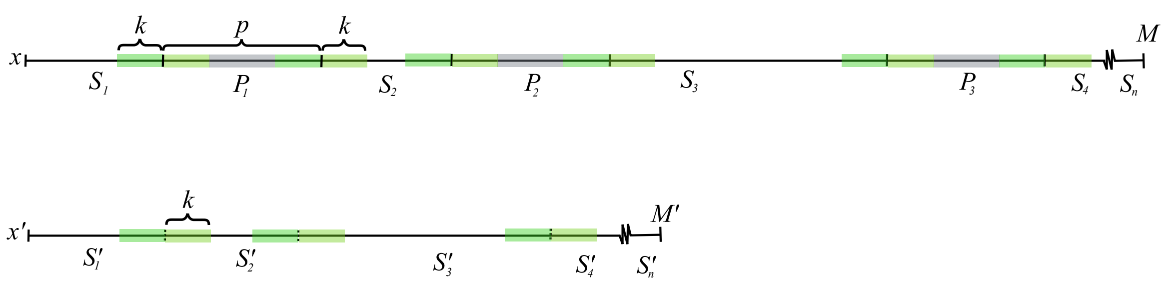

Figure 3.3: The procedure of removing segments that δ-shadow y for p steps.

Figure 3.3 gives a sketch of the procedure for constructing x'. In Figure 3.3, you can see $P_1, P_2, \ldots$ correspond to the occurrence of the symbols of y in x. The boundaries of the P_i blocks agree with each other k symbols to either side, so deleting the P_i blocks, we have that the S_i blocks match the S_i' blocks for at least k symbols to either side of any particular pivot symbol, with $2^{-k-1} \leq \delta$ (recall that two points are 2^{-k-1}-close if their initial symbol agrees and they agree k symbols to the left and

k symbols to the right of the initial symbol. I.e., you have to go $k+1$ symbols to find a disagreement).

The procedure of constructing x' from x is completely determined in the sense that x' converges as $M \to \infty$.

We will now examine the differences between $\sum_{P_j} f(T^i x)$ and $\sum f(T^i y)$. We already know that $d(T^i x, T^{i'} y) \leq \delta$ for $i \in P_j$, and so the Parallel Orbit Lemma gives us

$$\sum_{i \in P_j} |f(T^i x) - f(T^{i'} y)| < 4L\delta. \tag{3.3}$$

Similarly, for S_j and S'_j we have $d(T^i x, T^{i'} x') \leq \delta$ for $i \in S_j$, and so the Parallel Orbit Lemma gives us

$$\sum_{i \in S_j} |f(T^i x) - f(T^{i'} x')| < 4L\delta.$$

We are now ready to make some comparisons with $M \langle f \rangle (x) \approx \sum_{i=0}^{M-1} f(T^i x)$. First note that

$$\sum_{i=0}^{M-1} f(T^i x) = \sum_j \sum_{i \in P_j} f(T^i x) + \sum_j \sum_{i \in S_j} f(T^i x).$$

Since there are $\mathcal{M}$ P_j's, we may bound the number of S_j's by $\mathcal{M}+1$ (which would be the number of S_j's under the assumption that there were no two adjacent P_j's). Thus, by applying the triangle inequality $2(\mathcal{M}+1)$ times (the number of P_j's plus the max number of S_j's), we have

$$\left| \sum_{i=0}^{M-1} f(T^i x) - \sum_{i=0}^{M'-1} f(T^i x') - \mathcal{M} \sum_{i=0}^{p-1} f(T^i y) \right| < 2(\mathcal{M}+1)4L\delta = 8(\mathcal{M}+1)L\delta.$$

Now, consider the average contribution of each term in the summands. Let

$$\begin{aligned} A &= \frac{1}{p} \sum_{i=0}^{p-1} f(T^i y), \\ A'_M &= \frac{1}{M'} \sum_{i=0}^{M'-1} f(T^i x') = \frac{1}{M - p\mathcal{M}} \sum_{i=0}^{M-p\mathcal{M}-1} f(T^i x'), \text{ and} \\ D_M &= \frac{1}{M} \sum_{i=0}^{M-1} f(T^i x). \end{aligned}$$

We then have that

$$|MD_M - (M - \mathcal{M}p)A'_M - \mathcal{M}pA| < 8(\mathcal{M}+1)L\delta.$$

From this equation it becomes clear that D_M is very close to a fixed convex combination of A and A'_M.[1] The essential idea now becomes: A must be very close to D_M, lest A'_M become greater than D_M, contradicting maximality.

Since $D_M \to \langle f \rangle (x)$ and $A'_M \to \mathcal{A}' \leq \langle f \rangle (x)$ as $M \to \infty$, for any $\varepsilon > 0$, we have that for sufficiently large M,

$$D_M \geq \langle f \rangle (x) - \varepsilon$$
$$A'_M \leq \langle f \rangle (x) + \varepsilon.$$

Using this with the fact that $MD_M - (M - \mathcal{M}p)A'_M - \mathcal{M}pA \leq 8(\mathcal{M}+1)L\delta$ and $M - \mathcal{M}p \geq 0$ gives us

$$M\big(\langle f \rangle (x) - \varepsilon\big) - (M - \mathcal{M}p)\big(\langle f \rangle (x) + \varepsilon\big) - \mathcal{M}pA$$

$$= \mathcal{M}p(\langle f \rangle (x) - A) - (2M - \mathcal{M}p)\varepsilon < 8(\mathcal{M}+1)L\delta$$

Dividing both sides by $\mathcal{M}p$ and letting $M \to \infty$ (and, correspondingly, $M/\mathcal{M} \to 1/\eta$) gives us

$$\langle f \rangle (x) - A - (2/\eta - 1)\varepsilon < 8L\delta/p.$$

But, $2/\eta - 1$ is fixed and ε was arbitrary, so recalling that $A = \langle f \rangle (y)$, we get the final result

$$\langle f \rangle (x) - 8L\delta/p \leq \langle f \rangle (y) \leq \langle f \rangle (x).$$

□

Corollary 3.9. *If the Lipschitz function f is optimized by the measure recurrent point x, then for each periodic point $y \in \overline{\mathcal{O}x}$, $\langle f \rangle (y) = \langle f \rangle (x)$.*

Proof. Let x be an optimal, measure recurrent point for f and let $y \in \overline{\mathcal{O}x}$ be a periodic point with period p. Since there is some subsequence $T^{n_i}x \to y$, we have that for any fixed $\delta > 0$, $\{T^i x\}$ δ-shadows y for at least p steps. Thus, $\langle f \rangle (x) - 8L\delta/p \leq \langle f \rangle (y) \leq \langle f \rangle (x)$, but this is true for all $\delta > 0$ and so $\langle f \rangle (y) = \langle f \rangle (x)$. □

[1] Since $\mathcal{M} \approx \eta M$ in the sense that $\eta M/\mathcal{M} \to 1$, we may loosely rewrite our equation as $|MD_M - (M - \mathcal{M}p)A'_M - \mathcal{M}pA| \approx M|D_M - (1-\eta p)A'_M - \eta pA| < 8(\eta M + 1)L\delta$

When studying functions optimized by aperiodic orbits, Corollary 3.9 allows us to only consider functions whose optimizing orbits contain no periodic points in their orbit closure.

Another consequence of Proposition 3.8 is the following.

Proposition 3.10. *For each Lipschitz function f, there exists a measure recurrent optimal orbit whose closure is a minimal invariant set.*

Proof. First we will establish that every continuous dynamical system has a minimal subsystem. That is, if (X, T) is a dynamical system, there exists a $Y \subset X$ so that the orbit of every $y \in Y$ is dense in Y. Let $\mathcal{S}$ represent the set of all nonempty, closed, invariant subsets of X with respect to T. We know $\mathcal{S}$ is nonempty because $X \in \mathcal{S}$. Partially order $\mathcal{S}$ by inclusion. We know that because each element of $\mathcal{S}$ is closed, if $\{A_i\} \subset \mathcal{S}$ is a chain, then $\bigcap A_i \in \mathcal{S}$. Thus every chain has a lower bound and so by Zorn's lemma, there is a minimal $Y \in \mathcal{S}$. Suppose there is some $y \in Y$ such that $\{T^i y\}$ is not dense in Y. Then $\overline{\{T^i y\}} \neq Y$, but is closed and invariant because T is continuous, a contradiction. So Y is a minimal subsystem.

Let $\langle f \rangle (x)$ be optimal and $\omega(x) = \bigcap \overline{\bigcup \{T^i x\}}$ be the ω-limit set of x. $(\omega(x), T)$ is a continuous dynamical system, so pick a minimal subsystem (Y, T) and an ergodic measure, μ_Y, supported on Y [11]. Note that the minimality of Y ensures that it contains no isolated points or that Y is a periodic orbit. By the Ergodic Theorem for almost all $y \in Y$ (as defined by μ_Y), we get that μ_Y is generated by y, so pick such a $y \in Y$ that is also measure recurrent and then we have that $\langle f \rangle (y) = \int f \, \mathrm{d}\mu_Y$. Since Y is minimal, $\mathcal{O}y$ is dense in Y, so the measure recurrence of y gives us $\omega(y) = Y$. Therefore, showing $\langle f \rangle (y) = \langle f \rangle (x)$ would complete the proof.

Fix $\varepsilon > 0$. We then have, by the Birkhoff Ergodic Theorem, that there exists an N so that for $n > N$,

$$-\varepsilon < \frac{1}{n} \sum_{i=0}^{n-1} f(T^i y) - \langle f \rangle (y) < \varepsilon.$$

Further, by minimality, there exists an $M > N$ so that $d(T^M y, y) < \varepsilon$. Construct a periodic point z that ε-shadows y along $\{y, Ty, \ldots, T^{M-1} y\}$.

Now we will show that $|\langle f \rangle (y) - \langle f \rangle (z)|$ is small. Since y follows z in order for

M steps, the In Order Lemma gives us

$$\left|\frac{1}{M}\sum_{i=1}^{M} f(T^i y) - \frac{1}{M}\sum_{i=1}^{M} f(T^i z)\right| < 4L\varepsilon.$$

But $\frac{1}{M}\sum_{i=1}^{M} f(T^i y)$ is ε-close to $\langle f\rangle(y)$, so $|\langle f\rangle(y) - \langle f\rangle(z)| < (4L+1)\varepsilon$.

To complete the proof, we will show that $\langle f\rangle(z)$ is close to $\langle f\rangle(x)$. Since $Y \subset \omega(x)$ and $\mathcal{O}x$ is dense in $\omega(x)$, we know that there is an i such that $\{T^i x, T^{i+1}x, \ldots, T^{i+M}x\}$ ε-shadows $\{y, Ty, \ldots, T^M y\}$. Therefore we have that $\{z, Tz, \ldots, T^M z\}$ 2ε-shadows $\{T^i x, T^{i+1}x, \ldots, T^{i+M}x\}$, and so by Proposition 3.8, since $\langle f\rangle(x)$ is optimal, $\langle f\rangle(z)$ is within $8L\varepsilon/M$ of $\langle f\rangle(x)$, thereby showing that

$$|\langle f\rangle(y) - \langle f\rangle(x)| < (4L + 1 + 8L/M)\varepsilon.$$

But $4L+1+8L/M$ is constant and ε was arbitrary, so $\langle f\rangle(y) = \langle f\rangle(x)$ and the orbit closure of y is a minimal, invariant set. □

3.1.1 Stability of Periodic Points

We have seen that for any function, we can get very near to the optimum integral by averaging along a periodic point, but what if our function starts out being optimized by a periodic point? As it turns out, such optimizations are stable in the sense that, if a Lipschitz function is optimized by a periodic point, there exists an open set of functions arbitrarily close to the original, all optimized by the periodic point.

Before we show that a function optimized by a periodic point can be perturbed such that it lies in an open set of functions optimized by the same point, we present a slightly simpler proposition whose proof serves as a model for the more general Proposition 3.12.

Proposition 3.11. *For every periodic orbit, there exists an open set of Lipschitz functions optimized by that orbit.*

Proof. Let y be a periodic point with period p. Define $f(x) = -d(x, \mathcal{O}y)$. Notice that f is now a Lipschitz function with constant 1 and is optimized by y. Let $G = \{f + h : \|h\|_{\text{Lip}} < \varepsilon\}$ where $\varepsilon > 0$ is to be determined later. Let $\gamma = \min_{i\neq j}\{d(T^i y, T^j y)\}$ be the smallest distance between two points in $\mathcal{O}y$, as in the In Order Lemma.

First we will establish that for small ε, in all $\gamma/4$ neighborhoods of a point $T^i y \in \mathcal{O}y$, $g \in G$ achieves a maximum precisely on $T^i y$. Consider $x \notin \mathcal{O}y$ with $d(x, T^i y) <$

$\gamma/2$ for some i. Let $y' = T^i y$. We want to show $g(y') > g(x)$. Note that since x is closer to y' than to any other point in $\mathcal{O}y$, we have that

$$g(y') = f(y') + h(y') = h(y'); \qquad g(x) = f(x) + h(x) = -d(x, y') + h(x)$$

and so

$$g(y') - g(x) = d(x, y') + h(y') - h(x).$$

But, $\|h\|_{\text{Lip}} < \varepsilon$, so $-\varepsilon d(x, y') < h(y') - h(x) < \varepsilon d(x, y')$. This gives us that

$$(1 - \varepsilon)d(x, y') < d(x, y') + h(y') - h(x) = g(y') - g(x),$$

and so for $\varepsilon < 1$, $g(y') - g(x)$ is positive, giving us that $g(y') > g(x)$ and thus y' is still the maximum.

Let x be a measure recurrent point with $\mathcal{O}x \neq \mathcal{O}y$. Note that since y is periodic, x generates a measure distinct from the measure y generates. We will now show that x is not an optimal orbit.

