

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/58569>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

Table of Contents

Chapter 1. Invariant Measures for Families of Circle Maps	1
1.1 Introduction	3
1.2 Two examples	8
Theorem 1.1	
1.3 Proof of Theorem 1.1	13
Chapter 2. Expanding Maps, g -Measures and Generalized Baker's Transformations	21
2.1 Introduction	24
2.2 Connections between g -Measures, Expanding Maps and Generalized Baker's Transformations	33
2.3 Construction of Examples of Expanding Maps	37
2.4 Some Possible Approaches to Question 1	42
Chapter 3. Representation of g -Measures by Smooth Repellers	53
3.1 Introduction	53
3.2 Previously Known Results	60
3.3 Background for Smooth Repellers	64
3.4 Physical Motivation for Theorem 3.3	66
3.5 Differential Equations Background for Theorem 3.3	68
3.6 Proof of Theorem 3.3	70
3.7 Invariant Measures	72
Chapter 4. Representation of Markov Chains on Tori	77
4.1 Representation of Markov Chains on Tori	78
4.2 Degree of a Markov Chain on the Circle	82
4.3 A Markov Chain which cannot be represented by Homeomorphisms	85

by Anthony Nicholas Quas

Mathematics Institute,
University of Warwick,
Coventry,
CV4 7AL.

This thesis submitted in partial fulfilment of
the requirements of the degree of Doctor of
Philosophy at the University of Warwick

August 1993



Table of Contents

Chapter 1. Invariant Measures for Families of Circle Maps	3
1.1 Introduction	3
1.2 Two examples showing necessity of some conditions for Theorem 1.1	8
1.3 Proof of Theorem 1.1	13
Chapter 2. Expanding Maps, g -Measures and Generalized Baker's Transformations	21
2.1 Introduction	21
2.2 Connections between g -Measures, Expanding Maps and Generalized Baker's Transformations	35
2.3 Construction of Examples of Expanding Maps	42
2.4 Some Possible Approaches to Question 1	46
Chapter 3. Representation of Markov Chains on Manifolds	58
3.1 Introduction	58
3.2 Previously Known Results	60
3.3 Background for Smooth Representation	64
3.4 Physical Motivation for Theorem 3.3	65
3.5 Differential Equations Background for Theorem 3.3	68
3.6 Proof of Theorem 3.3	70
3.7 Invariant Measures for Smooth Markov Chains	72
Chapter 4. Representation of Markov Chains on Tori	77
4.1 Representation of Markov Chains on Tori	78
4.2 Degree of a Markov Chain on the Circle	82
4.3 A Markov Chain which cannot be Represented by Home- omorphisms	92
Bibliography	95

Table of Figures

Figure 1-1	Possible arrangement of intervals about a periodic point	6
Figure 1-2	Lift of an iterate of a circle map showing the types of periodic points	7
Figure 1-3	A family for which the conclusion of the theorem fails	9
Figure 1-4	'Speed Function' for the counterexample	10
Figure 1-5	Typical diagram of periodic points against parameter	16
Figure 1-6	Possible arrangement of chosen periodic points	19
Figure 2-1	Possible graph of ϕ	34
Figure 2-2	Fundamental partition of S under T_f	34
Figure 2-3	Graph of T	44
Figure 2-4	Configuration of J intervals	47
Figure 2-5	Construction of the map T	47
Figure 3-1	Possible paths in the domain and image	66

Acknowledgements

I would like to thank my supervisor, Peter Walters, for the encouragement and help which he offered during the work for this thesis. Sebastian Van Strien was of assistance in suggesting some of the questions which are addressed in Chapter 1. Thanks to Saul Jacka and Chris Bose for interesting conversations about some of the questions raised in Chapter 2, and to Chris Bose for bringing to my attention his work on generalized baker's transformations. I would also to acknowledge useful conversations and advice from Mario Micallef, Janice Booth and Jim Eells during the preparation of Chapter 3.

Declaration

I declare that the work contained in this thesis is original, except where otherwise indicated. In particular, Chapter 2 contains a survey of the known results about the concepts relevant to that chapter. Similarly, Chapter 3 borrows some known results from Differential Equations theory. These parts are clearly marked as being already known.

Some Problems in Ergodic Theory

Summary

The thesis consists of a study of problems in ergodic theory relating to one-dimensional dynamical systems, Markov chains and generalizations of Markov chains. It is divided into chapters, three of which have appeared in the literature as papers. Chapter 1 looks at continuous families of circle maps and investigates conditions under which there is a weak*-continuous family of invariant measures. Sufficient conditions are exhibited and the necessity of these conditions is investigated. Chapter 2 is about expanding maps of the interval and the circle, and their relation with g -measures and generalized baker's transformations. The g -measures are generalizations of Markov chains to stochastic processes with infinite memory and generalized baker's transformations are geometric realizations of these. The chapter is based around the question of whether there exist expanding maps preserving Lebesgue measure, for which Lebesgue measure is not ergodic. Results are known if the map is sufficiently differentiable (for example $C^{1+\alpha}$), but the C^1 case is still unclear. The chapter contains some partial solutions to this question. Chapter 3 is about representation of Markov chains on compact manifolds by measured collections of smooth maps. Given a measured collection of maps, a Markov chain is induced in a natural fashion. This chapter is about reversing this process. Chapter 4 describes a specialization of the setting of Chapter 3 to Markov chains on tori. In this case, it is possible to demand more of the maps of the representation than smoothness. In particular, they can be chosen to be local diffeomorphisms. The chapter also addresses the question of whether in general the maps can be taken to be diffeomorphisms and gives a counterexample showing that there exist Markov chains on tori which do not admit a representation by diffeomorphisms.

The invariant measures, which are central in the study of ergodic theory, have been shown to be given to ensure that the limit function f is constant almost everywhere with respect to the invariant measure for any L^1 function f . This turns out to be extremely important.

Chapter 1 of this dissertation considers the case where there is a continuously generalized family of circle maps (that is orientation-preserving homeomorphisms of the circle). Each circle map is known to have at least one invariant measure. In this chapter, I consider whether the invariant measure for the family of circle maps may be chosen to vary continuously with respect to the parameter. In general, I show that subject to certain conditions, this may be arranged by careful choice of the invariant measure. In an exceptional situation, the invariant measure is determined by the initial conditions. Typically, without special initial conditions, one would not expect to see continuous variation of the invariant measure as the parameter is varied continuously. The results of this chapter

Some Problems in Ergodic Theory

Anthony Quas

This dissertation concerns itself with problems of ergodic theory, the branch of dynamical systems theory which deals with problems of long-term averages of values of a measurement taken at discrete intervals of time. In general, the formulation is that T is a map from some measure space X to itself and f is an L^1 function $X \rightarrow \mathbb{R}$. The main objects of study in ergodic theory are the averages

$$A_n[f](x) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x).$$

Ergodic theory gives conditions for these to converge pointwise almost everywhere with respect to an appropriate measure (or in L^1) to a function \bar{f} . These measures are in fact the invariant measures, which are central in the study of ergodic theory. Further conditions can be given to ensure that the limit function \bar{f} is constant almost everywhere with respect to the invariant measure for any L^1 function f . This turns out to be extremely important.

Chapter 1 of this dissertation considers the case where there is a continuously parameterized family of circle maps (that is orientation-preserving homeomorphisms of the circle). Each circle map is known to have at least one invariant measure. In this chapter, I consider whether the invariant measures for the family of circle maps may be chosen to vary continuously with respect to the parameter. In general, I show that subject to certain conditions, this may be arranged by careful choice of the invariant measure. In an experimental situation, the invariant measure is determined by the initial conditions. Typically, without special initial conditions, one would not expect to see continuous variations of the long-term averages as the parameter is varied continuously. The results of this chapter

thus show that while typically long-term averages change discontinuously with respect to small changes in the parameter, they may in special circumstances vary continuously.

Chapter 2 deals with the relationships between the concepts of expanding maps, g -measures and generalized baker's transformations. The starting point is the question of uniqueness of absolutely continuous invariant measures for C^1 expanding maps of the interval and the circle. This leads naturally into an investigation of g -measures, which may be considered as a generalization of Markov chains to processes which depend on the entire past, not just the last outcome. These are known to have a geometric realization as generalized Baker's transformations. This realization is studied in Chapter 2 and made more explicit. Finally, I present some examples of possible constructions of C^1 maps which might have more than one absolutely continuous invariant measure. If correct, these would provide a solution to a question of Keane ([Kea]).

Chapter 3 looks at Markov chains on compact manifolds. Conditions are found for Markov chains, under which there exists a family of smooth maps from the manifold to itself and a probability measure on them such that applying the maps at random according to the probabilities specified by the measure reproduces the Markov chain. This is called a *representation* of a Markov chain. A representation of a Markov chain allows it to be viewed as a Random Dynamical System (RDS), as described in [Ki] and [AC].

Chapter 4 is a specialization of Chapter 3 to the case where the manifold is a torus. In this case, it is shown that a smooth Markov chain admits a representation by local diffeomorphisms. It is then natural to ask whether such a Markov chain in fact admits a representation by diffeomorphisms. It is shown that in general this is not the case.

Chapter 1. Invariant Measures for Families of Circle Maps

1. Introduction

This chapter considers the invariant measures of a continuous family of circle maps. There is some evidence (see below) that a continuous family of circle maps should have a continuous family of invariant measures. In fact, this does not always turn out to be the case, but in this chapter, we give conditions for the conclusion to hold and show the necessity of some of these. The results of this chapter have appeared in the literature as [Q2].

Let π denote the projection $\mathbb{R} \rightarrow S^1$ given by $x \mapsto \exp(2\pi ix)$. We will denote in the usual way intervals on the circle (for example, the interval $[a, b]$ is the closed interval starting at a and going anticlockwise round to b). By a *circle map*, we will always mean an orientation-preserving homeomorphism $T : S^1 \rightarrow S^1$. For a detailed introduction to the theory of circle maps, the reader is referred to [CFS], §3.3. The main results, however, are summarized below for convenience. The dynamical behaviour of circle maps is very well understood, and may be principally characterized by the *rotation number* of the map. This is a measure of the ‘average rotation’ that the map imparts to a point. To define the rotation number of a circle map $T : S^1 \rightarrow S^1$, we first need its lift $F : \mathbb{R} \rightarrow \mathbb{R}$. The *lift* of a continuous map ϕ of the circle (not necessarily a circle map) is a continuous map $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ defined by the equation $\pi \circ \Phi = \phi \circ \pi$. This is uniquely defined up to an additive integer constant. The *degree* of the map ϕ is given by $\Phi(x+1) - \Phi(x)$. This is always an

integer and is independent of the point $x \in \mathbb{R}$ and the lift chosen. In the case of a circle map, the degree is always 1. The rotation number of the circle map T is then given by

$$\rho(T) = \lim_{n \rightarrow \infty} \frac{F^n(x) - x}{n},$$

where F is a lift of T . This limit exists for all x , and is independent of x . The convergence of the limit is uniform in x and the rotation number is unique for a map T up to an additive integer constant (depending on the particular lift chosen to represent T). The notion of a circle map with rational rotation number is therefore well-defined, since this property does not depend on the lift chosen. It can be shown that a circle map has rational rotation number with denominator q say (with the fraction expressed in its lowest terms), if and only if it has periodic points of period q . Further, if this is the case, then each point of the circle converges monotonically to a periodic point under iteration of the map. From these facts, it follows that a circle map with irrational rotation number has no periodic orbits. Here, the dynamics are also well-understood: each point has the same ω -limit point set and this is either a Cantor set, or the whole circle. In the former case, the map is semi-conjugate to a rotation through 2π times the rotation number, and in the latter case, the map is conjugate to a rotation through 2π times the rotation number.

It may be shown by elementary means that the map taking a circle map to its rotation number is continuous with respect to the C^0 -topology on the space of circle maps; see for example [CFS], §3.3, theorem 2.

Suppose now that F is a lift of a circle map T . Write $R(x) = F(x) - x$ and $r : S^1 \rightarrow \mathbb{R}$; $y \mapsto R(\pi^{-1}(y))$. This is well-defined since R is periodic, and is the amount of rotation which the point y undergoes when it is acted upon by T . Now, we have

$$\frac{1}{n} \sum_{i=0}^{n-1} r(T^i(\pi(x))) = \frac{1}{n} (F^n(x) - x).$$

From this it follows that $\frac{1}{n} \sum_{i=0}^{n-1} r(T^i(y))$ converges uniformly to $\rho(T)$ as $n \rightarrow \infty$. But, for any invariant measure ν for T , $\int r(y) d\nu(y) = \int \frac{1}{n} \sum_{i=0}^{n-1} r(T^i(y)) d\nu(y)$, so taking limits, we get

$$\int r(y) d\nu(y) = \rho(T).$$

This shows that the rotation number of a circle map is numerically equal to the amount of rotation at each point integrated with respect to an invariant measure for the circle map. As T changes continuously, $r(y)$ and $\rho(T)$ both change continuously. This suggests the invariant measures also depend continuously on the circle map in some sense. The appropriate sense of continuity turns out to be weak*-continuity, and this chapter contains an investigation of the weak*-continuity of the invariant measures of circle maps.

In the statement of the theorems, we will need some definitions. We say that a family $(T_\alpha)_{\alpha \in J}$ of circle maps, with J a compact subinterval of \mathbb{R} is a *continuous family of circle maps* if the map $T : J \times S^1 \rightarrow S^1; (\alpha, \xi) \mapsto T_\alpha(\xi)$ is continuous.

Given a circle map T with rotation number p/q , let S be the lift of T^q fixing the preimages of the periodic points. Define $u(x) = S(x) - x$. Note that u satisfies the equation $u(x) = u(x + 1)$, since the degree of T^q is 1. The function $v : S^1 \rightarrow \mathbb{R}; \xi \mapsto u(\pi^{-1}(\xi))$ is then well-defined. Note that the zeros of v are precisely the periodic points of the map T . Then given a periodic point ξ , there may be a neighbourhood of ξ on which v takes the value 0 only at ξ itself. If such a neighbourhood exists, we say the periodic point is of *definite type*, and conversely, if no such neighbourhood exists, we say the periodic point is of *indefinite type*. If the periodic point is of definite type, it follows that there is an open interval I_1 clockwise from ξ with ξ as an endpoint on which the sign of v is constant, and a similar interval I_2 anticlockwise from ξ (See Figure 1.1).

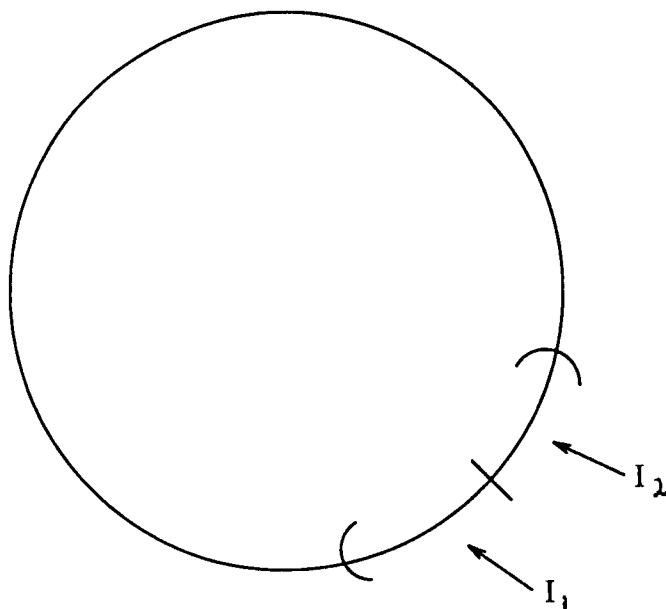


Figure 1.1 Possible arrangement of intervals about a periodic point.

We say that ξ is of type $++$, $+ -$, $- +$ or $--$ according to the sign of v on these two intervals. A *hyperbolic periodic point* is one of type $+ -$ or $- +$ (these are *stable* and *unstable* respectively). The types $++$ and $--$ of periodic point are *non-hyperbolic* and have stability on one side only. We call a map with non-hyperbolic periodic points (or sometimes its parameter value) *critical*. Note that if a point on a periodic orbit is of a particular type, then all the other points on the orbit are of that type (this follows since the maps are orientation-preserving homeomorphisms), so that it makes sense to say that a periodic orbit is of a specific type, or in particular hyperbolic or non-hyperbolic (see Figure 1.2).

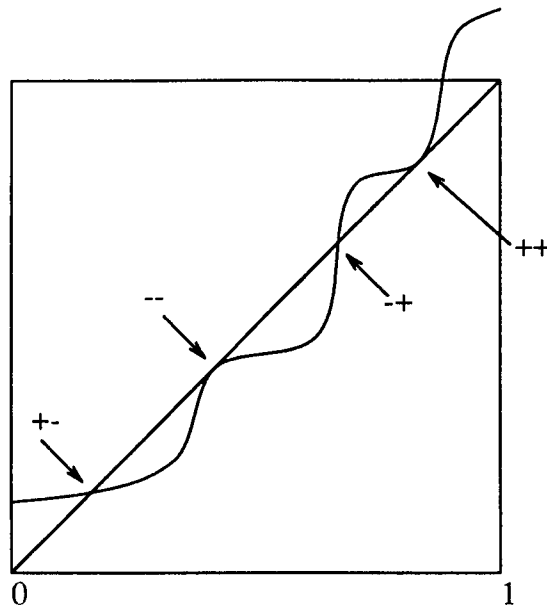


Figure 1.2 Lift of an iterate of a circle map showing the types of periodic points.

An *invariant Borel probability measure* (or briefly *invariant measure*) for a circle map T is a probability measure μ on the Borel σ -algebra Ω satisfying $\mu(T^{-1}(B)) = \mu(B)$ for any Borel set B . A circle map T is *uniquely ergodic* if it has exactly one invariant measure. There is a well-known theorem of ergodic theory (see [Wa3], theorem 6.18) saying that if T is a circle map with irrational rotation number, then T is uniquely ergodic. A map with rational rotation number is uniquely ergodic if and only if it has a unique periodic orbit (see Lemma 2).

We are now ready to state the theorem:

Theorem 1. Suppose that $(T_\alpha)_{\alpha \in J}$ is a continuous family of circle maps such that

- (i) for each non-trivial interval K on which the rotation number has a constant value, there are at most finitely many values of α in K for which T_α is critical, and
- (ii) for each $\alpha \in J$ such that T_α has rational rotation number, T_α has a finite number of periodic orbits. If there is more than one such orbit, then at least one of them is

hyperbolic.

Then there is a weak-continuously varying family of probability measures μ_α such that μ_α is an invariant measure for T_α .*

Part of Theorem 1 was previously known to Herman. In particular, Herman showed that the map taking a circle map with irrational rotation number to its unique invariant probability measure is weak*-continuous on the sets F_ρ , the collection of circle maps with rotation number equal to ρ (irrational). He in fact shows (see [He], proposition X.6.1), that the (semi-)conjugacy h conjugating a circle map $f \in F_\rho$ to the rotation by $2\pi\rho$ depends continuously on f . Since the invariant measure is given by $\mu(A) = \lambda(h(A))$, this implies that the map taking f to its invariant measure μ is weak*-continuous when restricted to F_ρ . This result can easily be recovered from the proof here.

2. Two examples showing necessity of some conditions for Theorem 1

Before embarking on a proof of Theorem 1, we first present two examples to show that some restrictions are necessary for the conclusions of the theorem to hold. In particular, we exhibit families which do not satisfy condition (ii) for which the conclusion fails. It seems likely that condition (i) is unnecessary for the conclusion of the theorem to hold, although any significant relaxation of this condition will necessarily make the construction of the invariant measures much harder than the one given in this proof.

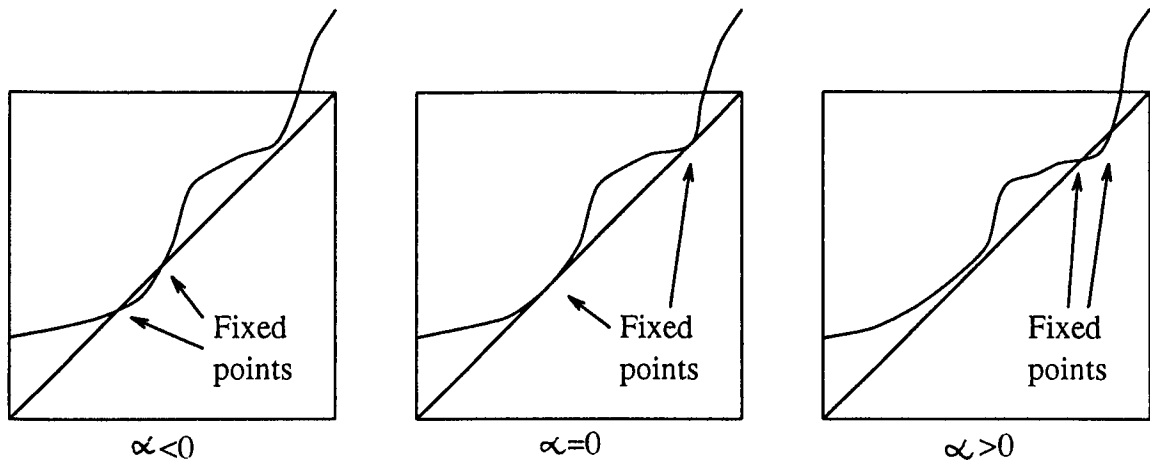


Figure 1.3 A family for which the conclusion of the theorem fails.