First, divide $\mathcal{O}x$ into in-order segments that $\gamma/4$-shadow $\mathcal{O}y$ for exactly p steps and segments we will refer to as *bad segments* that do not. By the In Order Lemma we have that in-order segments do indeed follow $\mathcal{O}y$ in order. As a consequence, if S' is an out of order segment, then for all $t \in S'$,

$$d(t, \mathcal{O}y) > \frac{\gamma}{2^{p+1}}$$

(since $d(t, y) \leq l/2$ implies that $d(Tt, Ty) \leq l$, if $d(t, \mathcal{O}y) \leq \gamma/(2^{p+1})$, then t would follow $\mathcal{O}y$ in order for p steps).

Observe that in the Shift, if x $\gamma/4$-shadows y forever, then $x \in \mathcal{O}y$. Since this was assumed not the case, bad segments must occur with some positive frequency.

Consider an in order segment $S = \{T^m x, T^{m+1}x, \ldots, T^{m+p-1}x\}$. Since S is in a $\gamma/4$ neighborhood of $\mathcal{O}y$, we have the relation

$$p \langle g \rangle (y) - \sum_{i=0}^{p-1} g(T^{i+m}x) = \sum_{i=0}^{p-1} \left[g(T^{m'+i}y) - g(T^{i+m}x) \right] > 0,$$

where m' satisfies $d(T^{m'}y, T^m x) \leq \gamma/4$.

Next, consider a bad segment. Let $T^m x$ be in a bad segment. We then have

$$\langle g \rangle (y) - g(T^m x) = \langle f + h \rangle (y) - (f + h)(T^m x) = \langle h \rangle (y) - f(T^m x) - h(T^m x)$$

$$= \langle h \rangle (y) + d(T^m x, \mathcal{O}y) - h(T^m x) \geq \gamma/(2^{p+1}) - 2\varepsilon.$$

The last inequality derives from the fact that the ε bound on the sup-norm of f ensures $\langle h \rangle (y) - h(T^m x) > -2\varepsilon$. Thus, if $\varepsilon < \gamma/(2^{p+3})$, then $\langle g \rangle (y) > g(T^m x) - \gamma/(2^{p+2})$.

Recalling that bad segments occur with positive frequency $\eta > 0$, we have

$$\langle f \rangle (y) > \langle f \rangle (x) + \frac{\eta\tau}{2^{p+1}},$$

and so $\langle f \rangle (y) > \langle f \rangle (x)$. □

Using an idea very similar to that of the proof for Proposition 3.11, we can prove a stronger statement: a function optimized by a periodic point is arbitrarily close to an open set of functions optimized by that point.

Proposition 3.12. *If f is a Lipschitz continuous function optimized by a periodic point y, then there exists an arbitrarily small perturbation $\tilde{f}$ of f (i.e. $\|\tilde{f} - f\|_{\text{Lip}}$ is small) such that there is an open set of functions containing $\tilde{f}$ all uniquely optimized by y.*

Proof. Let f be a Lipschitz continuous function with Lipschitz constant L and let y be a point of period p that optimizes f.

Fix $\sigma > 0$ and let

$$\tilde{f}(x) = f(x) - \sigma d(x, \mathcal{O}y).$$

Let $G = \{\tilde{f} + h : \|h\|_{\text{Lip}} < \varepsilon\sigma\}$ with $\varepsilon > 0$ to be determined later.

As in the proof of 3.11, let γ be the minimum distance between points in $\mathcal{O}y$.

Suppose x is a measure-recurrent optimal orbit for $g \in G$ and that $\mathcal{O}x \neq \mathcal{O}y$ and partition the orbit of x into in-order segments of length p and out-of-order segments. That is, if $T^m x$ is the start of an in-order segment, then $d(T^{i+m} x, \mathcal{O}y) \leq \gamma/4$ for $0 \leq i < p$ (which by the In Order Lemma gives us that $T^m x$ follows $\mathcal{O}y$ in order for p steps) and if $T^m x$ is in an out-of-order segment, $d(T^m x, \mathcal{O}y) > \gamma/(2^{p+1})$.

Consider an in-order segment. For ease of notation, let $\bar{d}(x) = \sigma d(x, \mathcal{O}y)$. We then have that

$$p\langle g\rangle(y)-\sum_{i=0}^{p-1}g(T^{i+m}x)=\sum_{i=0}^{p-1}\Big[f(T^iy)+h(T^iy)-\big[f(T^{i+m}x)-\bar{d}(T^{i+m}x)+h(T^{i+m}x)\big]\Big],$$

but $|h(T^{i+m}x)-h(T^iy)|\leq\varepsilon\bar{d}(T^{i+m}x)$, and $\varepsilon<1$, so $\bar{d}(T^{i+m}x)-\varepsilon\bar{d}(T^{i+m}x)\geq 0$ implies

$$p\langle g\rangle(y)-\sum_{i=0}^{p-1}g(T^{i+m}x)\geq\sum_{i=0}^{p-1}\Big[f(T^iy)-f(T^{i+m}x)\Big]=p\langle f\rangle(y)-\sum_{i=0}^{p-1}f(T^{i+m}x).$$

Next, consider T^mx to be a point in an out-of-order segment. We then have

$$\langle g\rangle(y)-g(T^mx)=\big\langle f-\bar{d}+h\big\rangle(y)-(f-\bar{d}+h)(T^mx)$$

$$=\langle f\rangle(y)-f(T^mx)+\bar{d}(T^mx)+\langle h\rangle(y)-h(T^mx),$$

but h is bounded in Lipschitz-norm and therefore in sup-norm by $\sigma\varepsilon$, so

$$|\langle h\rangle(y)-h(T^mx)|\leq 2\sigma\varepsilon.$$

This gives that

$$\langle g\rangle(y)-g(T^mx)\geq\langle f\rangle(y)-f(T^mx)+\bar{d}(T^mx)-2\sigma\varepsilon,$$

but being in an out-of-order segment supposes that $\bar{d}(T^mx)>\sigma\gamma/(2^{p+1})$, so for $\varepsilon<\gamma/(2^{p+3})$, we have

$$\langle g\rangle(y)-g(T^mx)>\langle f\rangle(y)-f(T^mx)+\frac{\sigma\gamma}{2^{p+2}}.$$

Noting that out-of-order segments occur with positive frequency, combining these two results we see

$$\langle g\rangle(y)-\langle g\rangle(x)>\langle f\rangle(y)-\langle f\rangle(x),$$

and so $\langle f\rangle(y)$ being optimum implies that the measure supported on $\mathcal{O}y$ is the optimizing measure for all functions in the open set G. $\square$

Propositions 3.11 and 3.12 summarize why we say functions optimized by periodic orbits can be stable. Of course, it is too much to ask that all functions optimized by a periodic point are stably optimized by that point. The function $f(x) = 0$ is optimized by the periodic point $y = 0$, but $\tilde{f}(x) = -\varepsilon d(x, 1)$ is an arbitrarily small perturbation of f such that $\tilde{f}$ is uniquely optimized by the measure supported on $\{1\}$. Even worse, there exists arbitrarily small perturbations of f that are optimized by measures supported on the orbit of aperiodic points. Thus, the best we can hope for is that a function optimized by a periodic point lies arbitrarily close to an open set of functions stably optimized by that periodic point.

If P is the set of all Lipschitz continuous functions optimized by a periodic point, then Proposition 3.12 gives that P contains an open dense set U, where each function in U is uniquely optimized by a measure supported on a periodic orbit. The next section will explore whether U is dense in the space of all Lipschitz continuous functions.

3.1.2 Instability of Aperiodic Points

We have seen that the set U of functions uniquely optimized by measures supported on a periodic orbit is dense in functions optimized by periodic points. In this section, we will see that a function optimized by an aperiodic point is not stable and that we can show that U is dense in portions of Lip, the space of all Lipschitz continuous functions.

In order to perturb a function optimized by an aperiodic point to a function optimized by a periodic point, we need some way of generating appropriate periodic points from aperiodic ones. This is accomplished through the use of *recursive segments*. To more easily discuss recursive segments, we will first introduce notation for general segments.

Notation 3.13. *For an aperiodic point x, the contiguous segment of $\mathcal{O}x$ from $T^a x$ to $T^{b-1}x$ is denoted by*

$$\left\{ {b \atop a} \right\} = \left\{ T^a x, T^{a+1}x, T^{a+2}x, \ldots, T^{b-2}x, T^{b-1}x \right\}.$$

It is important to note that $\left\{ {b \atop a} \right\} \subset \mathcal{O}x$ does not include $T^b x$. This is so that $\left| \left\{ {b \atop a} \right\} \right| = b - a$. Often times, we will refer to the segment $\left\{ {m+p \atop m} \right\}$ to make the length explicit.

Definition 3.14 (Recursive Segment)**.** *For a point x, the segment $\left\{ {b \atop a} \right\}$ of $\mathcal{O}x$ is called* recursive *if for all* $z', z \in \left\{ {b \atop a} \right\}$ *with* $z \neq z'$,

$$d(T^a x, T^b x) < d(z, z').$$

That is, a segment $\left\{ {b \atop a} \right\}$ *is recursive if the distance between the endpoints* $T^a x$ *and* $T^b x$ *is smaller than the distance between any points in the segment.*

Recursive segments are useful because they provide a natural place to break an aperiodic point into periodic points that closely shadow the original point. In this vein, there are several notational conventions that will be used repeatedly when referring to properties of recursive segments.

Notation 3.15. *If* $\left\{ {b \atop a} \right\}$ *is a recursive segment of* $\mathcal{O}x$*, then the distance between the endpoints of* $\left\{ {b \atop a} \right\}$ *is denoted*

$$\delta_{\left\{ {b \atop a} \right\}} = d(T^a x, T^b x).$$

The minimum distance between points in $\left\{ {b \atop a} \right\}$ *is denoted*

$$\gamma_{\left\{ {b \atop a} \right\}} = \min_{a \leq i < i' < b} d(T^i x, T^{i'} x).$$

We may now see that recursive segments are fairly common in the sense that for a point x and for any $\varepsilon > 0$, there exists a recursive segment $\left\{ {b \atop a} \right\}$ such that $\delta_{\left\{ {b \atop a} \right\}} < \varepsilon$. This is easily seen by construction: By compactness, we know that there exists i, j so that $d(T^i x, T^j x) < \varepsilon$. If $\left\{ {j \atop i} \right\}$ is not a recursive segment, then there exists i', j' with $i < i' < j' < j$ so that $d(T^{i'} x, T^{j'} x) \leq d(T^i x, T^j x)$. If $\left\{ {j' \atop i'} \right\}$ is a recursive segment, we are done. If not, repeat. Since $j - i$ is finite, this process will eventually terminate in a recursive segment with the desired properties.

Before we move on to classifying aperiodic points in terms of their recursive segments, it is worth pointing out the standard notation for the periodic point derived from a recursive segment.

Notation 3.16. *If* x *is a point and* $\left\{ {m+p \atop m} \right\}$ *is a recursive segment of* $\mathcal{O}x$*, then the periodic point derived from* $\left\{ {m+p \atop m} \right\}$*, denoted*

$$y_{\left\{ {m+p \atop m} \right\}},$$

is formed by repeating the first p *symbols of* $T^m x$*.*

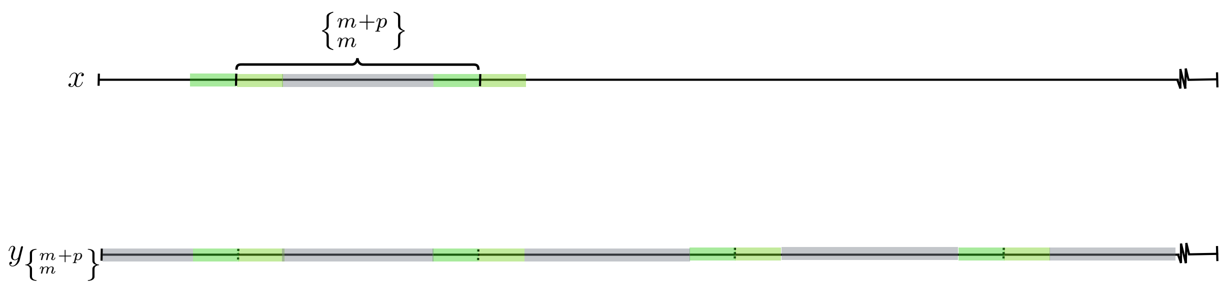

Figure 3.4: The periodic point $y_{\left\{ {m+p \atop m} \right\}}$ derived from the recursive segment $\left\{ {m+p \atop m} \right\}$. Shaded regions indicate symbols of agreement.

Notice that the derived periodic point $y_{\left\{ {m+p \atop m} \right\}}$ $\delta_{\left\{ {m+p \atop m} \right\}}$-shadows $\left\{ {m+p \atop m} \right\}$. Let k be the number of symbols that agree to either side of $T^m x$ and $T^{m+p}x$. That is, $\delta_{\left\{ {m+p \atop m} \right\}} = 2^{-k-1}$. We now have a familiar picture: $y_{\left\{ {m+p \atop m} \right\}}$ agrees to the left and right for k symbols with both $T^m x$ and $T^{m+p}x$. Further, for $0 < i < p-1$, $T^i y_{\left\{ {m+p \atop m} \right\}}$ agrees to the right and left with $T^{m+i}x$ for more than k symbols. Thus it is clear that $y_{\left\{ {m+p \atop m} \right\}}$ $\delta_{\left\{ {m+p \atop m} \right\}}$-shadows $\left\{ {m+p \atop m} \right\}$ (and consequentially, $\delta_{\left\{ {m+p \atop m} \right\}}$-shadows $\mathcal{O}x$ for p steps).

We will now classify aperiodic points in an effort to show the instability of functions optimized by such points. We need not consider the class of measure recurrent, aperiodic points x such that $\omega(x)$ contains a periodic point because of an immediate corollary to Corollary 3.9 and Proposition 3.12.