The first example for which the conclusion fails is illustrated graphically in Figure 1.3. The process which takes place is that a pair of fixed points vanish simultaneously with the birth of another pair of fixed points. The limits from the two sides in the parameter space of the invariant measures are concentrated on the dying pair (respectively new pair) for parameter values lower than (respectively greater than) the critical value (see Lemma 2).

The example shows that even in the one-dimensional case, there exist examples for which the unrestricted version of this theorem fails. The reliance of this proof on properties of circle maps suggests that this would fail more spectacularly in higher dimensions.

This example works by having a parameter value such that the probability measures for parameters on the left converge to a limit and similarly with parameters on the right, but that the two limits fail to agree. It is then natural to ask if this is the only way that the theorem could go wrong. In particular, if a parameter value is on the boundary of an interval on which the rotation number is rational, then this construction cannot be used. The question is then whether the condition (ii) needs to apply at the boundary of regions of constant rotation number.

The next example, which is more complicated, shows that the conclusion of the theorem need not hold even if condition (ii) fails only on the boundary of an interval in the parameter space on which the rotation number is constant. To write down the example, we regard the circle as the interval $[0,1) \bmod 1$. The maps which we consider are then of the form $T(x) = x + v(x) \bmod 1$. The form of the functions v which we are considering is shown in Figure 1.4.

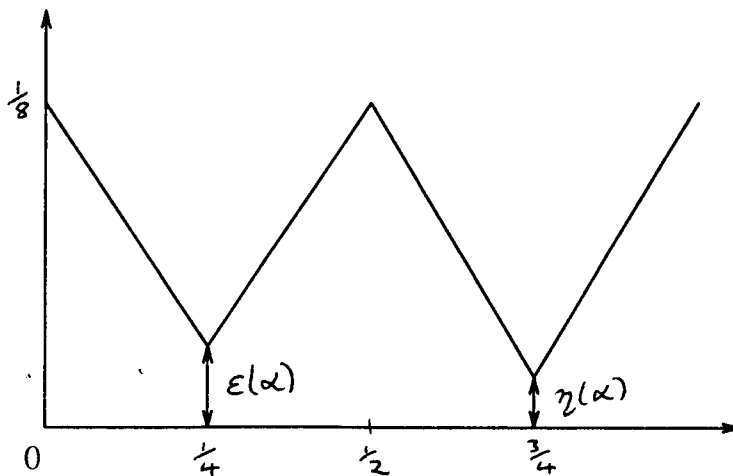


Figure 1.4 'Speed Function' for the counterexample.

The function v depends on the parameters ϵ and η . It is clear that if ϵ and η are allowed to vary continuously with respect to a parameter α say, then the family of circle maps given by $T_\alpha(x) = x + v_\alpha(x) \bmod 1$ is in fact a continuous family of circle maps. The family v_α is given explicitly by the expression

$$v_\alpha(x) = \begin{cases} \epsilon(\alpha) + 4\left(\frac{1}{8} - \epsilon(\alpha)\right)\left|x - \frac{1}{4}\right| & x \in [0, \frac{1}{2}] \\ \eta(\alpha) + 4\left(\frac{1}{8} - \eta(\alpha)\right)\left|x - \frac{3}{4}\right| & x \in [\frac{1}{2}, 1] \end{cases}$$

We then consider a family with the properties that $\epsilon(\alpha) \rightarrow 0$ and $\eta(\alpha) \rightarrow 0$ as $\alpha \rightarrow \alpha_0$, and investigate the limit of the invariant measures of the maps T_α as $\alpha \rightarrow \alpha_0$ and in particular, show that the limit exists if and only if $\log \epsilon / \log \eta$ has a well-defined limit

as $\alpha \rightarrow \alpha_0$. In this case, the limit is a measure μ concentrated at the points $\frac{1}{4}$ and $\frac{3}{4}$ with

$$\mu(\{\frac{1}{4}\}) = \lim_{\alpha \rightarrow \alpha_0} \frac{\log \epsilon(\alpha)}{\log \epsilon(\alpha) + \log \eta(\alpha)}, \quad \mu(\{\frac{3}{4}\}) = \lim_{\alpha \rightarrow \alpha_0} \frac{\log \eta(\alpha)}{\log \epsilon(\alpha) + \log \eta(\alpha)}. \quad (1)$$

It is clear that there exist examples of continuous functions $\epsilon(\alpha)$ and $\eta(\alpha)$ with the properties that $\epsilon(\alpha) \rightarrow 0$ and $\eta(\alpha) \rightarrow 0$ as $\alpha \rightarrow \alpha_0$, $\epsilon(\alpha) > 0$ and $\eta(\alpha) > 0$ for all $\alpha < \alpha_0$ such that the limit of $\log \epsilon / \log \eta$ fails to exist as $\alpha \rightarrow \alpha_0$.

It is a well-known fact of ergodic theory that each circle map T_α has some invariant measure, μ_α say (see [Wa3], corollary 6.9.1). To evaluate the limiting measure μ described above, we take a small set containing $\frac{1}{4}$, say $A = [\frac{1}{4} - \delta, \frac{1}{4} + \delta)$ and a similar one containing $\frac{3}{4}$, say $B = [\frac{3}{4} - \delta, \frac{3}{4} + \delta)$, and estimate $\mu_\alpha(A)$ and $\mu_\alpha(B)$ for $\alpha \rightarrow \alpha_0$. To do this, we note that if it takes between n and $n + 1$ steps 'for a point to go all the way around the circle' (that is if $0 \leq T_\alpha^{n+1}(0) < T_\alpha(0)$ and this is the first such n), and if it takes between m and $m + 1$ steps for a point to go through A (that is, if $T_\alpha^{m+1}(\frac{1}{4} - \delta) \geq \frac{1}{4} + \delta$ and this is the smallest such m), then

$$\mu_\alpha(A) = \int \chi_A d\mu_\alpha = \int \frac{1}{n} \sum_{i=0}^{n-1} \chi_A \circ T_\alpha^i d\mu_\alpha,$$

where χ_A is the characteristic function of the set A . This follows by invariance of the measure. But for each point, we have

$$\frac{m-1}{n} \leq \frac{1}{n} \sum_{i=0}^{n-1} \chi_A \circ T_\alpha^i(x) \leq \frac{m+1}{n},$$

It follows that $|\mu_\alpha(A) - \frac{m}{n}| \leq \frac{1}{n}$. There is of course a similar result for $\mu_\alpha(B)$. If we then show that the amount of steps in each cycle spent outside sets A and B is bounded above by some constant, then it is clear that we can evaluate the limit of $\mu_\alpha(A)$ as $\alpha \rightarrow \alpha_0$, by estimating the values of m and n , since these tend to infinity as $\alpha \rightarrow \alpha_0$. The only

calculation which we need to perform is to solve a simple recurrence relation to estimate the time spent in certain sets of a very simple form. Suppose then we are considering a set C of the form $[0, a)$ and the 'speed' function is given by $v(x) = c - (c - b)x/a$ where $c - b < a$, and we have $T(x) = x + v(x)$; then the recurrence relation is $x_{n+1} = c + (1 - (c - b)/a)x_n$. Let $\rho = 1 - (c - b)/a$. Then we are solving $x_{n+1} = c + \rho x_n$. The solutions are $x_n = c/(1 - \rho) + A\rho^n$. By substituting the initial conditions, we see that in fact $x_n = c(1 - \rho^n)/(1 - \rho)$. The number of steps thus spent in the set C is thus the rounded-up value of

$$\log(1 - (1 - \rho)a/c)/\log \rho = (\log b/c)/\log \rho.$$

We can now apply this to the collection of circle maps described above. In what follows, we will require ϵ and η to be bounded above by $\frac{1}{16}$. The first thing we show is that the amount of steps per cycle spent outside the sets A and B is bounded above by a constant as ϵ and η tend to zero. To show this, we note the symmetry of the situation: the number of steps taken to get from 0 to $\frac{1}{4} - \delta$ is the same as the number of steps to get from $\frac{1}{4} + \delta$ to $\frac{1}{2}$. This number is given by the round up of $\log(8 \cdot v(\frac{1}{4} - \delta))/\log(\frac{1}{2} + 4\epsilon)$. This is bounded above by $\log(4\delta)/\log(\frac{3}{4})$, so we see that the number of steps spent outside A and B is bounded above by $4\log(4\delta)/\log(\frac{3}{4})$. The number of steps in A is given by $-2\log \epsilon/\log(\frac{1}{2} + 4\epsilon)$ plus a term which is bounded, and similarly the number of steps in B is given by $-2\log \eta/\log(\frac{1}{2} + 4\eta)$ plus a bounded term. Set $m(\epsilon) = -2\log \epsilon/\log(\frac{1}{2} + 4\epsilon)$ and $p(\eta) = -2\log \eta/\log(\frac{1}{2} + 4\eta)$. Then given a constant $\sigma > 0$, there exists a τ such that if $\epsilon, \eta < \tau$, we get $|\mu_\alpha(A) - m(\epsilon)/(m(\epsilon) + p(\eta))| < \sigma$ and $|\mu_\alpha(B) - p(\eta)/(m(\epsilon) + p(\eta))| < \sigma$. By elementary analysis, we see that the assertion of equation (1) is now proved, and thus the example is complete.

3. Proof of Theorem 1

A useful lemma is the following:

Lemma 2. *The invariant Borel probability measures for a circle map T with rational rotation number p/q are precisely those measures which can be expressed in the form*

$$\mu(A) = \frac{1}{q} \sum_{i=0}^{q-1} \nu(T^{-i}A),$$

where ν is a probability measure concentrated on the fixed points of T^q .

Proof. Certainly any Borel probability measure of the form described is invariant for the circle map in question. Conversely, in the preliminary discussion, it was noted that each point of the circle converges monotonically under iteration of the map to a periodic orbit. From this, it follows that the only non-wandering points of the map are the periodic points. There is then a standard theorem telling us the non-wandering set has full measure (that is measure 1) with respect to any invariant Borel probability measure (see [Wa3], theorem 6.15). The remainder of the proof follows easily from the invariance of the measure.

Lemma 3. *Suppose that $(T_\alpha)_{\alpha \in J}$ is a continuous family of circle maps such that T_{α_0} has a hyperbolic periodic orbit of period q through a point $\xi \in S^1$. Then for each neighbourhood M of ξ , there exists a neighbourhood N of α_0 such that if $\beta \in N$, then T_β has a periodic point of period q in M .*

Proof. Suppose that we are given a neighbourhood M of ξ . Then there must exist a closed subinterval I of M with $\xi \in \text{Int}(I)$ with the property that $T_{\alpha_0}^q(I) \subset \text{Int}(I)$ or $T_{\alpha_0}^q(\text{Int}(I)) \supset (I)$ according to whether ξ is stable or unstable. But then for any T which is sufficiently close to T_{α_0} , the appropriate containment property persists ($T^q(I) \subset \text{Int}(I)$

or $T^q(\text{Int}(I)) \supset I$ respectively). But then it follows by Brouwer's fixed point theorem that T has a periodic point of period q in $\text{Int}(I)$.

Lemma 4. *Suppose X is a compact metric space and $T_0 : X \rightarrow X$ is a continuous map which is uniquely ergodic, with unique invariant measure ν_0 , say. Then for any weak*-neighbourhood N of ν_0 , there is a neighbourhood U of T_0 such that for $T \in U$ and ν any invariant measure for T , we have $\nu \in N$.*

Proof. By reducing N if necessary, we may first of all assume that N is a basic neighbourhood of ν_0 (that is $N = \{\mu : |\int f_i d\mu - \int f_i d\nu_0| < \epsilon_i, i = 1, \dots, n\}$ for a finite sequence $(f_i)_{1 \leq i \leq n}$ of continuous functions and $(\epsilon_i)_{1 \leq i \leq n}$ a finite sequence of positive bounds). We may further assume that $n = 1$ as for larger n , we may simply take the intersections of the resulting neighbourhoods U obtained from the proof below. We will therefore assume for this proof that the neighbourhood N is given by $N = \{\mu : |\int f d\mu - \int f d\nu_0| < \epsilon\}$. Now assume for a contradiction that for any neighbourhood U of T_0 , there is a map $T \in U$ and an invariant measure ν for T such that $|\int f d\nu - \int f d\nu_0| \geq \epsilon$. It follows that there exists a sequence of maps $(T_n)_{n \in \mathbb{N}}$ converging uniformly to T_0 , having invariant measures $(\nu_n)_{n \in \mathbb{N}}$ such that

$$\left| \int f d\nu_n - \int f d\nu_0 \right| \geq \epsilon, \forall n. \quad (2)$$

Since X is a compact metric space, the space of Borel probability measures on X is weak*-compact, hence weak*-sequentially compact. The sequence of measures (ν_n) therefore has a convergent subsequence, (ν_{n_i}) converging to μ , say. Since ν_{n_i} is invariant for T_{n_i} , we have for any continuous g , that

$$\int g \circ T_{n_i} d\nu_{n_i} = \int g d\nu_{n_i}, \forall i. \quad (3)$$

Now since T_{n_i} converges to T_0 uniformly, it follows that $g \circ T_{n_i}$ converges to $g \circ T_0$ uniformly, and hence, taking limits of (3) as $i \rightarrow \infty$, we see that $\int g \circ T_0 d\mu = \int g d\mu$ for any continuous function g . It follows that μ is an invariant measure for T_0 , yet taking the same limit in (2), we see that $\mu \neq \nu_0$. This contradicts our assumption that T_0 was uniquely ergodic, and hence proves the Lemma. \square

We now proceed to the proof of Theorem 1.

Proof of Theorem 1.

Set $C = \text{Cl}\{\alpha \in J : \rho(T_\alpha) \notin \mathbb{Q}\}$. We will show that for those values of α in C , the map T_α is uniquely ergodic. If $\rho(T_\alpha) \notin \mathbb{Q}$, then this is a standard ergodic theorem as noted earlier. If T_α has a hyperbolic periodic point, of period q say, then by Lemma 3, there is a neighbourhood of parameter values about α , such that for maps with parameters in the neighbourhood, there is a periodic point, and hence the rotation number is rational on a whole neighbourhood of parameter values about α . In particular, $\alpha \notin C$. It follows that if $\alpha \in C$ and $\rho(T_\alpha) \in \mathbb{Q}$ then T_α has no hyperbolic periodic points. We therefore see that T_α must have periodic points, and these must all be non-hyperbolic. By hypothesis (ii) of the theorem, we have that T_α has a unique periodic orbit. It now follows from Lemma 2, that there is a unique invariant probability measure.

We proceed by defining μ_α for $\alpha \in J \setminus C$. First, note that since C is a closed set, we have that $J \setminus C$ is an open subset of J and hence consists of a countable disjoint union of open subintervals of J , say J_1, J_2, \dots . Now fix such an interval J_i . Unless J_i is one of the end intervals, J_i is open in \mathbb{R} , so we write $J_i = (\alpha_i, \beta_i)$. In this case, T_{α_i} and T_{β_i} are uniquely ergodic, so the invariant measures are determined at the endpoints of the interval. If J_i is one of the end intervals, then we typically have that it is closed at one end or the

other. Now set $K_i = \text{Cl}(J_i) = [\alpha_i, \beta_i]$. The idea behind the construction is as follows. Plotting the periodic points of T_α against α for $\alpha \in K_i$ gives a graph similar to Figure 1.5. The invariant measure, being concentrated on the periodic points, must be chosen to be a superposition of ‘ δ -measures’, moving along the periodic point curves. Since these curves can terminate, it may be necessary to transfer to a new periodic curve. This must also be done continuously, so in the construction, one curve is being phased in, while another curve is being phased out.

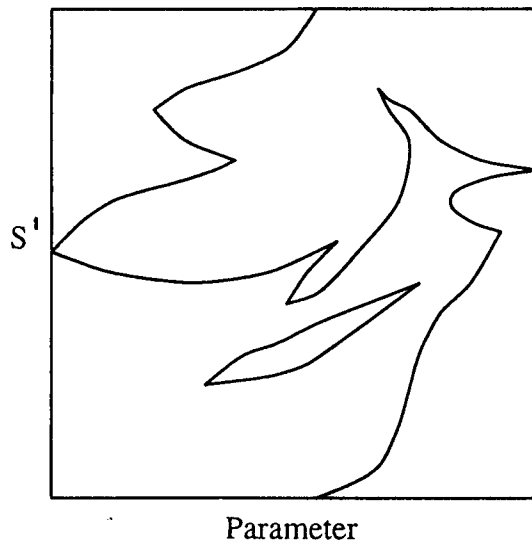


Figure 1.5 Typical diagram of periodic points against parameter.

We construct an open cover for K_i . Suppose the rotation number of the maps with parameters in K_i is p/q . This may be assumed by the continuity of ρ . Then let S_α be a lift of T_α^q fixing the preimages (under π) of the periodic points, and set $u_\alpha(x) = S_\alpha(x) - x$. Clearly the maps $(\alpha, x) \mapsto S_\alpha(x)$ and $(\alpha, x) \mapsto u_\alpha(x)$ are continuous on $K_i \times \mathbb{R}$. Given $\alpha \in K_i$, we seek a connected open neighbourhood N_α containing α , and a continuous map $\phi_\alpha : \text{Cl}(N_\alpha) \rightarrow \mathbb{R}$ taking each parameter value to a fixed point of S for that parameter value (that is $S_\beta(\phi_\alpha(\beta)) = \phi_\alpha(\beta)$).

If the periodic points of T_α are all non-hyperbolic, choose ξ to be any periodic point of T_α . Note that in this case, there is exactly one periodic orbit, so ξ is clearly bounded away from any other periodic points. If T_α has a hyperbolic periodic orbit, choose ξ to be a hyperbolic periodic point. In either case, there is a neighbourhood of ξ in which there are no other periodic points of T_α . Let x be a preimage (under π) of ξ . There is then a neighbourhood of x which contains no fixed points of S_α . Choose τ small such that $[x - \tau, x + \tau]$ is in this neighbourhood. Then set $\epsilon = \min(|u_\alpha(x + \tau)|, |u_\alpha(x - \tau)|)$. By continuity of u , there exists a $\delta_1 > 0$ such that $|\beta - \alpha| < \delta_1$ implies that $u_\beta \neq 0$ at $x - \tau$ and $x + \tau$. By hypothesis (i) of the theorem, we also have that there are finitely many values of β in $(\alpha - \delta_1, \alpha + \delta_1)$ with T_β critical, so it follows that there is a $\delta > 0$ such that $0 < |\beta - \alpha| < \delta$ implies that $u_\beta \neq 0$ at $x - \tau$ and $x + \tau$, and that T_β has no non-hyperbolic periodic orbits (note that there may be a non-hyperbolic orbit at α itself, but if so, it must lie outside $[x - \tau, x + \tau]$ or T_α must have no hyperbolic periodic orbits).

We then define $N_\alpha = \{\beta : |\alpha - \beta| < \delta\} \cap K_i$ and define ϕ_α on this reduced interval by the equation

$$\phi_\alpha(\beta) = \sup\{y \in [x - \tau, x + \tau] : u_\beta(y) = 0\}.$$

We claim that ϕ_α is continuous. If ϕ_α is not continuous, there exists a sequence $(\beta_i)_{i \in \mathbb{N}}$ of points in N_α tending to some $\beta \in N_\alpha$ such that $\phi_\alpha(\beta_i)$ fails to converge to $\phi_\alpha(\beta)$. By passing to a subsequence, we may assume that the $\phi_\alpha(\beta_i)$ converge to some other limit. If $\phi_\alpha(\beta)$ is smaller than this, then we get a contradiction by noting that $u_\beta(\lim \phi_\alpha(\beta_i)) = 0$, so that $\phi_\alpha(\beta)$ was in fact not the supremum of those fixed points in the range of interest. Conversely if $\phi_\alpha(\beta) > \lim \phi_\alpha(\beta_i)$, we must have $\beta \neq \alpha$ as T_α has only a single periodic point in $[x - \tau, x + \tau]$. But then T_β cannot be critical, so $\pi(\phi_\alpha(\beta))$ must be a hyperbolic periodic

point. The orbit at $\pi(\phi_\alpha(\beta))$ therefore persists for parameter values near β , which is a contradiction by construction of ϕ_α . This shows that ϕ_α is continuous, so for each $\alpha \in K_i$, there exists a neighbourhood N_α of α in K_i , and a continuous function ϕ_α defined on N_α , such that for all $\beta \in N_\alpha$, $\phi_\alpha(\beta)$ is a fixed point of S_β . By reducing the neighbourhood N_α if necessary, we may assume the additional properties that ϕ_α is continuous on the closure of N_α and that the only neighbourhoods containing α_i and β_i are N_{α_i} and N_{β_i} . We have then found an open cover of K_i , and so may apply compactness of K_i to pick a finite subcover. We may assume that this subcover is minimal by inclusion (that is there is no smaller subcover, each of whose sets is a member of our chosen subcover). We label the sets in the open cover in the order of the left-most point from left to right as N_1, N_2, \dots, N_k , and write $N_j = (a_j, b_j)$ for $1 < j < k$; $N_1 = [a_1, b_1)$; $N_k = (a_k, b_k]$, where we have taken $b_k = \beta_i$ and $a_1 = \alpha_i$. Let ϕ_j be the ϕ -function associated to the interval N_j . We then have

$$a_1 < a_2 < b_1 \leq a_3 < b_2 \leq a_4 < \dots \leq a_k < b_{k-1} < b_k$$

by the minimality of the cover. To see this, note that clearly the sequence of a_i is increasing by construction. The sequence of b_i must also be increasing, since otherwise one of the intervals would be completely contained in another. We need that $N_j \cup N_{j+2} \not\subset N_{j+1}$ giving the condition that $b_j \leq a_{j+2}$, and the condition $a_{j+1} < b_j$ arises from the requirement that the collection be a cover.