Corollary 3.17. *If x is a measure recurrent, aperiodic point and $\omega(x)$ contains a periodic point y, then if $\langle f \rangle (x)$ is optimal, there exists an arbitrarily small perturbation of f, $\tilde{f}$ such that $\tilde{f}$ is uniquely optimized by the measure supported on $\mathcal{O}y$.*

Proof. Let x be a measure recurrent, aperiodic point such that $\omega(x)$ contains a periodic point y. Further, let f be such that $\langle f \rangle (x)$ is optimal. By Corollary 3.9, $\langle f \rangle (y) = \langle f \rangle (x)$. By Proposition 3.12, there exists an arbitrarily small perturbation of f, $\tilde{f}$ such that $\tilde{f}$ is uniquely optimized by the measure supported on $\mathcal{O}y$. □

We will now classify measure recurrent, aperiodic points whose orbit closure contains no periodic points into two, possibly overlapping, classes. This categorization is due to Yuan and Hunt[13].

Definition 3.18 (Class I & II)**.** *Let x be a measure recurrent, aperiodic point where $\omega(x)$ contains no periodic points. We then say x belongs to the following categories if*

Class I: For all $Q > 0$, there exists a recursive segment $\left\{ {m+p \atop m} \right\} \subset \mathcal{O}x$ so that

$$\gamma_{\left\{ {m+p \atop m} \right\}} > Q\delta_{\left\{ {m+p \atop m} \right\}}.$$

Class II: $\{T^i x\}$ cannot be exponentially approximated by periodic orbits; i.e., for all $\Gamma > 1$, there exist $c_0 > 0$ so for any recursive segment $\left\{ {m+p \atop m} \right\} \subset \mathcal{O}x$,

$$\delta_{\left\{ {m+p \atop m} \right\}} \geq \frac{c_0}{\Gamma^p}.$$

These two classes may overlap, but the following proposition shows the categorization into class I and class II is complete.

Note, an equivalent, though slightly less elegant statement of class II, is that for all $\Gamma > 1$ there exists a $c_0 > 0$ and an $N > 0$ so that for all recursive segments $\left\{ {m+p \atop m} \right\}$ with $p > N$, then $\delta_{\left\{ {m+p \atop m} \right\}} \geq \frac{c_0}{\Gamma^p}$. Clearly, the definition of class II implies this alternate formulation. If x satisfies the alternate formulation, since N is finite and $\delta_{\left\{ {m+p \atop m} \right\}} > 0$ for all recursive segments $\left\{ {m+p \atop m} \right\}$ (as stipulated by the fact that $\omega(x)$ contains no fixed points), there must be a c_1 that satisfies

$$\delta_{\left\{ {m+p \atop m} \right\}} \Gamma^p \geq c_1$$

for all $\left\{ {m+p \atop m} \right\}$ with $p < N$. Replacing c_0 with $\min\{c_0, c_1\}$ gives the definition of class II.

Proposition 3.19. *Let x be a measure recurrent, aperiodic point such that $\omega(x)$ has no fixed points. If x is not in class I, it is in class II.*

Proof. Firstly we establish that any segment $\left\{ {n \atop m} \right\}$ of length greater than two strictly contains a recursive segment: Let $\left\{ {n \atop m} \right\}$ be a recursive segment of length greater than two. If we identify m' and n' with $m \leq m' < n' < n$ so that $d(T^{m'}x, T^{n'}x) = \gamma_{\left\{ {n \atop m} \right\}}$, then $\left\{ {n' \atop m'} \right\} \subset \left\{ {n \atop m} \right\}$ is strict. Since any segment contains a recursive segment, we have that $\left\{ {n' \atop m'} \right\}$ contains a recursive segment and so $|\left\{ {n' \atop m'} \right\}| < |\left\{ {n \atop m} \right\}|$ implies that $\left\{ {n \atop m} \right\}$ strictly contains a recursive segment.

Suppose x is not class I. Then there exists a Q so that for all recursive segments $\left\{ {m+p \atop m} \right\}$, $\gamma_{\left\{ {m+p \atop m} \right\}} \leq Q\delta_{\left\{ {m+p \atop m} \right\}}$. From the definition of recursive segment, Q is necessarily greater than 1.

Let $\left\{ {m_0+p_0 \atop m_0} \right\}$ be a recursive segment. We then know that there exists a recursive segment $\left\{ {m_1+p_1 \atop m_1} \right\} \subset \left\{ {m_0+p_0 \atop m_0} \right\}$ strictly contained in $\left\{ {m_0+p_0 \atop m_0} \right\}$ (i.e. $m_0 \leq m_1 < m_1+p_1 <$

$m_0 + p_0)$. Note that $\delta_{\left\{ {m_1+p_1 \atop m_1} \right\}} \le \gamma_{\left\{ {m_0+p_0 \atop m_0} \right\}} \le Q\delta_{\left\{ {m_0+p_0 \atop m_0} \right\}}$. Continue this process n times until $|\left\{ {m_n+p_n \atop m_n} \right\}| = 2$. We then have a chain

$$\left\{ {m_n+p_n \atop m_n} \right\} \subset \left\{ {m_{n-1}+p_{n-1} \atop m_{n-1}} \right\} \subset \cdots \subset \left\{ {m_1+p_1 \atop m_1} \right\} \subset \left\{ {m_0+p_0 \atop m_0} \right\}$$

with

$$\delta_{\left\{ {m_{i+1}+p_{i+1} \atop m_{i+1}} \right\}} \le Q\delta_{\left\{ {m_i+p_i \atop m_i} \right\}}$$

and therefore $\delta_{\left\{ {m_n+p_n \atop m_n} \right\}} \le Q^n \delta_{\left\{ {m_0+p_0 \atop m_0} \right\}}$, which gives us the suggestive inequality

$$\delta_{\left\{ {m_0+p_0 \atop m_0} \right\}} \ge \frac{\delta_{\left\{ {m_n+p_n \atop m_n} \right\}}}{Q^n}.$$

If we can show that n grows slowly enough as a function of p_0, we may conclude $\{T^i\}$ is in class II (As is, we have exponential approximation for a specific Q, however we need such a bound for all $Q > 0$. If n grows sub-linearly as a function of p_0, we will have that for an arbitrary Q' and large enough p_0, $1/Q^n > 1/(Q')^{p_0}$).

Let

$$\tau_k = \inf_{0<|i-j|<k} d(T^i x, T^j x).$$

We claim that $\tau_k \ne 0$ for any k.

Proof. If $\tau_2 = 0$, then $\omega(x)$ contains a fixed point. Suppose $\tau_n = 0$ for some $n > 2$. Let k' be the smallest value such that $\tau_{k'} = 0$. We may then find r' such that $d(T^{r'}x, T^{r'+k'}x)$ is arbitrarily small and therefore we may find r and $2 < k \le k'$ so that $\left\{ {r+k \atop r} \right\}$ is a recursive segment with $\delta_{\left\{ {r+k \atop r} \right\}}$ arbitrarily small. We then know that $\left\{ {r+k \atop r} \right\}$ strictly contains a recursive subsegment $\left\{ {m+p \atop m} \right\}$ with

$$\delta_{\left\{ {m+p \atop m} \right\}} \le Q\delta_{\left\{ {r+k \atop r} \right\}}.$$

Since Q is fixed and $\delta_{\left\{ {r+k \atop r} \right\}}$ may be made arbitrarily small, $\delta_{\left\{ {m+p \atop m} \right\}}$ may be made arbitrarily small. Since there are only a finite number of $p < k'$, by the pigeon hole principle, there exists a $p < k'$ with $\tau_p = 0$, a contradiction. □

We will now get a lower bound for $p_i - p_{i+1}$ in terms of $\delta_{\left\{ {m_i+p_i \atop m_i} \right\}}$. More specifically, we will show that for small $\delta_{\left\{ {m_i+p_i \atop m_i} \right\}}$, $p_i - p_{i+1}$ must be large.

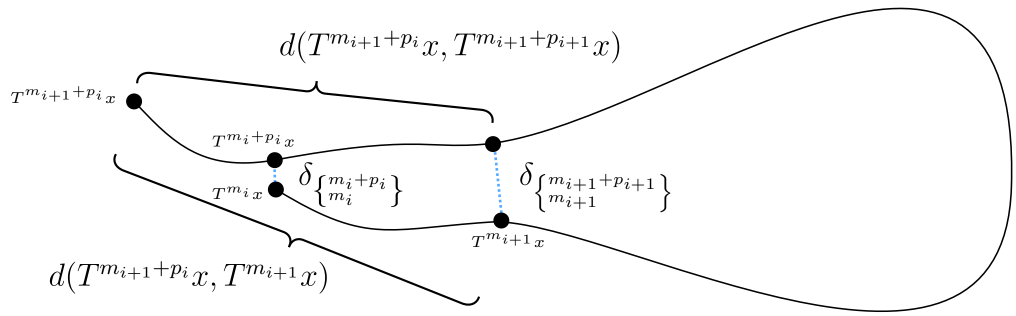

Figure 3.5: The orbit of x near $T^{m_i}x$.

Following the suggested triangle inequality in the figure, we see

$$d(T^{m_{i+1}+p_i}x, T^{m_{i+1}+p_{i+1}}x) \leq d(T^{m_{i+1}+p_i}x, T^{m_{i+1}}x) + \delta_{\left\{m_{i+1}+p_{i+1} \atop m_{i+1}\right\}},$$

but $d(T^{m_{i+1}+p_i}x, T^{m_{i+1}}x) \leq 2^{m_{i+1}-m_i}\delta_{\left\{m_i+p_i \atop m_i\right\}}$. Since $p_i - p_{i+1} \geq m_{i+1} - m_i$, we get

$$d(T^{m_{i+1}+p_i}x, T^{m_{i+1}}x) \leq 2^{p_i-p_{i+1}}\delta_{\left\{m_i+p_i \atop m_i\right\}}.$$

Further, we know that $\delta_{\left\{m_{i+1}+p_{i+1} \atop m_{i+1}\right\}} \leq Q\delta_{\left\{m_i+p_i \atop m_i\right\}}$ and so

$$d(T^{m_{i+1}+p_i}x, T^{m_{i+1}+p_{i+1}}x) \leq (2^{p_i-p_{i+1}} + Q)\delta_{\left\{m_i+p_i \atop m_i\right\}}.$$

But, $T^{m_{i+1}+p_i}x$ and $T^{m_{i+1}+p_{i+1}}x$ are exactly $p_i - p_{i+1}$ steps apart, so $\tau_{p_i-p_{i+1}} \leq d(T^{m_{i+1}+p_i}x, T^{m_{i+1}+p_{i+1}}x)$ gives us

$$\tau_{p_i-p_{i+1}} \leq (2^{p_i-p_{i+1}} + Q)\delta_{\left\{m_i+p_i \atop m_i\right\}},$$

and so

$$p_i - p_{i+1} \geq \log\left(\frac{\tau_{p_i-p_{i+1}}}{\delta_{\left\{m_i+p_i \atop m_i\right\}}} - Q\right).$$

Let $q > 1$ be arbitrary and pick δ^* so that

$$\log\left(\frac{\tau_q}{\delta^*} - Q\right) > q.$$

Consider the case where $\delta_{\left\{ m_i + p_i \atop m_i \right\}} < \delta^*$. We then get,

$$p_i - p_{i+1} \geq \log \left(\frac{\tau_{p_i - p_{i+1}}}{\delta_{\left\{ m_i + p_i \atop m_i \right\}}} - Q \right).$$

If it happens that $p_i - p_{i+1} \leq q$, then $\tau_{p_i - p_{i+1}} \geq \tau_q$, and so

$$p_i - p_{i+1} \geq \log \left(\frac{\tau_{p_i - p_{i+1}}}{\delta_{\left\{ m_i + p_i \atop m_i \right\}}} - Q \right) \geq \log \left(\frac{\tau_{p_i - p_{i+1}}}{\delta^*} - Q \right) > q,$$

a contradiction. Thus,

$$p_i - p_{i+1} > q.$$

We will now show that the number of recursive subsegments that shrink by at least q is logarithmic, thereby completing the proof.

Since we are in a compact space, for each δ, there is an $N(\delta) \in \mathbb{N}$ so that in any sequence of $N(\delta)$ points, at least two of them are δ-close. Fix q and the corresponding δ^* and $N(\delta^*)$. For any recursive segment $\left\{ m_0 + p_0 \atop m_0 \right\}$ with $p_0 > N(\delta^*)$, we have that $\delta_{\left\{ m_0 + p_0 \atop m_0 \right\}} < \gamma_{\left\{ m_0 + p_0 \atop m_0 \right\}} \leq \delta^*$.

Since $\delta_{\left\{ m_{i+1} + p_{i+1} \atop m_{i+1} \right\}} > \delta_{\left\{ m_i + p_i \atop m_i \right\}}$ and $\delta_{\left\{ m_0 + p_0 \atop m_0 \right\}} < \delta^*$, let w be the smallest number such that

$$\delta_{\left\{ m_w + p_w \atop m_w \right\}} > \delta^*.$$

We then have that the lengths of our recursive segments shrink by at least q a minimum of w times, which gives the trivial lower bound on p_0 of

$$p_0 > wq,$$

which means

$$w < \frac{p_0}{q}.$$

Further, $\delta_{\left\{ m_{i+1} + p_{i+1} \atop m_{i+1} \right\}}$ grows by at most a factor of Q with respect to $\delta_{\left\{ m_i + p_i \atop m_i \right\}}$ and so we have

$$\delta_{\left\{ m_0 + p_0 \atop m_0 \right\}} > \delta^* Q^{-w}.$$

Fix a $\Gamma > 1$. Since q was arbitrary, we may pick q such that $\Gamma \geq Q^{1/q}$. Recalling that

$w < p_0/q$ gives our final inequality,

$$\delta_{\left\{{m_0+p_0 \atop m_0}\right\}} > \frac{\delta^*}{\Gamma^{p_0}},$$

and so $\{T^i x\}$ is in class II. $\square$

We are now prepared to show the instability of functions optimized by class I points.

Proposition 3.20. *If f is a Lipschitz continuous function optimized by a measure recurrent, aperiodic point x in class I, then there exists an arbitrarily small perturbation $\tilde{f}$ of f such that $\tilde{f}$ is optimized by a periodic point.*

Proof. Let x be a measure recurrent, aperiodic point in class I and let f be a Lipschitz continuous function optimized by x. Let L be the Lipschitz constant of f. Without loss of generality, we may assume $\langle f \rangle (x) = 0$.

Fix $\varepsilon > 0$. Since x is in class I, we may find recursive segments $\left\{{m+p \atop m}\right\}$ of x such that $\gamma_{\left\{{m+p \atop m}\right\}}/\delta_{\left\{{m+p \atop m}\right\}}$ is as large as we like. Fix Q' large, to be determined later, and fix a recursive segment $\left\{{m+p \atop m}\right\}$ such that $Q = \gamma_{\left\{{m+p \atop m}\right\}}/\delta_{\left\{{m+p \atop m}\right\}} \geq Q'$. Let y be the periodic point constructed from $\left\{{m+p \atop m}\right\}$. For ease of notation, let

$$\gamma = \gamma_{\left\{{m+p \atop m}\right\}} \qquad \text{and} \qquad \delta = \delta_{\left\{{m+p \atop m}\right\}}.$$

Define the perturbation function g by

$$g(t) = d(t, \mathcal{O}y)$$

and let

$$\tilde{f} = f - \varepsilon g.$$

We will show that $\tilde{f}$ is optimized by y.

Suppose z is a measure-recurrent point with $\mathcal{O}z \neq \mathcal{O}y$. The strategy of the proof will be as follows: Either $\mathcal{O}z$ is far from $\mathcal{O}y$ quite often, in which case the penalty incurred by g will show that z is not optimal, or z will follow $\mathcal{O}y$ for long stretches before it deviates. The averages along the long, closely-shadowing stretches will be very close to the average along $\mathcal{O}y$, however, since γ/δ will be chosen to be large, the time when we do not follow $\mathcal{O}y$ closely (and consequently, in order), we will incur a penalty from g that will more than compensate for the time we spend close to $\mathcal{O}y$.

We will first handle the second case—where z follows $\mathcal{O}y$ in order for long periods of time. Ensure that γ/δ is large enough so that $\delta < \gamma/4$. We therefore know that z cannot δ-shadow $\mathcal{O}y$ forever (lest $\mathcal{O}z = \mathcal{O}y$). We will proceed by constructing a point z' that δ-shadows parts of $\mathcal{O}z$ by splicing together the segments of $\mathcal{O}z$ that are far away from $\mathcal{O}y$ (more than $\gamma/4$ away).

In order to appropriately splice together segments to ensure a true δ-shadow, we will cut $\mathcal{O}z$ at places that are within $\delta/2$ of $\mathcal{O}y$. Since the goal in constructing z' is to create a segment that is as far from $\mathcal{O}y$ as possible as often as possible, it will be useful to get an upper bound on how many steps $T^i z$ needs to be close to $\mathcal{O}y$ before we make the splice (We know if we follow $\mathcal{O}y$ in order, we will get geometrically close and so this will give us a logarithmic upper bound).

Let ℓ be the largest number of steps such that there exists a j with $T^{j+i}z$ remaining farther than $\delta/(2^{p+2})$ from $\mathcal{O}y$ but closer than $\gamma/4$ from $\mathcal{O}y$ with i varying from 0 to $\ell - 1$. More precisely,

$$\ell = \max\left\{ l : \exists\, j \text{ with } \frac{\delta}{2^{p+2}} < d(T^{j+i}z, \mathcal{O}y) \leq \frac{\gamma}{4} \text{ for } 0 \leq i < l \right\}.$$

The In Order Lemma gives us that if we $\gamma/4$-shadow $\mathcal{O}y$ for ℓ steps,

$$d(T^{j+k}z, \mathcal{O}y) \leq 2^{-\min\{k, \ell-k\}} \frac{\gamma}{4}.$$

Since $d(T^{j+k}z, \mathcal{O}y) > \delta/(2^{p+2})$ we have that $\delta/(2^{p+2}) \leq 2^{-\left\lfloor \frac{\ell}{2} \right\rfloor} \frac{\gamma}{4}$ and so

$$\ell < 2 \log\left(\frac{\gamma}{\delta}(2^{p+1})\right) \leq 2 \log Q + 2(p+1).$$

We have chosen ℓ such that if $T^i z$ $\gamma/4$-shadows $\mathcal{O}y$ for ℓ steps, $T^i z$ $\delta/2$-shadows $\mathcal{O}y$ for an entire period. We are now ready to define z'.

As suggested by the figure, construct minimal segments W_j with the following properties: Each W_j contains a point that is more than $\gamma/4$ away from $\mathcal{O}y$; the size of W_j is a multiple of p; for each j there exists a $y_j \in \mathcal{O}y$ such that the right endpoint of W_j and the left endpoint of W_{j+1} both come within $\delta/2$ of y_j; and, the collection of W_j's contains all points of $\mathcal{O}z$ farther than $\gamma/4$ from $\mathcal{O}y$.

Recall that we have assumed that z follows $\mathcal{O}y$ in order for long stretches, which is now taken to mean there are segments of $\mathcal{O}z$ that $\gamma/4$-shadow $\mathcal{O}y$ for at least 2ℓ steps. Since $\gamma/4$-shadowing $\mathcal{O}y$ for ℓ steps ensures that you $\delta/2$-shadow $\mathcal{O}y$ for an

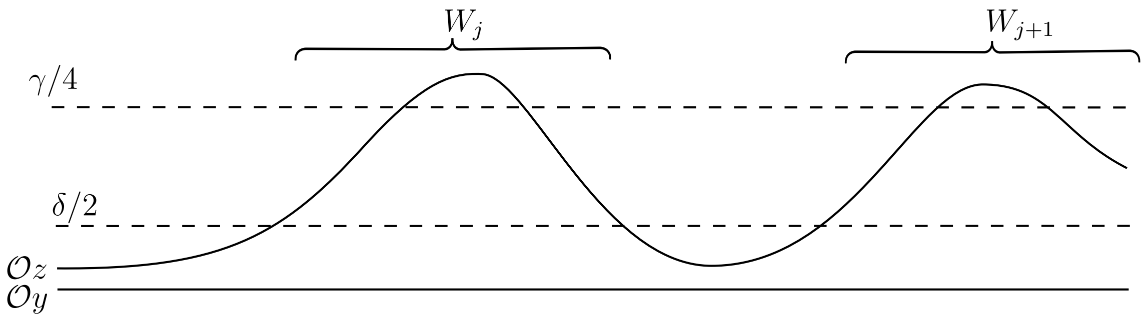

Figure 3.6: The distance between $\mathcal{O}y$ and $\mathcal{O}z$.

entire period and $|\mathcal{O}y|$ is finite, we know by the pigeon hole principle that we can construct W_j's with the desired properties. A further consequence of this argument is that the consecutive amount of time W_j spends within $\gamma/4$ of $\mathcal{O}y$ is bounded by ℓ. That is, every subsegment of W_j longer than length ℓ must contain a point farther than $\gamma/4$ from $\mathcal{O}y$.

Note that the construction of the W_j's is not unique, but once one W_j is fixed, the rest are determined.

Next, label the segment between W_{j-1} and W_j by S_j. Note that by construction both $|W_j|, |S_j| \in p\mathbb{Z}^+$.

We will now define z' as the concatenation of the leading symbol of each point in W_j. We will use adjacency to indicate this type of concatenation. In this notation we have

$$z' = \cdots W_{-1}W_0W_1\cdots \qquad \text{and} \qquad z = \cdots S_{-1}W_{-1}S_0W_0S_1W_1\cdots .$$

Notice that we have constructed z' in such a way that z' δ-shadows parts of $\mathcal{O}z$ and remains δ-close to $\mathcal{O}z$ at all times. Further, points T^iz' such that $d(T^iz', \mathcal{O}y) > \gamma/4$ exist with a frequency at least $\frac{1}{2\ell+1} > \frac{1}{4\ell}$ portion of the time and so

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} g(T^iz') > \frac{\gamma}{16\ell}.$$

We must take a lim inf in this case since $\langle g\rangle(z')$ may not exist because z' may not be measure recurrent, but since this notation is convenient, we will extend notation by defining

$$\langle F \rangle_{\inf}(z') = \liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} F(T^i z')$$

and

$$\langle F \rangle (z') = \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} F(T^i z')$$

for all functions F. Thus, with the extended notation, $\langle g \rangle_{\inf}(z') > \frac{\gamma}{16\ell}$. Recall that $\langle f \rangle (y) > -8L\delta/p$ and so we have

$$\begin{aligned}\langle f \rangle (z') - \varepsilon \langle g \rangle_{\inf}(z') - \langle f \rangle (y) &< \langle f \rangle (z') - \varepsilon \langle g \rangle_{\inf}(z') + \frac{8L\delta}{p} \\ &\leq \langle f \rangle (z') - \frac{\varepsilon\gamma}{16\ell} + \frac{8L\delta}{p} \\ &\leq \frac{8L\delta}{p} - \frac{\varepsilon\gamma}{16\ell}.\end{aligned}$$

Since z' follows $\mathcal{O}z$ in order along W_j, for each j, the In Order Lemma gives us

$$-4L\delta \leq \sum_{i\in W_j} \tilde{f}(T^i z) - \tilde{f}(T^{i'} z') \leq 4L\delta,$$

and since z δ-shadows $\mathcal{O}y$ in order along S_j, the In Order Lemma gives us

$$-4L\delta \leq \sum_{i\in S_j} \tilde{f}(T^i z) - \tilde{f}(T^{i'} y) \leq 4L\delta.$$

Recall that we are using simplifying notation: the i' corresponding to $i \in S_j$ is the

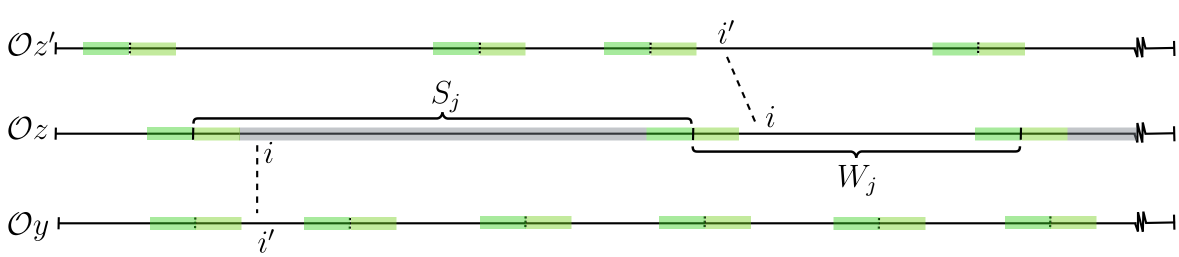

Figure 3.7: Picture showing the correspondence between i and i' for $i \in S_j$ and $i \in W_j$.

unique i' such that $d(T^i z, T^{i'} y) \leq 2^{-i-1}$ and i, i' when $i \in W_j$ means $i \in W_j$ and the i' such that $T^{i'} z'$ corresponds to the point in $\mathcal{O}z'$ created from $T^i z$.

Rewriting $\left\langle \tilde{f} \right\rangle (z')$ as a limit, we see

$$\limsup_{j\to\infty} \left[\frac{\displaystyle\sum_{i' \in W_{-j},\ldots,W_j} \tilde{f}(T^{i'} z')}{|W_{-j}| + \cdots + |W_j|} \right] - \left\langle \tilde{f} \right\rangle (y) < \frac{8L\delta}{p} - \frac{\varepsilon\gamma}{16\ell}.$$

Notice that $\left| \sum_{i \in W_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} z') \right] \right| \leq 4L\delta$ gives

$$\left| \sum_{i \in W_{-j},\ldots,W_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} z') \right] \right| \leq (2j+1)4L\delta,$$

and because W_j were chosen such that $|W_j| \geq p$, we get

$$\limsup_{j\to\infty} \frac{1}{|W_{-j}| + \cdots + |W_j|} \left| \sum_{i \in W_{-j},\ldots,W_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} z') \right] \right| \leq \frac{4L\delta}{p}.$$

Thus, by the triangle inequality we may conclude

$$\limsup_{j\to\infty} \left[\frac{\displaystyle\sum_{i \in W_{-j},\ldots,W_j}}{|W_{-j}| + \cdots + |W_j|} \tilde{f}(T^i z) \right] - \left\langle \tilde{f} \right\rangle (y) < \frac{12L\delta}{p} - \frac{\varepsilon\gamma}{16\ell}.$$

Analyzing the average along multiple S_j segments, we see that since

$$\sum_{i \in S_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} y) \right] \leq 4L\delta,$$

$$\sum_{i \in S_{-j},\ldots,S_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} y) \right] \leq b4L\delta,$$

where $b = \left| \{ |S_r| > 0 : -j \leq r \leq j \} \right| \leq 2j+1$ (if $|S_r| = 0$, we clearly get no contribution from the $\sum_{i \in S_r}$ term). Since $|S_r|$ is a multiple of p, we have $|S_r| < p$ implies $|S_r| = 0$. Thus, there are at least b S_r's with $|r| \leq j$ satisfying $|S_r| \geq p$. This gives us

$$\frac{1}{|S_{-j}| + \cdots + |S_j|} \sum_{i \in S_{-j},\ldots,S_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} y) \right] \leq \frac{4L\delta}{p},$$

and so pulling out $(|S_{-j}|+\cdots+|S_j|)\left\langle \tilde{f}\right\rangle(y)$ from the sum gives us

$$\frac{1}{|S_{-j}|+\cdots+|S_j|}\left[\sum_{i\in S_{-j},\ldots,S_j}\tilde{f}(T^i z)\right]-\left\langle \tilde{f}\right\rangle(y)\leq\frac{4L\delta}{p}.$$

Reintroducing the S_j segments and combining our previous equations, we see

$$\left\langle \tilde{f}\right\rangle(z)-\left\langle \tilde{f}\right\rangle(y)=\lim_{j\to\infty}\left[\frac{\sum_{i\in W_{-j}S_{-j}\cdots S_jW_j}\tilde{f}(T^i z)}{|W_{-j}S_{-j}\cdots S_jW_j|}\right]-\left\langle \tilde{f}\right\rangle(y)<\frac{16L\delta}{p}-\frac{\varepsilon\gamma}{16\ell},$$

and so $\frac{32L\delta}{p}-\frac{\varepsilon\gamma}{16\ell}<0$ would prove the non-optimality of z. Notice that

$$\frac{16L\delta}{p}-\frac{\varepsilon\gamma}{16\ell}<0$$

is satisfied if

$$16L-\frac{p\varepsilon\gamma}{\delta\ell}=16L-\frac{pQ}{\ell}\varepsilon<0$$

is satisfied. Recalling that $\ell<2\log Q+2(p+1)$, we see that for large enough Q (that is for Q satisfying $\log Q>1$),

$$\frac{p}{\ell}>\frac{p}{2\log Q+2(p+1)}>\frac{1}{8\log Q}.$$

The last inequality can be deduced by the following argument: Observe that if a,b,p satisfy $a,pb\geq 1$ then $\frac{p}{a+pb}\geq\frac{p}{a+pba}\geq\frac{p}{pba+pba}=\frac{1}{2ba}$. Thus, since $\frac{p}{2\log Q+2(p+1)}>\frac{p}{4\log Q+2p}=\frac{p}{2(2\log Q+p)}$, we see $\log Q,p>1$ gives the inequality.