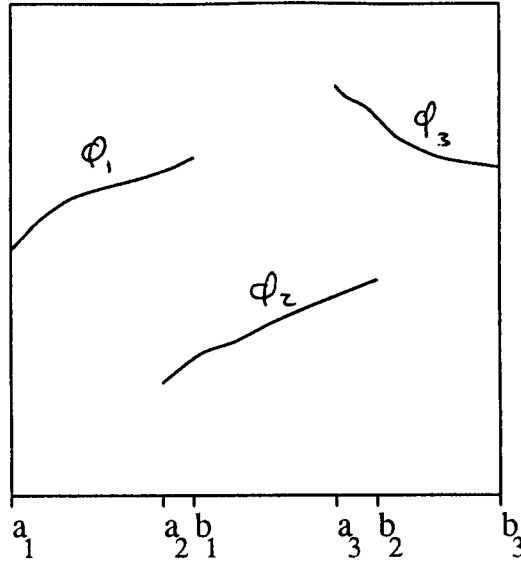


Figure 1.6 Possible arrangement of chosen periodic points.

In Figure 1.6, an example of such a configuration is shown. We are then in a position to construct the invariant measures for the T_α with $\alpha \in K_i$. Define

$$\nu_\alpha = \begin{cases} \frac{b_j - \alpha}{b_j - a_{j+1}} \delta_{\pi(\phi_j(\alpha))} + \frac{\alpha - a_{j+1}}{b_j - a_{j+1}} \delta_{\pi(\phi_{j+1}(\alpha))} & \text{if } \alpha \in [a_{j+1}, b_j] \\ \delta_{\pi(\phi_{j+1}(\alpha))} & \text{if } \alpha \in [b_j, a_{j+2}], \end{cases}$$

where δ_ζ is the Dirac δ -measure with unit mass concentrated at ζ and where we take $a_{k+1} = b_k$ and $b_0 = a_1$. Given a continuous function $f \in C(S^1)$,

$$\int f d\nu_\alpha = \begin{cases} \frac{b_j - \alpha}{b_j - a_{j+1}} f(\pi(\phi_j(\alpha))) + \frac{\alpha - a_{j+1}}{b_j - a_{j+1}} f(\pi(\phi_{j+1}(\alpha))) & \text{if } \alpha \in [a_{j+1}, b_j] \\ f(\pi(\phi_{j+1}(\alpha))) & \text{if } \alpha \in [b_j, a_{j+2}]. \end{cases}$$

Continuity is clear everywhere except at the a_i and b_i , and this can be checked by comparing the expressions and using the fact that the ϕ -functions were chosen to be continuous on the closures of the subintervals. We can then see that the family $(\nu_\alpha)_{\alpha \in K_i}$ is a continuous family of probability measures, and ν_β is concentrated on the periodic points of T_β . Forming

$$\mu_\alpha = \frac{1}{q} \sum_{i=0}^{q-1} \nu_\alpha \circ T_\alpha^{-i}$$

gives a continuous family of invariant measures on K_i by Lemma 2. Notice that by the construction, since we forced $N_1 = N_{\alpha_i}$ and $N_k = N_{\beta_i}$, the limit of the measures as they approach the end-points is just the required measure, in the (usual) case where this is uniquely ergodic.

Repeating this process inductively, we will be able to define a family of invariant Borel probability measures, one measure for each parameter in the set $J \setminus C$, and so since we have already shown the uniqueness of the probability measures for maps with parameters lying in C , we have defined the whole family of invariant measures. The family thus constructed has already been shown to be continuous on all intervals contained in $J \setminus C$, and therefore, since $J \setminus C$ is an open set, it follows that the map $M : \alpha \mapsto \mu_\alpha$ is continuous for $\alpha \in J \setminus C$. It remains to show continuity at points of C , but this is a straightforward application of Lemma 4, so we are done. \square

Chapter 2. Expanding Maps, g-Measures and Generalized Baker's Transformations

1. Introduction

In this chapter, we discuss the relationship between expanding maps, g -measures and generalized baker's transformations.

Throughout this chapter, we take a definition of expanding maps which is slightly different from the traditional one, as we include also the possibility that the map is not differentiable. I will denote the unit interval. For a subinterval J of I , $|J|$ will denote the length of the interval J .

Note that in what follows, we will often refer to maps which are piecewise monotone and continuous or piecewise monotone and C^k . These mean that the map is piecewise strictly monotonic and on each of those pieces the map is continuous or C^k respectively. In the latter case, the map is assumed to have a C^k extension to the closure of any interval of monotonicity.

Definition 1. *Let $T : I \rightarrow I$ be piecewise monotone and continuous. T is expanding if there exists a constant $C > 1$ such that whenever J is a subinterval of I , for which the restriction of T to J is a homeomorphism, we have $|T(J)| \geq C|J|$.*

Definition 2. *An expanding map T will be called Markov if it has the additional properties:*

- (i) There is a finite partition of I into subintervals $I = I_0 \cup \dots \cup I_{n-1}$ such that the restriction of T to each of these subintervals is a homeomorphism.
- (ii) $\text{Cl}(T(\text{Int } I_i))$ is a non-empty union of some of the $\text{Cl}(I_j)$.

Note that in this situation, we can define an associated *Markov matrix* A by setting $A_{ij} = 1$ if $\text{Cl}(T(\text{Int } I_i)) \supset I_j$ and 0 otherwise. This matrix is said to be *mixing* if the entries of A^n are all strictly positive for some $n > 0$.

Note also that this is also at variance with the definition of the Markov property given in [Ma], where additional continuity/differentiability properties are required of T .

There is another situation, in which we frequently find ourselves, so this will be given its own definition.

Definition 3. A map T from the interval to itself will be called a *full map* if it is expanding, has a finite partition of I into subintervals $I = I_0 \cup \dots \cup I_{l-1}$ such that the restriction of T to the interior of each subinterval is an orientation-preserving homeomorphism onto $(0,1)$. T will be called a C^k *full map* if it has a C^k extension to each of the intervals $\text{Cl}(I_i)$. The *degree* of the map is l , the number of branches.

The reason for this nomenclature is that the symbolic dynamics associated with T take place on the full l -shift (see below).

Let T be an expanding map of the interval. An *absolutely continuous invariant measure* (or ACIM) for T is a Borel probability measure which is absolutely continuous with respect to Lebesgue measure and is invariant under T .

Many authors have studied the existence and the number of such measures for expanding maps T of the interval. Krzyżewski and Szlenk showed that for C^2 expanding maps of compact manifolds (that is C^2 maps whose Jacobian is everywhere bounded below by some

$C > 1$), there is a unique ACIM (see [KS] and [Kr1]). This applies to those maps of the interval which are obtained from C^2 expanding endomorphisms of the circle. Lasota and Yorke showed in [LaY] that any piecewise C^2 expanding map of I has an ACIM. Kowalski ([Ko]) improved this by showing that the same conclusion holds if the map is piecewise C^{1+1} (that is the map has Lipschitz derivative). Mañé's book ([Ma]) gives a refined proof showing that this remains true if the map is piecewise $C^{1+\alpha}$. Wong ([Wo]) found that the conclusion holds when the assumption is altered to assuming that the map is piecewise C^1 with the reciprocal of the derivative, $1/T'$, of bounded variation.

Krzyżewski ([Kr2]) managed to show that the same conclusions do not in general hold for C^1 maps by showing that for any manifold M , there exist C^1 expanding maps of M which do not have any ACIM. His proof however was not constructive, so there was still some interest in constructing an explicit example of such a C^1 map in (for example) the simple case of the circle. This was done by Gora and Schmitt (see [GS]).

Various authors then turned their attentions to the number of ergodic ACIMs in the piecewise C^2 case (where ACIMs are known to exist). Papers on this include [LiY], [BS] and [BB]. These in particular imply that if T is a C^2 full map, then there is a unique ACIM for the map T . Such an ACIM would therefore necessarily be ergodic.

A natural question which remains is the following:

Question 1. *Does there exist a C^1 full map with more than 1 ACIM?*

This question has been recently answered for C^0 maps and even for Lipschitz maps in the affirmative: There exist relatively simple examples of Lipschitz full maps which preserve Lebesgue measure, but for which Lebesgue measure is not ergodic. This was first answered by Bose in [Bos1] using generalized baker's transformations. I have since

found a simpler, but less geometric proof of this result using g -measures, which is presented below. This chapter contains a demonstration of the relationship between the two approaches, and in general exhibits the connection between expanding maps, generalized baker's transformations and g -measures.

The general question which remains then is to see what constraints are imposed on a system by assuming that it is a C^1 full map of I , preserving Lebesgue measure. In particular, is such a system automatically ergodic? The results described below stem from an attempt, as yet unsuccessful, to answer Question 1.