Since $Q/\log Q$ is unbounded in Q and $\varepsilon/8$ is constant, for large enough Q we have that

$$0>16L-\frac{Q}{8\log Q}>16L-\frac{pQ}{\ell}\varepsilon,$$

and so $\left\langle \tilde{f}\right\rangle(z)-\left\langle \tilde{f}\right\rangle(y)<0$.

Finally, we must show that if $\mathcal{O}z$ is far from $\mathcal{O}y$ quite often, then $\left\langle \tilde{f}\right\rangle(z)-\left\langle \tilde{f}\right\rangle(y)<0$. This condition means that there are no segments of $\mathcal{O}x$ that $\gamma/4$-shadow $\mathcal{O}y$ for a length of at least 2ℓ. Thus, T^iz is farther than $\gamma/4$ with a frequency of at least $1/(2\ell+1)>1/(4\ell)$. We now see z fulfills the same properties that z' did in the preceding proof, and so we have

$$\left\langle \tilde{f} \right\rangle (z) - \left\langle \tilde{f} \right\rangle (y) \leq \frac{8L\delta}{p} - \frac{\varepsilon\gamma}{16\ell} < \frac{16L\delta}{p} - \frac{\varepsilon\gamma}{16\ell},$$

which we have already seen is less than zero for large enough Q, completing the proof. □

Now that we see that we may perturb a function optimized by a point in class I to one optimized by a periodic point, it is a simple application of Proposition 3.12 to derive the following corollary.

Corollary 3.21. *If f is a Lipschitz continuous function optimized by a measure recurrent, aperiodic point x in class I, then there exists an arbitrarily small perturbation $\tilde{f}$ of f such that there is an open set of functions containing $\tilde{f}$ all optimized by y.*

Proof. Fix $\varepsilon > 0$. By Proposition 3.20 we may perturb f to a function f' with $\|f - f'\|_{\text{Lip}} < \varepsilon/2$ such that f' is optimized by a periodic point y. By Proposition 3.12 we may perturb f' to a function $\tilde{f}$ with $\|f' - \tilde{f}\|_{\text{Lip}} < \varepsilon/2$ such that there exists an open set of functions containing $\tilde{f}$ that are all optimized by y. Applying the triangle inequality completes the proof. □

We will now turn our attention to functions optimized by points in class II. The definition of class II given by Yuan and Hunt[13] is suitable for any dynamical system. We will give alternative classifications of class II applicable to a shift space in terms of the following functions.

Notation 3.22. *For some point x, let $R_x(p)$ be the length of the longest repeated sub-word whose starting positions occur within the same p-block.*

$$R_x(p) = \sup_{0<|i-j|\leq p} \{k : (x)_i^{i+k-1} = (x)_j^{j+k-1}\}.$$

Let $G_x(l)$ be the smallest gap between matches of at least l symbols. That is,

$$G_x(l) = \min_{i \neq j} \{|i - j| : (x)_i^{i+l-1} = (x)_j^{j+l-1}\}.$$

Using the R and G notation, we present the following alternative classification of class II.

Proposition 3.23. *For a measure recurrent, aperiodic point x that contains no periodic points in its orbit closure, the following are equivalent:*

1. $\{T^i x\}$ *is in class II.*
2. $R_x(p)/p \to 0$ *as* $p \to \infty$.
3. $G_x(l)/l \to \infty$ *as* $l \to \infty$.

Proof. First note that because $\omega(x)$ contains no periodic points, $R_x(p) < \infty$ for all p. Now we will show the equivalence of 2 and 3. We claim that if $G_x(l)$ is the smallest distance between agreement of l symbols, then

$$R_x(G_x(l)) = l.$$

Since in some $G_x(l)$-block, we have at least l symbols of agreement and so $R_x(G_x(l)) \geq l$. Further, if $R_x(G_x(l)) > l$, then there would be some block of length $G_x(l) - 1$ with l symbols of agreement, contradicting the minimality of $G_x(l)$.

Similarly,

$$G_x(R_x(p)) \leq p,$$

since $R_x(p)$ is the largest number of symbols of agreement in some p-block, $G_x(R_x(p))$ cannot possibly be larger than p. Note that $G_x(R_x(p)) = p$ whenever R_x makes a jump; i.e., whenever $R_x(p) - R_x(p-1) \geq 1$.

We will first show that both $G_x(t), R_x(t) \to \infty$ as $t \to \infty$. Since $R_x(t)$ is the length of the longest sub-word that re-occurs in a t-block, by the pigeon hole principle, $R_x(t) \to \infty$. The proof that $G_x(t) \to \infty$ relies on the fact that $\omega(x)$ contains no periodic points. Suppose that G_x were bounded. Since G_x is non-decreasing, this means that eventually $G_x(t) = M$ for all large t. Thus, for every large $t > M$, there is a string of symbols of length t that repeats with period M. Because x is measure-recurrent, taking t larger and larger, we find that $\omega(x)$ must contain a point of period M (though the minimal period may be some divisor of M).

Now suppose that $R_x(t)/t \to 0$. Then, since $G_x(t) \to \infty$,

$$\frac{R_x(G_x(t))}{G_x(t)} = \frac{t}{G_x(t)} \to 0$$

and so $G_x(t)/t \to \infty$. Suppose $G_x(t)/t \to \infty$, then $t/G_x(t) \to 0$ and so

$$\frac{R_x(t)}{G_x(R_x(t))} \to 0,$$

but $G_x(R_x(t)) \leq t$ and so $\frac{R_x(t)}{G_x(R_x(t))} \geq \frac{R_x(t)}{t}$, which shows $R_x(t)/t \to 0$.

Now we will show the equivalence of class II with conditions 2 and 3. Suppose $\{T^i x\}$ is in class II, then for all $\Gamma > 1$, there is an N and c_0 so that for any recursive segment $\{{m+p \atop m}\}$ with $p > N$,

$$\delta_{\{{m+p \atop m}\}} \geq \frac{c_0}{\Gamma^p}.$$

For a recursive segment $\{{m+p \atop m}\}$, let $k_{\{{m+p \atop m}\}}$ be the number of symbols of agreement between the endpoints $T^m x$ and $T^{m+p} x$ of $\{{m+p \atop m}\}$. That is,

$$k_{\{{m+p \atop m}\}} = -\log \delta_{\{{m+p \atop m}\}}.$$

Notice that class II gives us

$$k_{\{{m+p \atop m}\}} = -\log \delta_{\{{m+p \atop m}\}} \leq p \log \Gamma + \log c_0$$

and so $k_{\{{m+p \atop m}\}}/p < \log \Gamma + \frac{1}{p} \log c_0$. Since this is true for all recursive segments of length p, we see that because

$$k_{\{{m+p \atop m}\}} + 1 \leq G_x(p) = \max_n \left\{ k_{\{{n+p \atop n}\}} + 1 \right\},$$

we have

$$\frac{G_x(p)}{p} \leq \log \Gamma - \frac{1}{p} \log c_0.$$

Fix $\varepsilon > 0$. Since Γ was arbitrary, we may choose Γ close enough to 1 and p large so that $\log \Gamma - \frac{1}{p} \log c_0 < \varepsilon$, and so for large enough p,

$$\frac{G_x(p)}{p} < \varepsilon.$$

Thus $G_x(p)/p \to 0$.

Similarly, if $G_x(p)/p \to 0$, for any Γ we may choose N so that $p > N$ implies

$$\frac{G_x(p)}{p} \leq \log \Gamma.$$

Thus we have

$$-\log \delta_{\{{m+p \atop m}\}} \le k_{\{{m+p \atop m}\}} \le G_x(p) \le \log \Gamma^p.$$

Exponentiating both sides gives us the class II condition. □

We are now prepared to give an instability result for functions optimized by points in class II. Unlike Proposition 3.20, this proposition does not show that arbitrarily small perturbations of a function optimized by a class II point become optimized by a periodic point, but it does show that arbitrarily small perturbations of a function optimized by a class II point ensure the function is no longer optimized by the original point.

Proposition 3.24. *If $\{T^i x\}$ is a measure recurrent, aperiodic orbit in class II with no periodic points in $\omega(x)$ and if f is optimized by x, then there exists an arbitrarily small Lipschitz perturbation $\tilde{f}$ of f such that $\left\langle \tilde{f} \right\rangle (x)$ is not optimal.*

Proof. Let f be a Lipschitz continuous function optimized by the class II point x. For a periodic point $y_{\{{m+p \atop m}\}}$ derived from the recursive segment $\{{m+p \atop m}\}$ of x, define the function

$$g_{\{{m+p \atop m}\}}(z) = d(z, \mathcal{O}y_{\{{m+p \atop m}\}}).$$

We will proceed by contradiction to show that if all perturbations $f - \varepsilon g_{\{{m+p \atop m}\}}$ are still optimized by x, then $\{T^i x\}$ is not in class II.

Suppose there exists an $\varepsilon > 0$ such that for any recursive segment $\{{m+p \atop m}\}$,

$$\left\langle f - \varepsilon g_{\{{m+p \atop m}\}} \right\rangle (x) \ge \left\langle f - \varepsilon g_{\{{m+p \atop m}\}} \right\rangle (y_{\{{m+p \atop m}\}}).$$

For ease of notation, let $g = g_{\{{m+p \atop m}\}}$, $y = y_{\{{m+p \atop m}\}}$, and $\delta = \delta_{\{{m+p \atop m}\}}$. From Proposition 3.8 we have that

$$\langle f \rangle (y) - \langle f \rangle (x) > -\frac{8L\delta}{p}.$$

By construction of g we have

$$\langle g \rangle (y) - \langle g \rangle (x) = - \langle g \rangle (x),$$

and so

$$\langle f - \varepsilon g \rangle (y) - \langle f - \varepsilon g \rangle (x) > \varepsilon \langle g \rangle (x) - \frac{8L\delta}{p}.$$

Thus, since we assume $f - \varepsilon g$ was optimized by x, $\langle f - \varepsilon g \rangle (y) - \langle f - \varepsilon g \rangle (x) \le 0$

implies

$$\langle g \rangle (x) < \frac{8L\delta}{\varepsilon p}.$$

Since $g(z) = d(z, \mathcal{O}y)$, this means that the average distance of $T^i x$ from $\mathcal{O}y$ is very small.

Let k_δ be the number of symbols that agree to either side of the endpoints of $\left\{ {m+p \atop m} \right\}$. That is, $\delta = 2^{-k_\delta - 1}$. We will now find an upper bound for $G_x(2k_\delta + 3)$. That is, we will find an upper bound on n such that there exists some i so $d(T^i x, T^{i+n} x) < 2^{-k_\delta - 2}$. In plainer terms, we are getting a bound on how far we need to search to find agreement that is one symbol better than the agreement between the endpoints of $\left\{ {m+p \atop m} \right\}$.

Let $A = 16L/\varepsilon$. We have now established a general result about any g (recall that for each recursive segment, there exists a possibly different g; further, k_δ also depends on the recursive segment), namely

$$\frac{p}{A} \langle g \rangle (x) < 2^{-k_\delta - 2}.$$

Since $16L$ and ε are fixed, for large enough p, $p > A$.

We will now consider only g's generated by recursive segments with $p > A$. Note that the choice of recursive segment also affects y and k_δ. Because $\langle g \rangle (x)$ is the average distance of $T^i x$ from $\mathcal{O}y$ and $\langle g \rangle (x) < \frac{A}{p} 2^{-k_\delta - 2}$, we have that $d(T^i x, \mathcal{O}y) > 2^{-k_\delta - 2}$ at most an A/p portion of the time and so $d(T^i x, \mathcal{O}y) < 2^{-k_\delta - 2}$ occurs at least $1 - A/p$ portion of the time. Thus, there must be a segment of $\mathcal{O}x$ with length no more than $w = (p+1)/(1 - A/p)$ that contains $p+1$ points within $2^{-k_\delta - 2}$ of $\mathcal{O}y$. By the pigeon hole principle, this means that there exist i, j with $0 < |i - j| < w$ such that $d(T^i x, T^j x) < 2^{-k_\delta - 2}$.

Since A is fixed, we have that $w < p + A + 2$ for p large enough. This can be seen by taking the Taylor expansion: $w = (p+1)(1 + A/p + O(1/p^2)) = p + 1 + A + O(1/p)$ and so for large enough p, the inequality is ensured. To complete the proof, notice that for any p, we may find a recursive segment $\left\{ {m+p \atop m} \right\}$ so that k_δ is maximal. That is,

$$p = G_x(2k_\delta).$$

We have shown that in this case, there are points in $\mathcal{O}x$ no more than $w < p + A + 2$

steps apart that agree for $2(k_\delta + 1) = 2k_\delta + 2$ symbols. Thus,

$$G_x(2k_\delta + 3) \leq p + A + 2,$$

and so the derivative of G_x is bounded. This gives $G_x(t)/t < \infty$, and so x is not in class II. □

Our results about Lipschitz functions optimized by aperiodic points may be summarized by instability: if f is optimized by an aperiodic point x, there is an arbitrarily small perturbation $\tilde{f}$ of f such that $\tilde{f}$ is not optimized by x. This is true for all aperiodic points x. However, if we restrict to the aperiodic points in class I, we get the stronger result that there are arbitrarily small perturbations $\tilde{f}$ such that there is an open set of functions containing $\tilde{f}$ that is optimized by a periodic point, thus showing that the set of Lipschitz functions uniquely optimized by invariant measures supported on periodic points contains an open set that is dense in functions optimized by periodic points and aperiodic points in class I.

3.2 Summable Variation

We will now extend the results about Lipschitz continuous functions to the more general class of functions of summable variation, first explored in the context of ergodic theory by Walters in [10].

Definition 3.25 (Variation)**.** *We say the variation of a function over a range of I symbols is the maximum a function changes in a distance of 2^{-I}. That is, if f is a function*

$$\text{var}_I(f) = \sup\{|f(x) - f(y)| : d(x, y) \leq 2^{-I}\}.$$

With this notation we see that a Lipschitz continuous function f with Lipschitz constant L may be defined by the property that $\text{var}_I(f) \leq L2^{-I}$. Note that in a shift space, we have special freedom because distances can only take values of the form 2^{-k}.

Definition 3.26 (Summable Variation)**.** *We say that a function f is of summable variation if*

$$\sum_{I=0}^{\infty} \text{var}_I(f) < \infty.$$

We will also introduce some simplifying notation for discussing the tail sum of the variation of a function.

Notation 3.27. $\mathrm{V}_k(f)$ *represents the tail sum of the variation of f over distances smaller than 2^{-k}. That is*

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

Functions of summable variation form a much larger class than Lipschitz functions. Since the geometric sequence 2^{-k} is summable, we see that all Lipschitz functions are functions of summable variation, but there are many summable sequences that are not geometric.

Recall that the In Order Lemma gave us that if x followed y in order for some time, x would get geometrically close to y. This is equivalent to x agreeing with y for more and more symbols. Thus, just as geometric closeness gives geometric summability for the difference between two Lipschitz functions, we get a very similar property for the sum of the differences between functions of summable variation along x and y.

Lemma 3.28 (Parallel Orbit Lemma B)**.** *For a function of summable variation f, if x 2^{-r}-shadows a point y for k steps, then for $r > 0$,*

$$\sum_{i=0}^{k-1} \left| f(T^{m+i}x) - f(T^{m'+i}y) \right| \leq 2\mathrm{V}_r(f).$$

Proof. Suppose x, y are points such that x 2^{-r}-shadows $\mathcal{O}y$ for k steps. For $r > 0$, expansivity gives us that for some segment $\{T^m x, T^{m+1}x, \ldots, T^{m+k-1}x\}$, there exists an m' so that

$$d(T^{m+i}x, T^{m'+i}y) \leq 2^{-\min\{i,k-i\}-r}.$$

We then have

$$\sum_{i=0}^{k-1} |f(T^{m+i}x) - f(T^{m'+i}y)| \leq 2 \sum_{I=r}^{r+\lceil k/2 \rceil} \mathrm{var}_I(f) \leq 2\mathrm{V}_r(f).$$

□

The Parallel Orbit Lemma B is the summable-variation analog of the Parallel Orbit Lemma.

We can now derive the analog of Proposition 3.8 for summable variation.

Proposition 3.29. *Let f be a function of summable variation, $\{T^i x\}$ a measure recurrent, optimal orbit for f, y a point of period p, and $r > 0$. If $\{T^i x\}$ 2^{-r}-shadows $\mathcal{O}y$ for one period (i.e., there exists an m such that $d(T^{i+m}x, T^i y) \leq 2^{-r}$ for $0 \leq i < p$), then*

$$\langle f \rangle (x) - 4\mathrm{V}_r(f)/p \leq \langle f \rangle (y) \leq \langle f \rangle (x).$$

The proof of Proposition 3.29 is essentially the same as the proof for Proposition 3.8 with $L\delta$ replaced by $\mathrm{V}_r(f)$. Similarly, with the same method of proof we get the analog of Corollary 3.9.

Corollary 3.30. *If the function of summable variation f is optimized by the measure-recurrent point x, then for each periodic point $y \in \omega(x)$, $\langle f \rangle (y) = \langle f \rangle (x)$.*

Thus, it initially appears that all the propositions stated in terms of Lipschitz functions directly apply to those of summable variation. However, Lipschitz functions have the distinguishing property that there exists a "sharpest" function. That is, for $L = 1$, the function $g(z) = d(z, \mathcal{O}y)$ lies above all other Lipschitz functions (with $L \leq 1$) that are zero on $\mathcal{O}y$. In the space of functions of summable variation, even when putting a bound on $V_0(f)$, there is no such "sharpest" function. Thus, we will introduce a family of norms on subsets of the set of functions of summable variation such that we may exploit the existence of a "sharpest" function.

To this end, we will introduce a new norm.

Definition 3.31. *Let $A = \{A_0, A_1, \ldots\}$ be a monotone-decreasing, positive sequence. We define the A-norm of f by*

$$\|f\|_A = \sup_I \left\{ \frac{\mathrm{var}_I f}{A_I} \right\} + \|f\|_\infty.$$

Proposition 3.32. *The A-norm is a norm.*

Proof. Since $A_i \neq 0$, we have that $\|f\|_A$ is well defined. Since A is a positive sequence and $\mathrm{var}_I f \geq 0$ for all I, $\|f\|_A \geq 0$, so the A-norm is positive. It is clear that if $f \neq 0$, $\|f\|_\infty \neq 0$ and so $\|f\|_A \neq 0$, so the A-norm is definite. Further, since $\mathrm{var}_I af = |a|\mathrm{var}_I f$ for any $a \in \mathbb{R}$, we have that $\|af\|_A = |a|\|f\|_A$. Lastly, note that $\mathrm{var}_I(f + g) \leq \mathrm{var}_I f + \mathrm{var}_I g$, which gives that $\|f + g\|_A \leq \|f\|_A + \|g\|_A$. □

Notice that if $A_i = 2^{-i}$ then $\|f\|_A = \|f\|_{\mathrm{Lip}}$.

A-norms also have the nice property that they generate Banach spaces.

Proposition 3.33. *If A is a sequence satisfying the properties of an A-norm, then $X = \{f : \|f\|_A < \infty\}$ is a Banach space.*

Proof. Let A be a sequence satisfying the properties of an A-norm and let $X = \{f : \|f\|_A < \infty\}$. Let f_n be a Cauchy sequence in X (with respect to $\|\cdot\|_A$). Since $\|f\|_\infty \leq \|f\|_A$ and f_n is a Cauchy sequence in $\|\cdot\|_A$, f_n is a Cauchy sequence in $\|\cdot\|_\infty$ and so there exists an f such that $f_n \to f$ pointwise.

For two points x, y, let $k_{x,y}$ be such that $d(x,y) = 2^{-k_{x,y}}$. Fix $\varepsilon > 0$. We then have that there exists an M such that for all $n \geq M$, $\|f_n - f_M\|_A < \varepsilon$. This gives us gives us

$$\frac{|(f_n - f_M)(x) - (f_n - f_M)(y)|}{A_{k_{x,y}}} < \varepsilon \tag{3.4}$$

for all x, y. Since $f_n \to f$ pointwise, letting $n \to \infty$, inequality (3.4) gives us $\sup_I \frac{\mathrm{var}_I(f - f_M)}{A_I} \leq \varepsilon$ and so $\|f - f_M\|_A \leq \varepsilon$. Thus, we may conclude $\|f\|_A \leq \|f_M\|_A + \varepsilon$ and so $f \in X$. □

We can now state the analog of Proposition 3.12.

Proposition 3.34. *If f is in the Banach space generated by some A-norm and is optimized by a periodic point y, then there exists an arbitrarily small perturbation $\tilde{f}$ of f (i.e. $\|\tilde{f} - f\|_A$ is small) such that there is an open set of functions containing $\tilde{f}$ all uniquely optimized by the periodic orbit measure generated by y.*

Proof. Let f be in the Banach space generated by the A-norm $\|\cdot\|_A$, and let y be a point of period p that optimizes f.

We first define our "sharp" function g by

$$g(x) = A_n \quad \text{if } d(x, \mathcal{O}y) = 2^{-n}.$$

Fix $\sigma > 0$ and let

$$\tilde{f}(x) = f(x) - \sigma g(x).$$

Let $Q = \{\tilde{f} + h : \|h\|_A < \varepsilon\sigma\}$ with $\varepsilon > 0$ to be determined later.

As usual, let γ be the minimum distance between points in $\mathcal{O}y$.

Suppose x is a measure-recurrent optimal orbit for $q = f - \sigma g + h \in Q$ (and so $\|h\|_A < \varepsilon\sigma$) and that $\mathcal{O}x \neq \mathcal{O}y$ and partition the orbit of x into in-order segments of length p and segments that we will call *bad segments*. That is, if $T^m x$ is the start of an in-order segment, then $d(T^{i+m}x, \mathcal{O}y) \leq \gamma/4$ for $0 \leq i < p$ (which by the In Order

Lemma gives us that $T^m x$ follows $\mathcal{O}y$ in order for p steps) and if $T^m x$ is in a bad segment, $d(T^m x, \mathcal{O}y) > \gamma/(2^{p+1})$.

Consider an in-order segment. We then have that

$$p\langle q\rangle(y) - \sum_{i=0}^{p-1} q(T^{i+m}x) = \sum_{i=0}^{p-1}\Big[f(T^i y) + h(T^i y) - \big[f(T^{i+m}x) - \sigma g(T^{i+m}x) + h(T^{i+m}x)\big]\Big].$$

Notice that since we are in an in-order segment, $\sigma g(T^{i+m}x) = \sigma A_{|\log d(T^{i+m}x, T^i y)|}$. Further, since $\|h\|_A < \varepsilon\sigma$ we have that $\mathrm{var}_I h < \varepsilon\sigma A_I$. Thus $|h(T^{i+m}x) - h(T^i y)| \leq \varepsilon\sigma A_{|\log d(T^{i+m}x, T^i y)|}$. So, for $\varepsilon < 1$, we have $[h(T^i x) - h(T^{i+m}y)] + \sigma g(T^{i+m}x) \geq 0$, which implies

$$p\langle q\rangle(y) - \sum_{i=0}^{p-1} q(T^{i+m}x) \geq \sum_{i=0}^{p-1}\Big[f(T^i y) - f(T^{i+m}x)\Big] = p\langle f\rangle(y) - \sum_{i=0}^{p-1} f(T^{i+m}x) \geq 0.$$

Next, consider $T^m x$ to be a point in a bad segment. We then have

$$\langle q\rangle(y) - q(T^m x) = \langle f - \sigma g + h\rangle(y) - (f - \sigma g + h)(T^m x)$$

$$= \langle f\rangle(y) - f(T^m x) + \sigma g(T^m x) + \langle h\rangle(y) - h(T^m x),$$

but h is bounded in A-norm and therefore in sup-norm by $\sigma\varepsilon$, so

$$|\langle h\rangle(y) - h(T^m x)| \leq 2\sigma\varepsilon.$$

This gives that

$$\langle q\rangle(y) - q(T^m x) \geq \langle f\rangle(y) - f(T^m x) + \sigma g(T^m x) - 2\sigma\varepsilon,$$

but being in a bad segment supposes that $\sigma g(T^m x) > \sigma A_{|\log(\gamma/(2^{p+1}))|}$, so for $\varepsilon < A_{|\log(\gamma/(2^{p+1}))|}/4$, we have

$$\langle q\rangle(y) - q(T^m x) > \langle f\rangle(y) - f(T^m x) + \frac{\sigma A_{|\log(\gamma/(2^{p+1}))|}}{2}.$$

Noting that bad segments occur with positive frequency, combining these two results we see

$$\langle q\rangle(y) - \langle q\rangle(x) > \langle f\rangle(y) - \langle f\rangle(x),$$

and so $\langle f \rangle (y)$ being optimum implies that the measure supported on $\mathcal{O}y$ is the optimizing measure for all functions in the open set Q. □

Using the more general framework that we have set up for functions of summable variation, we now present a class of functions for which the set of those uniquely optimized by measures supported on the orbit of periodic points contains an open dense subset. Proposition 3.34 already takes care of the periodic case, so we only need to deal with perturbing functions optimized by aperiodic orbits.

Proposition 3.35. *Let X be the Banach space generated by an A-norm $\|\cdot\|_A$ where A satisfies $A_n/A_{n+1} \to \infty$. If $f \in X$ is optimized by an aperiodic point, there exists an arbitrarily small perturbation $\tilde{f}$ of f such that $\tilde{f}$ is optimized by a periodic point.*

Before we prove Proposition 3.35, we need to make some subtle observations about following in order and our metric that will allow us to squeeze the necessary factors from our equations.