We now summarize the theory of g -measures. Let A be a mixing Markov $l \times l$ matrix as described above. We will assume the indices of A run from 0 to $l - 1$. Then Σ_A is the space of sequences defined by

$$\Sigma_A = \{x \in \{0, 1, \dots, l - 1\}^{\mathbb{Z}^+} : A_{x_i, x_{i+1}} = 1, \forall i \geq 0\}.$$

This space is endowed with the induced topology on Σ_A of the product topology on $\{0, 1, \dots, l - 1\}^{\mathbb{Z}^+}$ by giving it the metric

$$d(x, y) = \begin{cases} 0 & \text{if } x = y; \\ 2^{-n} & \text{if } x_i = y_i \text{ for } i = 0, 1, \dots, n - 1, \text{ but } x_n \neq y_n. \end{cases}$$

We then consider the map $\sigma : \Sigma_A \rightarrow \Sigma_A$ defined by $\sigma(x)_n = x_{n+1}$. The map σ is commonly known as the *shift map*. The topological space (X, d) , together with the map σ acting on it is known as a *mixing subshift of finite type*. We will often work with the special case where A is the $l \times l$ matrix consisting entirely of 1s. In this case, Σ_A is the space of all sequences of symbols of $\{0, 1, \dots, l - 1\}$, and is denoted now by Σ_l . This space (together with the map σ) is known as the *full shift on l symbols*. Most of this chapter will concentrate on this restricted situation. We will be looking at those Borel probability measures which

are invariant under the action of the shift map. These measures are called *shift-invariant*. Suppose now that Σ_A is a mixing subshift of finite type. Let $a = (a_0, a_1, \dots, a_{s-1})$ be a finite (possibly empty) word satisfying $A_{a_i, a_{i+1}} = 1$ for each i , let $x \in \Sigma_A$ and suppose $A_{a_{s-1}, x_0} = 1$, then denote by ax , the sequence in Σ_A given by concatenating a onto the front of x :

$$(ax)_i = \begin{cases} a_i & \text{if } i < s; \\ x_{i-s} & \text{if } i \geq s. \end{cases}$$

If $x \in \Sigma_A$, then we define $[x]^n$ to be the n th cylinder about x : $[x]^n = \{y : d(x, y) < 2^{-n}\}$. If $f \in C(\Sigma_A)$, the n th variation of f is given by $\text{var}_n(f) = \max\{|f(x) - f(y)| : x, y \in \Sigma_A, d(x, y) < 2^{-n}\}$. The function f is Lipschitz if there exists a $C > 0$ such that $\text{var}_n(f) \leq C \cdot 2^{-n}$ for all n . It is Hölder if there exists a $C > 0$ and a $\beta < 1$ such that $\text{var}_n(f) \leq C \cdot \beta^n$.

We are now in a position to start defining g -measures.

Let $g : \Sigma_A \rightarrow [0, 1]$ be a Borel-measurable function such that $\sum_{y \in \sigma^{-1}(x)} g(y) = 1$ for all $x \in \Sigma_A$. Write \mathcal{G} or $\mathcal{G}(\Sigma_A)$ for the set of all such functions. The subclass of those g which are bounded below by a positive number will be denoted by \mathcal{G}^+ . We will usually restrict attention to the subclass of those $g \in \mathcal{G}$ which are continuous and strictly bounded away from 0. We will write \mathcal{G}^0 for this class of functions. Given $g \in \mathcal{G}^0$, define the *Ruelle-Perron-Frobenius operator* $\mathcal{L}_g : C(\Sigma_A) \rightarrow C(\Sigma_A)$ as follows:

$$\mathcal{L}_g f(x) = \sum_{y \in \sigma^{-1}(x)} g(y) f(y).$$

This is a positive operator and it satisfies $\mathcal{L}_g 1 = 1$ for all $g \in \mathcal{G}^0$. Since \mathcal{L}_g is a linear map defined on the Banach space of continuous functions on Σ_A , it has a dual map \mathcal{L}_g^* which maps the space of finite signed measures on Σ_A into itself. The defining relation for

\mathcal{L}_g^* is then $\int \mathcal{L}_g f d\mu = \int f d\mathcal{L}_g^* \mu$. The above facts noted about \mathcal{L}_g imply that \mathcal{L}_g^* maps the probability measures on Σ_A into themselves.

A *g-measure* for $g \in \mathcal{G}^0$ is simply a probability measure ν such that $\mathcal{L}_g^* \nu = \nu$.

The following Lemma records some elementary properties of *g-measures*.

Lemma 1. *The following properties of g-measures hold.*

- (i) *For each $g \in \mathcal{G}^0$, there is at least one g-measure;*
- (ii) *Any g-measure is shift-invariant;*
- (iii) *Any g-measure is fully supported on Σ_A ;*
- (iv) *A g-measure may be characterized by the property*

$$\lim_{n \rightarrow \infty} \frac{\mu([x]^n)}{\mu([\sigma(x)]^{n-1})} = g(x); \quad (1)$$

- (v) *For any given $g \in \mathcal{G}^0$, the g-measures form a non-empty convex set. The extreme points of this set are ergodic.*
- (vi) *All g-measures are non-atomic.*

Proof. To show (i) holds, let μ be any probability measure on Σ_A . Form the averages

$$\mu^{(n)} = \frac{1}{n} \sum_{i=0}^{n-1} \mathcal{L}_g^{*i} \mu.$$

Then since Σ_A is compact, there is a weak*-convergent subsequence of $\mu^{(n)}$, say $\mu^{(n_i)}$ converging to some measure ν . Then for any continuous function f , we have $|\int f d\mu^{(n_i)} - \int f d\mathcal{L}_g^* \mu^{(n_i)}| \leq 2\|f\|/n$. Taking limits, it follows that $\int f d\nu = \int f d\mathcal{L}_g^* \nu$. The measure ν is therefore a *g-measure*, completing the proof of (i).

Now note that $\mathcal{L}_g(f \circ \sigma) = f$ for any continuous f . Using this, and supposing that ν is a *g-measure*, we have

$$\int f \circ \sigma d\nu = \int f \circ \sigma d\mathcal{L}_g^* \nu = \int \mathcal{L}_g(f \circ \sigma) d\nu = \int f d\nu$$

for any continuous function f . It follows that ν is shift-invariant, showing (ii).

We prove (iii) by contradiction. Suppose ν is a g -measure which is not fully supported. Then there must be an open subset U of Σ_A such that $\nu(U) = 0$. We may therefore assume that U is a basic set: an n -cylinder for some $n > 0$. So now write $U = [z]^n$. Write χ_U for the characteristic function of U . We are assuming that A is a mixing Markov matrix, so let k be chosen such that A^k has strictly positive entries. Then pick $x \in \Sigma_A$. We have $A_{z_n, x_0}^k > 0$. It follows that there exists a sequence y_0, \dots, y_{k-1} such that $A_{y_i, y_{i+1}} = 1$ for each i , $A_{z_n, y_0} = 1$ and $A_{y_{k-1}, x_0} = 1$. In particular, the concatenation zyx is a member of Σ_A . Now, we have

$$\begin{aligned} \mathcal{L}_g^{n+k+1} \chi_U(x) &= \sum_{u \in \sigma^{-(n+k+1)}(x)} \chi_U(u) g(u) g(\sigma(u)) \dots g(\sigma^{n+k}(u)) \\ &\geq \chi_U(w) g(w) g(\sigma(w)) \dots g(\sigma^{n+k}(w)) \text{ where } w = zyx \\ &= g(w) g(\sigma(w)) \dots g(\sigma^{n+k}(w)). \end{aligned}$$

But now, since g is strictly bounded away from 0, this is strictly positive. Since this is true for all x , we now have that $\int \mathcal{L}_g^{n+k+1} \chi_U d\nu > 0$, but this implies $\int \chi_U d\nu > 0$, which is the desired contradiction.

To prove (iv), let μ be any probability measure. Now extend the definition of g by saying that $g : \Sigma_l \rightarrow [0, 1]$, where $g(x) = 0$ if $x \in \Sigma_l \setminus \Sigma_A$. Similarly, we may regard μ as a measure on Σ_l by the natural inclusion. We now have

$$\begin{aligned} \mathcal{L}_g^* \mu([ix]^{n+1}) &= \int \chi_{[i]^0}(\chi_{[x]^n} \circ \bar{\sigma}) d\mathcal{L}_g^* \mu \\ &= \int \mathcal{L}(\chi_{[i]^0}(\chi_{[x]^n} \circ \sigma)) d\mu = \int_{[x]^n} g(iy) d\mu(y). \end{aligned}$$

In particular, if μ is a g -measure, then the desired conclusion holds using the continuity of

g. Notice that this implies the important equation

$$d\mathcal{L}_g^*\mu(ix) = g(ix)d\mu(x). \quad (2)$$

Conversely, suppose (1) holds. Then write $\mu_i(A)$ for the quantity $\mu(iA)$. Then μ_i is a measure. The equation (1) implies that $d\mu_i/d\mu(x) = g(ix)$, or $d\mu(ix) = g(ix)d\mu(x)$. Comparing with (2), we see that μ has the same derivative as $\mathcal{L}_g^*\mu$. This implies that $\mathcal{L}_g^*\mu = \mu$, and part (iv) is proved.

In showing that (v) holds, note that it is clear that the set of *g*-measures is a non-empty (by (i)) convex set since the operator \mathcal{L}_g^* is affine. Suppose now that μ is an extreme point of this set of *g*-measures, and suppose for a contradiction that μ is not ergodic. Then there exists a set B such that $\mu(B) \in (0, 1)$ and such that $B = \sigma^{-1}(B)$. Now form a new measure ν in the usual way by defining $\nu(A) = \mu(A \cap B)/\mu(B)$. Now we have

$$\mu = (1 - \mu(B)) \left(\frac{\mu - \mu(B)\nu}{1 - \mu(B)} \right) + \mu(B)\nu,$$

so that if we can prove that ν is still a *g*-measure, then we are done. To show this, let f be a continuous function and note that

$$\mathcal{L}_g(f\chi_B)(x) = \sum_{y \in \sigma^{-1}(x)} g(y)f(y)\chi_B(y).$$

Note that if $y \in \sigma^{-1}(x)$, then $\chi_B(y) = 1$ if and only if $\chi_B(x) = 1$ for μ -almost all x . It follows that $\mathcal{L}_g(f\chi_B)(x) = \mathcal{L}_g f(x)\chi_B(x)$ for μ -almost all x . We now get

$$\begin{aligned} \int f d\mathcal{L}_g^*\nu &= \int \mathcal{L}_g f d\nu = \int (\mathcal{L}_g f) \frac{\chi_B}{\mu(B)} d\mu \\ &= \int \mathcal{L}_g \left(f \frac{\chi_B}{\mu(B)} \right) d\mu = \int f \frac{\chi_B}{\mu(B)} d\mu = \int f d\nu. \end{aligned}$$

It follows that ν is a g -measure, so by the earlier comments, we have achieved the desired contradiction, proving (v).