Notice that our metric on doubly-infinite sequences disregards some information. If x and y satisfy $d(x, y) = 2^{-k-1}$ then we are guaranteed that $(x)_{-k}^{k} = (y)_{-k}^{k}$, however it could be the case that $(x)_{-k}^{k+2} = (y)_{-k}^{k+2}$ or $(x)_{-k-5}^{k} = (y)_{-k-5}^{k}$, etc.. For this reason, it behooves us to work with not just $d(x, y)$ but also the number of symbols of agreement between x and y.

Along these lines, we define the notion of a weakly-recursive segment. Let $\Lambda(x, y)$ be the largest number of symbols of agreement between x and y starting from position zero. That is,

$$\Lambda(x, y) = \sup\{i : (x)_0^{i-1} = (y)_0^{i-1}\}.$$

Definition 3.36 (Weakly-recursive Segment)**.** *Given a point x, we say $\{x, Tx, T^2x, \ldots, T^{p-1}x\} \subset \mathcal{O}x$ is a* weakly-recursive segment *if for all $0 \leq i < j < p$,*

$$\Lambda(x, T^p x) > \Lambda(T^i x, T^j x).$$

Simply put, a segment is weakly recursive if the number of symbols matching starting from the two endpoints is larger than the number of symbols matching starting from any two points between the two endpoints. Also note that if $\left\{ {b \atop a} \right\}$ is a recursive segment with $\delta_{\left\{ {b \atop a} \right\}} = 2^{-k}$, then $\left\{ {b-k \atop a-k} \right\}$ is a weakly-recursive segment. However, there are weakly-recursive segments that do not give recursive segments.

Weakly-recursive segments have analogous attributes to a recursive segment's γ and δ. If $\left\{ {m+p \atop m} \right\}$ is a weakly-recursive segment of x, we define

$$\hat{k}_\delta = \Lambda(T^m x, T^{m+p} x) \qquad \text{and} \qquad \hat{k}_\gamma = \max_{0 \le i < j < p} \Lambda(T^{m+i} x, T^{m+j} x).$$

Lastly, we can create a derived periodic point from any weakly-recursive segment $\left\{ {m+p \atop m} \right\}$ of x by concatenating the symbols $(T^m x)_0^{p-1}$.

Using careful analysis, we can now slightly strengthen Proposition 3.29.

Corollary 3.37 (of Proposition 3.29)**.** *Let f be a function of summable variation, x a measure-recurrent optimal orbit for f, and y a period p point derived from a weakly-recursive segment. If $\hat{k}_\delta = 2k$ or $\hat{k}_\delta = 2k+1$,*

$$\langle f \rangle (x) - 4\mathrm{V}_{k+1}(f)/p \le \langle f \rangle (y) \le \langle f \rangle (x).$$

Proof. Notice, if $\hat{k}_\delta = 2k+1$, then y 2^{-k-1}-shadows $\mathcal{O}x$; specifically, there exists some m so $d(T^{k+i}y, T^{m+i}x) \le 2^{-k-1}$ for $0 \le i < p$. Since $(y)_0^{2k} = (x)_m^{m+2k}$, $T^k y$ agrees with $T^{m+k}x$ for a central block of length at least $2k+1$ symbols, and in particular $d(T^k y, T^{m+k}x) \le 2^{-k-1}$. Similarly $(T^p y)_0^{2k} = (T^p x)_m^{m+2k}$, and so $d(T^{m+p+k}x, T^{p+k}y) \le 2^{-k-1}$. Since $d(T^{m+k+i}x, T^{k+i}y)$ is closer than 2^{-k-1} for all $0 < i < p$, we have that x 2^{-k-1}-shadows $\mathcal{O}y$ for one period. So, by application of Proposition 3.29, we have the result.

Consider the case where $\hat{k}_\delta = 2k$. We will now show that we stay 2^{-k-1}-close to $\mathcal{O}x$ and that y in fact 2^{-k-1}-shadows $\mathcal{O}x$ for p steps.

Without loss of generality, assume that $\Lambda(x, y) \ge p + 2k$. That is, the point of $\mathcal{O}y$ that x is close to is y itself.

Figure 3.8: Comparison of the symbols of x with those of y.

Let us now find lower bounds on how many symbols we need to go from various positions in $\mathcal{O}y$ to find a symbol that disagrees with x. Taking a symmetric neighborhood about $T^k y$, we see $(T^k y)_{-k}^k = (T^k x)_{-k}^k$ and so $d(T^k y, T^k x) \le 2^{-k-1}$.

Similarly, $(T^{k+p-1}y)_{-k}^{k} = (T^{k+p-1}x)_{-k}^{k}$ and so $d(T^{k+p-1}y, T^{k+p-1}x) \leq 2^{-k-1}$. Further, we observe that for any $k \leq m \leq k+p-1$,

$$(T^m y)_{-k}^{k} = (T^m x)_{-k}^{k},$$

and so y 2^{-k-1}-shadows $\mathcal{O}x$ for p steps. Thus, an application of Proposition 3.29 completes the proof. □

We will now get a handle on how close you can remain to a periodic point derived from a weakly-recursive segment and *not* follow it in order. It will turn out that we only need to consider two cases. Case 1: y is derived from a weakly-recursive segment and $\hat{k}_\delta = 2k$ for some k; Case 2: y is derived from a weakly-recursive segment with $\hat{k}_\delta = 2k+1$ and $\hat{k}_\gamma \leq 2k-1$. (Briefly, we only need to deal with these two cases, because if $\hat{k}_\delta = 2k+1$ and $\hat{k}_\gamma = 2k$, it reduces to Case 1.)

Consider Case 1: the worst possible scenario (in terms of being able to remain close to $\mathcal{O}y$ yet not follow it in order) occurs when there is a repeated subword of length $2k-1$ (long repeated subwords allow us to agree for many symbols but still jump between positions of $\mathcal{O}y$). In this situation, we know that if x agrees with y for $2k+1$ symbols, Tx agrees for $2k > 2k-1$ symbols and so we must follow y in order. In other words, if our distance from y is at most 2^{-k-1}, we follow y in order. Therefore, if we do not follow y in order at some point we are at least a distance of 2^{-k} from $\mathcal{O}y$.

Case 2: Again, since the longest repeated subword in y is length at most $2k-1$, we have that being 2^{-k-1}-close to $\mathcal{O}y$ forces us to follow $\mathcal{O}y$ in order. Therefore, if we are out of order at some point, we must become a distance of at least 2^{-k} away.

We are now ready to prove Proposition 3.35.

Proof (of Proposition 3.35). Let $A = \{A_n\}$ satisfy $A_n/A_{n+1} \to \infty$ and let X be the Banach space generated by $\|\cdot\|_A$. Suppose $f \in X$ is optimized by some measure recurrent, aperiodic point x. If $\omega(x)$ contains a periodic point, by Corollary 3.30 we are done, so assume $\omega(x)$ contains no periodic points.

Notice that because $A_n/A_{n+1} \to \infty$ and hence bounded away from one, A_n tends to zero at least as quickly as a geometric sequence. Thus, there exists some constant L' such that

$$\sum_{i=k}^{\infty} A_i < L' A_k \tag{3.5}$$

for all k. In the geometric case, $A_n/A_{n+1} = C$ is constant, so $\sum_{i=k}^{\infty} A_i \leq \left[\frac{C}{C-1}\right] A_k$. Thus, because for any constant C, eventually $A_n/A_{n+1} > C$, we can pick $L' \geq C/(C-1)$ such that equation (3.5) is also satisfied in the finite number of cases before $A_n/A_{n+1} > C$. Let $L = \max\{L', \|f\|_A\}$. We now have

$$\mathrm{var}_k f \leq LA_k \qquad \text{and} \qquad V_k f \leq L^2 A_k.$$

Fix $\varepsilon > 0$ and a weakly-recursive segment $\left\{{m+p \atop m}\right\}$ of $\mathcal{O}x$ with $\hat{k}_\delta$ large (to be determined later) and let y be the derived periodic point. Let k_δ be such that $\hat{k}_\delta = 2k$ or $\hat{k}_\delta = 2k+1$ and let $\hat{k}_\gamma$ be the length of the longest repeated subword in y. Assume further that if $\hat{k}_\delta$ is odd, then $\hat{k}_\gamma \leq \hat{k}_\delta - 2$. This assumption will be justified later. Define the perturbation function g by

$$g(t) = A_n \quad \text{if } d(t, \mathcal{O}y) = 2^{-n},$$

and let

$$\tilde{f} = f - \varepsilon g.$$

We will now show that provided $\hat{k}_\delta$ is sufficiently large and satisfies a justifiable assumption, $\langle f \rangle (y) \geq \langle f \rangle (z)$ for all measure recurrent points z. Let z be a measure recurrent point with $\mathcal{O}z \neq \mathcal{O}y$. Very similarly to the proof of Proposition 3.20, partition $\mathcal{O}z$ into segments W_j and S_j with the W_j's satisfying the following: Each W_j contains a point that does not follow $\mathcal{O}y$ in order (i.e., there exists an i with $T^i x$ "in W_j" such that the closest point in $\mathcal{O}y$ to $T^i x$ is y' but the closest point in $\mathcal{O}y$ to $T^{i+1}x$ is not Ty'); the size of W_j is a multiple of p; for each j there exists a $y_j \in \mathcal{O}y$ such that the right endpoint of W_j and the left endpoint of W_{j+1} both come within 2^{-k-2} of y_j; the collection of W_j's contains all points of $\mathcal{O}z$ that do not follow $\mathcal{O}y$ in order for at least one step; and, each W_j is of minimal length.

Let z' be the concatenation of the W_j's and note that $T^i z'$ is always within 2^{-k-1} of $\mathcal{O}z$ and z' 2^{-k-1}-shadows segments of $\mathcal{O}z$. Label the segment between W_{j-1} and W_j by S_j. It should be noted that it is possible that z never becomes close to $\mathcal{O}y$ and so our partition results in a single W_0 element. In this case, sums over S_j's are taken to be zero.

Recall that since $\hat{k}_\delta = 2k$ or $2k+1$, if $T^i z$ does not follow $\mathcal{O}y$ in order for one step, $d(T^i z, \mathcal{O}y) \geq 2^{-k}$.

Let ℓ be the maximum amount of time a point can be 2^{-k-1}-close to $\mathcal{O}y$ but remain at least $2^{-k-1}/(2^{p+2})$ from $\mathcal{O}y$. Because of our assumption that if $\hat{k}_\delta$ is odd, $\hat{k}_\gamma \leq \hat{k}_\delta - 2$, being 2^{-k-1}-close to $\mathcal{O}y$ implies we follow $\mathcal{O}y$ in order. Thus, using similar analysis to that in the proof of Proposition 3.20 we have

$$\ell < 2\log\left(\frac{2^{-k-1}}{2^{-k-1}}(2^{p+3})\right) \leq 2(p+3).$$

Since the W_j's, from which z' is constructed, were chosen to have minimal length (as to spend the most time away from $\mathcal{O}y$ as possible), we know $T^i z'$ must satisfy $d(T^i z', \mathcal{O}y) \geq 2^{-k}$ at least $\frac{1}{2\ell+1} > \frac{1}{4\ell}$ portion of the time and so

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} g(T^i z') > \frac{A_k}{4\ell}.$$

We must take a lim inf in this case since $\langle g\rangle(z')$ may not exist because z' may not be measure recurrent, but since this notation is convenient, we will extend the notation by defining

$$\langle F\rangle_{\inf}(z') = \liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} F(T^i z')$$

and

$$\langle F\rangle(z') = \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} F(T^i z')$$

for all functions F. Recall Theorem 2.10, which says that $\langle f\rangle(x') \leq \langle f\rangle(x)$ still holds even though $\langle f\rangle(x')$ is now defined in terms of a lim sup. Corollary 3.37 gives us $\langle f\rangle(y) \geq \langle f\rangle(x) - 4\mathrm{V}_{k+1}f/p$, and so

$$\begin{aligned} &\left\langle \tilde{f}\right\rangle(z') - \left\langle \tilde{f}\right\rangle(y) = \\ \langle f\rangle(z') - \varepsilon\langle g\rangle_{\inf}(z') - \langle f\rangle(y) &\leq \langle f\rangle(z') - \varepsilon\langle g\rangle_{\inf}(z') + \frac{4V_{k+1}f}{p} - \langle f\rangle(x) \\ &\leq \langle f\rangle(z') - \langle f\rangle(x) - \frac{\varepsilon A_k}{4\ell} + \frac{4V_{k+1}f}{p} \\ &\leq \frac{4V_{k+1}f}{p} - \frac{\varepsilon A_k}{4\ell}. \end{aligned} \tag{3.6}$$

Since z' follows $\mathcal{O}z$ in order along W_j, for each j, the Parallel Orbit Lemma B gives us

$$-2V_{k+1}f \leq \sum_{i\in W_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} z')\right] \leq 2V_{k+1}f,$$

and since z 2^{-k-1}-shadows $\mathcal{O}y$ in order along S_j, the In Order Lemma gives us

$$-2V_{k+1}f \leq \sum_{i\in S_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} y)\right] \leq 2V_{k+1}f.$$

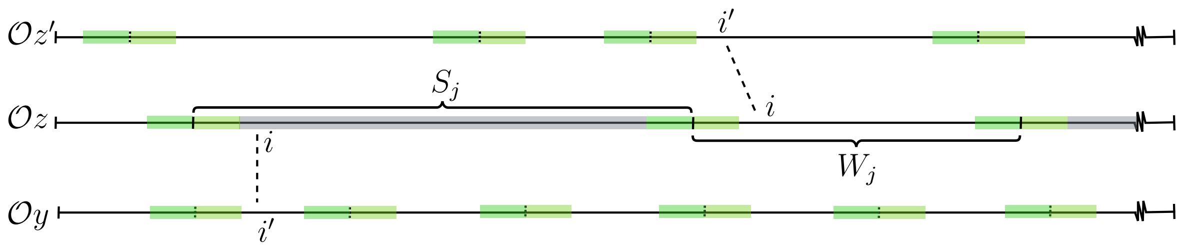

Figure 3.9: Picture showing the correspondence between i and i' for $i \in S_j$ and $i \in W_j$.