Now, suppose x is a non-periodic point of Σ_A . Then it is easy to check that $\sigma^{-m}(x) \cap \sigma^{-n}(x) = \emptyset$ for each $m > n \geq 0$: Suppose not. Assume that $y \in \sigma^{-m}(x) \cap \sigma^{-n}(x)$. Then $\sigma^m(y) \in \{x\} \cap \{\sigma^{m-n}(x)\}$, which establishes a contradiction. Using this, we see that any atoms of an invariant probability measure must be concentrated on its periodic points, for otherwise, if x is a non-periodic atom, then the sets $\sigma^{-n}(x)$ have equal positive measure and are disjoint, which contradicts the finiteness of the measure. Now suppose ν is a g -measure and that x is an atom of ν . Then x must be periodic since g -measures are shift-invariant. Next, let n be such that A^n has strictly positive entries where A is the associated Markov matrix. Then there are at least l elements of $\sigma^{-n}(x)$. Only one of these can be periodic (namely the one which has the n terms which are added on copied from x itself). Let y be one of the non-periodic preimages of x . From (1), we can see that $\nu(\{y\}) = g(y)g(\sigma(y)) \dots g(\sigma^{n-1}(y))\nu(\{x\})$. Hence $\nu(y) > 0$, which is a contradiction by the argument above, thus proving (vi).

This completes the proof of the Lemma. \square

Note that as yet, we have only defined g -measures for $g \in \mathcal{G}^0$. However, we use (1) to define g -measures for general $g \in \mathcal{G}$. Since (vi) only relies upon (1) and the fact that g is positive, it follows that the conclusion of (vi) holds for $g \in \mathcal{G}^+$.

We will now describe a more probabilistic interpretation of g -measures. We consider sequences $(X_n)_{n \in \mathbb{Z}}$ of random variables taking values in the set $\{0, \dots, l-1\}$, often regarding their values as outcomes of a sequence of experiments, one performed at each integer time. Strictly, one should consider the X_n as maps from some probability space Ω

to $\{0, \dots, l-1\}$, and write $X_n(\omega)$ for X_n , but as we will be using the same probability space throughout, we often prefer to simply write X_n . We will look at the evolution of the random variables by specifying the probabilities of the various outcomes of the ‘present’ experiment (that is X_0) conditional on the ‘past’ (that is $(X_n)_{n<0}$). The simplest non-trivial examples of this are given by Markov chains, where the probabilities of the outcomes of the present experiment are completely determined by outcome of the previous one (that is $\mathbb{P}(X_n = i | X_{n-1} = j_1, X_{n-2} = j_2, \dots)$ is independent of j_2, j_3, \dots). One can similarly consider the so-called ‘finite range’ processes or *k-step Markov chains*, where the probabilities are determined by the outcomes of the previous k experiments.

We will look at a generalization of these to ‘infinite range’ processes. Let $(X_n)_{n \in \mathbb{Z}}$ be a sequence of random variables taking values in $\{0, \dots, l-1\}$. Suppose the sequence satisfies

$$\mathbb{P}(X_n = i | X_{n-1} = a_1, X_{n-2} = a_2, \dots) = g(i, a_1, a_2, \dots), \quad (3)$$

where $g \in \mathcal{G}^0(\Sigma_l)$. If we now fix an n , then we get a natural map $\rho_n : \Omega \rightarrow \Sigma_l$ given by $\rho_n(\omega)_i = X_{n-i}(\omega)$. If we have a probability distribution on the subsequence $(X_m)_{m \leq n}$, then this pushes forward (under ρ_n) to a probability measure μ on Σ_l . If the evolution at the $n+1$ st stage is governed by g , as in (3), then the induced probability distribution on $(X_m)_{m \leq n+1}$ pushes forward under ρ_{n+1} to $\mathcal{L}_g^* \mu$. This follows by (2), which just says that the probability of adding an i on the front of the sequence x is given by $g(ix)$. It then follows that the stationary distributions for the random variables correspond exactly to g -measures: If \mathbb{P} is a stationary probability distribution on Ω , satisfying (3), then by stationarity, we have $\rho_n(\mathbb{P})$, the push-forward of the distribution on those symbols before the n th is independent of n . Call this measure ν , say. It follows that $\mathcal{L}_g^* \nu = \nu$, so ν is a

g -measure. Clearly this also works in reverse.

The first use of g -measures was in the 1930s to describe the so-called Learning Models, where people were interested in finding a mathematical description of the processes of learning. These were studied by Doeblin and Fortet in [DF], where they were called chains with complete connections. Karlin (see [Kar]) also looked at these and claimed he had a proof that for each $g \in \mathcal{G}^0$, there is a unique g -measure. This proof was however incorrect, and this statement is now known to be false. Keane ([Kea]) invented the name ‘ g -measures’ and showed that for a large class of g , there exists a unique g -measure, and this measure has strong ergodic properties with respect to the shift transformation. In fact, in [Kea], he works on the circle, using instead of the map σ , the map $T : x \mapsto 2x \pmod{1}$. The results may be readily translated to the situation which we are discussing. In this context, Keane’s results state that if g is Lipschitz then there exists a unique g -measure, which is strong-mixing. Keane asked whether there exists a unique g -measure for each $g \in \mathcal{G}^0$, which is very closely related to the question left open by Karlin’s wrong proof. Walters (see [Wa1]) then showed that there is a unique g -measure when g has summable variation (that is $\sum_{n=1}^{\infty} \text{var}_n(g) < \infty$). This holds in particular, when g is Hölder continuous. Palmer, Parry, and Walters took up the question of uniqueness of g -measures in [PPW], but their attempt yielded only some preliminary results. More recently, Berbee ([Be]) considered the question, providing weaker conditions than those of Walters, under which there exists a unique g -measure. It may be noted that the development of results for g -measures is similar to the development of results for expanding maps. The reasons for this are discussed in the next section.

Hulse ([Hu]) applied some ideas of statistical mechanics to find a new class of g

which have unique g -measures. This paper was interesting as the result followed from general statistical mechanical restrictions on g , rather than strong continuity conditions. In particular, Hulse introduced the definition of attractive g -functions. He worked mainly on Σ_2 , and introduced a partial order \preceq on it:

$$x \preceq y \text{ if } x_i \leq y_i, \forall i \geq 0.$$

A g -function is then *attractive* if $g(1x) \geq g(1y)$ whenever $x \succeq y$. This says (in the probabilistic interpretation) that the more 1s that one has in the past, the more likely one is to get a 1 at the present. One important consequence of this shown in [Hu] is that the sequence $\mathcal{L}_g^{*n} \delta_i$ is weak*-convergent to a g -measure, where δ_i is the probability measure concentrated on the point of Σ_2 whose terms are all equal to i . Normally, to get a g -measure, one is compelled to take subsequences of Césaro averages as in Lemma 1, but then one typically has very bad control of the resulting measure. In [Kal], Kalikow introduced the concept of bounded uniform martingales (which he gave the unfortunate acronym b.u.m.), which is equivalent to the concept of g -measures. Finally, in [BK], Bramson and Kalikow used this and attractive g -measures to provide an example of a $g \in \mathcal{G}^0$ for which there is more than one g -measure. This finally solved the main problem, which had been a major conjecture for a considerable time. It does not however solve the problem of Keane in its original form, as there is in general a difficulty in lifting functions from Σ_2 to the circle, so the example of Bramson and Kalikow may not be lifted into the context of Keane.

The third concept, which we shall require in this chapter is that of generalized baker's transformations, as introduced by Bose (see [Bos1]). For the purposes of describing this, we will consider generalized baker's transformations with two slices, although it is possible to

look at generalized baker's transformations with more slices. Let f_0 and f_1 be measurable functions on $[0,1]$ such that $f_0 + f_1 = 1$ almost everywhere and $f_i \geq C$ almost everywhere, with respect to Lebesgue measure λ for $i=0$ and 1 . Then define

$$\phi_0^*(x) = \int_0^x f_0(t)dt$$

$$\phi_1^*(x) = \phi_0^*(1) + \int_0^x f_1(t)dt.$$

These maps are homeomorphisms of the interval onto their images, and the union of their images is the whole interval. There is therefore a 2-branched expanding Lebesgue measure-preserving map ϕ , of which ϕ_0 and ϕ_1 are the two inverse branches. Then the *generalized baker's transformation* is defined as follows:

$$T_f(x, y) = \begin{cases} (\phi(x), f_0(\phi(x))y) & \text{if } x < c, \\ (\phi(x), 1 - (f_1(\phi(x))(1 - y))) & \text{if } x \geq c, \end{cases}$$

where $c = \phi_0^*(1)$. We will refer to this also as the *generalized baker's transformation based on ϕ* , as the map ϕ is easily seen to determine the whole transformation, by noting that $\phi'(x) = 1/f_i'(\phi(x))$ almost everywhere, where i is 0 if $x < c$ and 1 if $x \geq c$. Note that if we consider the projection p sending points of S onto their first coordinate then $p \circ T_f = T_f \circ p$, so the pair $(T_f, \lambda \times \lambda)$ may be factored through this projection. The result of this is just the pair (ϕ, λ) . We call ϕ the *vertical projection* of T_f .

This situation is illustrated in Figures 2.1 and 2.2 below:

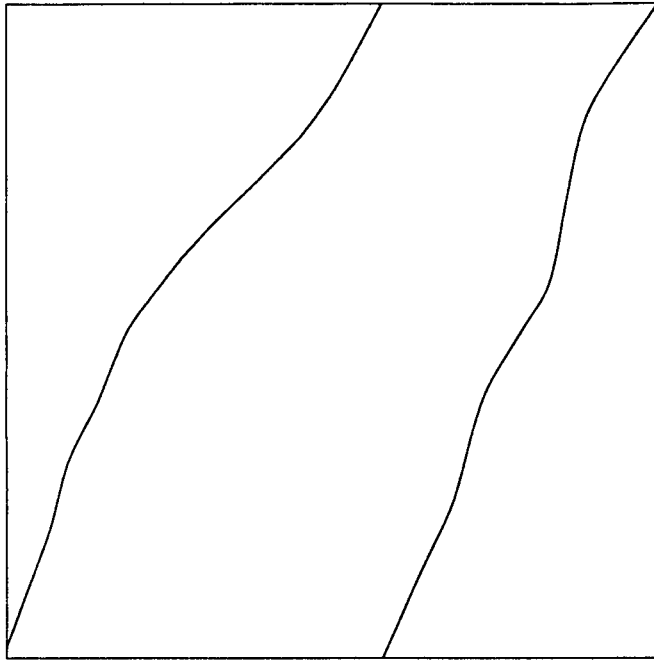


Figure 2.1 Possible graph of ϕ .