Recall that we are using simplifying notation: the i' corresponding to $i \in S_j$ is the unique i' such that $d(T^i z, T^{i'} y) \leq 2^{-i-1}$ and i, i' when $i \in W_j$ means $i \in W_j$ and the i' such that $T^{i'} z'$ corresponds to the point in $\mathcal{O}z'$ created from $T^i z$.

Rewriting $\left\langle \tilde{f} \right\rangle (z')$ as a limit, from equation (3.6), we see

$$\limsup_{j\to\infty} \left[\frac{\displaystyle\sum_{i'\in W_{-j},\ldots,W_j} \tilde{f}(T^{i'} z')}{|W_{-j}| + \cdots + |W_j|} \right] - \left\langle \tilde{f} \right\rangle (y) \leq \frac{4V_{k+1}f}{p} - \frac{\varepsilon A_k}{4\ell}. \tag{3.7}$$

Notice that $\left|\sum_{i\in W_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} z')\right]\right| \leq 2V_{k+1}f$ gives

$$\left| \sum_{i\in W_{-j},\ldots,W_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} z')\right] \right| \leq (2j+1)2V_{k+1}f,$$

and because W_j were chosen such that $|W_j| \geq p$, we get

$$\limsup_{j\to\infty} \frac{1}{|W_{-j}| + \cdots + |W_j|} \left| \sum_{i\in W_{-j},\ldots,W_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} z')\right] \right| \leq \frac{2V_{k+1}f}{p}. \tag{3.8}$$

Thus, by combining equations (3.7) and (3.8), we may conclude

$$\limsup_{j\to\infty} \left[\frac{\sum\limits_{i\in W_{-j},\ldots,W_j} \tilde{f}(T^i z)}{|W_{-j}| + \cdots + |W_j|} \right] - \left\langle \tilde{f} \right\rangle (y) \leq \frac{6V_{k+1}f}{p} - \frac{\varepsilon A_k}{4\ell}. \tag{3.9}$$

Analyzing the average along multiple S_j segments, we see that since

$$\sum_{i\in S_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} y)\right] \leq 2V_{k+1}f,$$

$$\sum_{i\in S_{-j},\ldots,S_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} y)\right] \leq b2V_{k+1}f,$$

where $b = \left|\{|S_r| > 0 : -j \leq r \leq j\}\right| \leq 2j+1$ (if $|S_r| = 0$, we clearly get no contribution from the $\sum_{i\in S_r}$ term). Since $|S_r|$ is a multiple of p, we have $|S_r| < p$ implies $|S_r| = 0$. Thus, there are at least b S_r's with $|r| \leq j$ satisfying $|S_r| \geq p$. This gives us

$$\frac{1}{|S_{-j}| + \cdots + |S_j|} \sum_{i\in S_{-j},\ldots,S_j} \left[\tilde{f}(T^i z) - \tilde{f}(T^{i'} y)\right] \leq \frac{2V_{k+1}f}{p},$$

and so pulling out $(|S_{-j}| + \cdots + |S_j|) \left\langle \tilde{f} \right\rangle (y)$ from the sum gives us

$$\frac{1}{|S_{-j}| + \cdots + |S_j|} \left[\sum_{i\in S_{-j},\ldots,S_j} \tilde{f}(T^i z) \right] - \left\langle \tilde{f} \right\rangle (y) \leq \frac{2V_{k+1}f}{p}. \tag{3.10}$$

Reintroducing the S_j segments and combining equations (3.9) and (3.10), we see

$$\left\langle \tilde{f} \right\rangle (z) - \left\langle \tilde{f} \right\rangle (y) = \lim_{j\to\infty} \left[\frac{\sum\limits_{i\in W_{-j}S_{-j}\cdots S_j W_j} \tilde{f}(T^i z)}{|W_{-j}S_{-j}\cdots S_j W_j|} \right] - \left\langle \tilde{f} \right\rangle (y) \leq \frac{8V_{k+1}f}{p} - \frac{\varepsilon A_k}{4\ell},$$

and so

$$\frac{8V_{k+1}f}{p} - \frac{\varepsilon A_k}{4\ell} < 0$$

would prove the non-optimality of z. Notice that this inequality is satisfied if

$$8V_{k+1}f - \frac{p\varepsilon A_k}{4\ell} < 0$$

is satisfied. Reintroducing our bound for ℓ, we see that

$$\frac{\varepsilon p}{4\ell} \geq \frac{\varepsilon p}{8(p+3)} \geq \frac{\varepsilon p}{8(4p)} = \frac{\varepsilon}{2^5}.$$

Recalling that $V_{k+1}f \leq L^2 A_{k+1}$, we now see it is sufficient if

$$A_{k+1} - A_k \left[\frac{\varepsilon}{2^8 L^2}\right] < 0 \tag{3.11}$$

is satisfied.

Since $A_n/A_{n+1} \to \infty$ we have that for any constant $c > 0$, all large enough n satisfy

$$A_{n+1} < cA_n.$$

Thus, we have shown that if there exist weakly-recursive segments such that either $\hat{k}_\delta$ is arbitrarily large and even or there exist recursive segments such that $\hat{k}_\delta$ is arbitrarily large and odd with $\hat{k}_\gamma \leq \hat{k}_\delta - 2$, then equation (3.11) can be satisfied.

We complete the proof with the observation that our demand on weakly-recursive segments where if $\hat{k}_\delta$ is odd then $\hat{k}_\gamma \leq \hat{k}_\delta - 2$ can always be met. Pick a weakly-recursive segment S with $\hat{k}_\delta$ arbitrarily large. If $\hat{k}_\delta$ is even, we are done. If $\hat{k}_\delta$ is odd and $\hat{k}_\gamma \leq \hat{k}_\delta - 2$ we are done. In the remaining case, we have that $\hat{k}_\delta$ is odd and $\hat{k}_\gamma = \hat{k}_\delta - 1$. Thus, $\hat{k}_\gamma$ is even and so we may find a new weakly-recursive segment $S' \subset S$ such that $\hat{k}'_\delta = \hat{k}_\gamma$ is even (where $\hat{k}'_\delta$ is the $\hat{k}_\delta$ corresponding to the weakly-recursive segment S'). □

We call functions f where $\|f\|_A < \infty$ and A satisfies $A_n/A_{n+1} \to \infty$ *super-continuous*. It is clear that these functions are a sub-class of Lipschitz continuous functions which are already a sub-class of continuous functions. Super-continuous functions are in some sense Lipschitz functions whose Lipschitz constant gets finer and finer as one looks at smaller and smaller scales. It is also worth noting that if the definition of super-continuity is properly extended to the reals with Euclidean distance, the only super-continuous functions are constant functions (since $d(x, y) = d(x + z, y + z)$ when d is the Euclidean distance, any super-continuous function must locally flatten out at every point and so must be constant). However, in the case of the Shift, the distance function is actually an ultrametric (i.e. a metric satisfying the stronger triangle inequality $d(x, y) \leq \max\{d(x, z), d(z, y)\}$) and so many non-constant examples of super-continuous functions exist.

The proof of Proposition 3.35 essentially enforces the important consequence of class I points onto all points through the use of super-continuity. The importance of class I points is that we can find recursive segments such that going away from that recursive segment (i.e., not following it in order) means you must get very far away and so you incur a large penalty from the perturbation function. Super-continuity turns that around slightly and allows for the creation of a penalty function from a recursive segment such that even following the weakly-recursive segment out of order for one step incurs a large penalty.

It is now a direct application of Proposition 3.34 to obtain the following.

Proposition 3.38. *Let X be the Banach space generated by an A-norm $\|\cdot\|_A$ where A satisfies $A_n/A_{n+1} \to \infty$. If $f \in X$ is optimized by an aperiodic point, there exists an arbitrarily small perturbation $\tilde{f}$ of f such that there is an open set of functions containing $\tilde{f}$ all optimized by the same periodic point.*

We may now formally state that functions optimized by periodic orbits are dense in super-continuous function.

Theorem 3.39. *Let X be the Banach space generated by an A-norm $\|\cdot\|_A$ where A satisfies $A_n/A_{n+1} \to \infty$, and let $P \subset X$ be the set of functions uniquely optimized by measures supported on the orbit of a periodic point. P then contains an open set that is dense in X.*

Proof. Let f be a super-continuous function optimized by a periodic point y. Proposition 3.34 gives us that there exists an open set O_f of functions uniquely optimized by the periodic orbit measure generated y with $f \in \overline{O_f}$. Proposition 3.38 gives that functions optimized by periodic points are dense in all functions, and so letting

$$P = \bigcup_f O_f,$$

where f ranges over all functions optimized by periodic points, completes the proof. $\square$

Chapter 4

Conclusion

In this thesis, we have presented the concept of ergodic optimization. That is, for a function f we search for an invariant measure μ such that $\int f\,\mathrm{d}\mu \geq \int f\,\mathrm{d}\nu$ for all invariant measures ν. The main conjecture in this field is that functions tend to be optimized by simple measures, with the simplest measures being those supported on the orbit of a periodic point, a.k.a. periodic orbit measures.

It has been shown by Bousch and others[2, 4] that in various spaces, this conjecture holds true. That is, the set of functions optimized by periodic orbit measures is open and dense. However, in the general space of Hölder/Lipschitz functions, this conjecture is still open.

Yuan and Hunt made a great deal of progress towards showing this conjecture true in the case of Lipschitz functions, and this thesis spends a great deal of time presenting proofs of many of the theorems and statements given without proof by Yuan and Hunt in [13]. These streamlined proofs include graphics and attempt to be logically simple, resorting to long equations and arguments by contradiction as little as possible. We also attempt to outline the basic strategy of each proof so that a general understanding of the methods used is possible without a meticulous read.

Yuan and Hunt's strategy for showing the set of functions uniquely optimized by periodic orbit measures contains an open and dense set was to divide up aperiodic points into two categories: class I and class II. Class I points are those aperiodic points that contain arbitrarily good approximations of periodic points, and so given a function optimized by a class I point x, there is a natural way of picking periodic points to be the support of an optimizing measure. Penalizing against going away from the chosen periodic point allows us to force the optimizing measure to be supported on that periodic point. Class II says that the point does not approximate periodic points very well in the sense that the closer you come to approximating a periodic point, the larger your period must be (in fact, your period must be super-linear in your return distance). The lack of good periodic approximations means that there is no apparent natural way to choose a periodic point to become the support of an optimizing measure. However, the assumption that "all small perturbations of a class-II-optimized point are still optimized by the same point" produces a contradiction, and so while not showing that a class-II-optimized function can be perturbed to be optimized by a periodic point, Yuan and Hunt showed that class-II-optimized functions are unstable.

For this reason, Yuan and Hunt fell short of the goal of finding an open, dense set of functions uniquely optimized by periodic orbit measure. However, it should be noted that class II points are not well understood. Though they have been constructed, it is an open question as to whether measure recurrent points that are not in class I exist. If they do not, then Yuan and Hunt's methods do produce an open, dense set.

Based on the idea of making everything appear like a class I segment, we introduced the A-norm and used it to define super-continuous functions. These functions may be informally described as Lipschitz functions whose Lipschitz constant decreases to zero when comparing points that are closer and closer together. In the realm of super-continuous functions, it is possible to create a penalty function that puts a large negative bias against deviating from a periodic point even by the smallest amount. This allowed us to prove that in the super-continuous functions, the set of functions uniquely optimized by periodic orbit measures contains an open, dense set with respect to the A-norm.

It has now been shown that in certain spaces larger than the space of Lipschitz functions, namely the space of Walters functions, and in other spaces smaller than the space of Lipschitz functions, namely the space of super-continuous functions, that the set of functions optimized by periodic orbit measures contains an open, dense set. Yet, the in between space, the space of Lipschitz functions, somehow remains elusive in this regard. Though there would be no apparent contradiction with Lipschitz functions not satisfying this property (seeing as the surrounding cases each use norms different from the Lipschitz norm), the evidence certainly suggests that this outcome is unlikely. Even so, math has seen stranger things and this author is eager to see if he or others eventually resolves this issue.

Bibliography

[1] Thierry Bousch. Le poisson n'a pas d'arêtes. *Ann. Inst. Henri Poincaré, Probabilités et Statistiques*, 36:197–224, 2000.

[2] Thierry Bousch. La condition de walters. *Ann. scientifiques de l'Ecole Normale Supérieure*, 34:287–311, 2001.

[3] Richard Durrett. *Probability: Theory and Examples, Third Edition*. Duxbury Press, Boston, MA, 2004.

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

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

[6] Oliver Jenkinson. Optimiziation and majorization of invariant measure. *Electronic Research Announcements of the American Mathematical Society*, 13:1–12, 2007.

[7] I. D. Morris. A sufficient condition for the subordination principle in ergodic optimization. *Bull. London Math. Soc.*, 39:214–220, 2007.

[8] Anthony Quas. Ergodicity and mixing properties. *Encyclopedia of Complex Systems*, 2009.

[9] Walter Rudin. *Real and Complex Analysis: Third Edition*. McGraw-Hill Book Company, New York, 1986.

[10] Peter Walters. Ruelle's operator theorem and g-measures. *Transactions of the American Mathematical Society*, 214:375–387, 1975.

[11] Peter Walters. *Ergodic Theory*. Springer-Verlag, New York, 1982.

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

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

Index