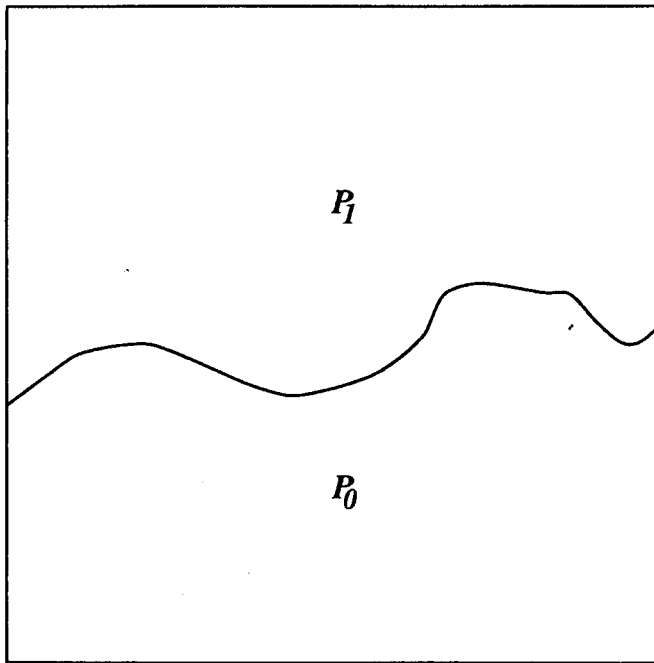


Figure 2.2 Fundamental partition of S under T_f .

The transformation operates as follows. The square is divided into two rectangles: $R_0 \equiv \{(x, y) : x < c\}$ and $R_1 \equiv \{(x, y) : x \geq c\}$. The rectangle R_0 is then stretched

(non-uniformly) horizontally to have width 1, with x moving to $\phi(x)$. There is a corresponding vertical contraction so that areas are preserved. The image of R_0 is then called P_0 . Meanwhile, R_1 is flipped vertically and then stretched (also with x moving to $\phi(x)$) in such a way as to fill the remainder of the square lying over the image of R_0 . The image of R_1 is P_1 .

In [Bos1], Bose shows how to find generalized baker's transformations which are measurably isomorphic to stationary stochastic processes. Recently, Rahe (see [Ra]) has related generalized baker's transformations and the work of Kalikow to represent generalized baker's transformations as uniform martingales. [Bos2] uses generalized baker's transformations to construct examples of C^0 expanding, Lebesgue measure-preserving maps of I with varying degrees of ergodic properties.

2. Connections between g -Measures, Expanding Maps and Generalized Baker's Transformations

In this section, we show how the concepts of expanding maps, g -measures and generalized baker's transformations are related. First, we show how to use Walters' result [Wa1] to give a quick proof of a simple, but archetypal result for expanding maps. This is somewhat similar to the rather more general proof given in [Wa2].

Proposition 2. *Suppose $T : I \rightarrow I$ is a Markov expanding map which is piecewise $C^{1+\alpha}$, (that is the derivative is piecewise Hölder continuous with exponent α). Suppose further that T has a mixing associated Markov matrix A . Then T preserves a unique ACIM.*

Proof. Let the constant in the definition of expanding maps be C , where $C > 1$. We use standard symbolic dynamics arguments to get a topological semi-conjugacy π from (Σ_A, σ) to (I, T) such that $T^n(\pi(x)) \in I_{x_n}$ for each $n \geq 0$. This semi-conjugacy is one-to-one off

a countable set, which is, of course, of Lebesgue measure 0. Lebesgue measure λ therefore pulls back under π to give a measure μ defined on Σ_A . The triples (T, I, λ) and (σ, Σ_A, μ) are therefore measure-theoretically isomorphic, although μ is not a shift-invariant measure. We then define $g(x)$ to be $1/T'(\pi(x))$, taking the appropriate one-sided derivatives at the endpoints of the intervals I_j (that is if x is the left hand endpoint of I_j , then $T'(x)$ is taken to be the right derivative of T at x).

Now take any function f on I . Then we have by the change of variables formula,

$$\begin{aligned} \int_0^1 f(x)dx &= \sum_i \int_{I_i} f(x)dx = \sum_i \sum_{\{j:A_{ij}=1\}} \int_{I_j} \frac{f(T_i^{-1}(y))}{|T'(T_i^{-1}(y))|} \\ &= \sum_j \int_{I_j} \sum_{\{i:A_{ij}=1\}} \frac{f(T_i^{-1}(y))}{|T'(T_i^{-1}(y))|} \\ &= \int_I \sum_{\{i:A_{ij}=1\}} \frac{f(T_i^{-1}(y))}{|T'(T_i^{-1}(y))|}, \end{aligned}$$

where T_i is the restriction of T to I_i . This says (under the isomorphism) that

$$\int \mathcal{L}_g f d\mu = \int f d\mu \quad (4)$$

for any continuous function f on Σ_A .

Next, we check that g is Hölder continuous. If $d(x, y) < 2^{-n}$, then $\pi(x)$ and $\pi(y)$ lie in the same n -cylinder. But the n -cylinders are mapped homeomorphically by T^n into I . In particular, since T expands distances by at least C , we have that the length of the n -cylinder is bounded above by C^{-n} . This means that $|\pi(x) - \pi(y)| \leq C^{-n}$. It follows that $|g(x) - g(y)| \leq K \cdot (C^{-n})^\alpha$ where K and α are the Hölder constant and exponent. This is however of the form $k\beta^n$ for some $\beta < 1$, so we see g is Hölder continuous. Note that

in general $g \notin \mathcal{G}$ so that the dual of \mathcal{L}_g , although well-defined does not map probability measures to probability measures. The remainder of the proof will follow from the Ruelle-Perron-Frobenius operator theorem (see [PP]). This says that there is a \tilde{g} cohomologous to g (that is $\tilde{g}(x) = \kappa g(x)h(x)/h \circ \sigma(x)$ for a constant $\kappa > 0$ and a continuous function $h > 0$), and an equilibrium measure ν such that:

- (i) $\mathcal{L}_{\tilde{g}}1 = 1$,
- (ii) $\mathcal{L}_{\tilde{g}}^n f$ converges uniformly to $\int f d\nu$ and
- (iii) ν is ergodic and shift-invariant.

By a straightforward calculation, we see that $h \cdot \mathcal{L}_{\tilde{g}} f = \kappa \mathcal{L}_g(h \cdot f)$. Taking $f = 1$ and integrating with respect to μ , using (4), we see that $\kappa = 1$. It follows that $h \cdot \mathcal{L}_{\tilde{g}}^n f = \mathcal{L}_g^n(h \cdot f)$. Integrating with respect to μ and taking the limit as $n \rightarrow \infty$ gives $\int f d\nu = \int f \cdot h d\mu$. In particular, ν is absolutely continuous with respect to μ . By the isomorphism π , ν lifts to an ACIM for T . Since the lifted measure remains ergodic, it follows that ν is unique as claimed. This completes the proof of the proposition. \square

This illustrates how problems about expanding maps give rise to problems in g -measures. The following Lemma provides a connection in the other direction.

Lemma 3. *Given a g -measure ν on Σ_2 , with $g \in \mathcal{G}^+$, there exists a continuous surjection $\pi : \Sigma_2 \rightarrow S^1$ and a degree 2 full Lipschitz map $T : S^1 \rightarrow S^1$ such that*

- (i) T preserves Lebesgue measure λ and
- (ii) (T, λ) is measure-theoretically isomorphic under π to (σ, ν) . Suppose further that

$g \in \mathcal{G}^0$ and that g satisfies

$x \sim y \Rightarrow g(x) = g(y)$, where

$$x \sim y \text{ if } \begin{cases} x = y \\ x = a0111\dots, y = a1000\dots & \text{for some finite (possibly empty) word } a, \\ x = a1000\dots, y = a0111\dots & \text{for some finite (possibly empty) word } a, \\ x = 00000\dots, y = 11111\dots & \text{or} \\ x = 11111\dots, y = 00000\dots \end{cases} \quad (5)$$

Then the map T has the additional property that it is C^1 .

We will find it convenient to write $\mathcal{G}^{\text{comp}}$ for the set of those $g \in \mathcal{G}^0$ satisfying (5). We call these g *compatible*.

Proof. Define a total order on Σ_2 , the lexicographic ordering:

$$x < y \Leftrightarrow \exists n \geq 0 \text{ such that } x_0 = y_0, \dots, x_{n-1} = y_{n-1} \text{ and } x_n < y_n.$$

Now, set $[x, y] = \{z : x \leq z \leq y\}$ and define the open intervals analogously. We will at this point record for later use the following equation, which follows from (1). Suppose x and y lie in Σ_2 and have the same first term. Suppose also $x \leq y$. Then we have

$$\nu(x, y] = \int_{(\sigma(x), \sigma(y)]} g(x_0 z) d\nu(z). \quad (6)$$

We will regard the circle as the quotient of the interval $[0,1]$ by the relation $0=1$. Write o for the sequence in Σ_2 whose terms are all 0. Now define $\pi : \Sigma_2 \rightarrow S^1$ by $\pi(x) = \nu[o, x] \pmod{1}$. Using elementary properties of g -measures (that they are non-atomic and of full support), we have that $\pi(x) = \pi(y) \Leftrightarrow x \sim y$.

To check that π is surjective, note that π is continuous (since ν has no atoms), so that $\pi(\Sigma_2)$ is compact and hence closed. The set $\pi(\Sigma_2)$ also contains the set $\pi(\{a000\dots : a$

is a finite sequence}), which is dense in S^1 , so π is surjective. We also want to check that the metric topology on S^1 coincides with the quotient topology it inherits from the projection $\pi : \Sigma_2 \rightarrow S^1$. We have already noted that π is continuous with respect to the metric topology on S^1 . This implies that the open sets in the metric topology are open in the quotient topology. We have to check the converse. Suppose A is open in the quotient topology on S^1 , that is $\pi^{-1}(A)$ is open in Σ_2 . This implies that $\pi^{-1}(A)$ is a union of cylinders in Σ_2 . Pick $\zeta \in A$. Then $\pi^{-1}(\zeta)$ consists of a \sim -equivalence class. If this class has only one member, then since $\pi^{-1}(A)$ consists of cylinders, it must contain a cylinder which contains $\pi^{-1}(\zeta)$. It is easy to see that ζ must be contained in the interior of the image under π of this cylinder, hence $\zeta \in \text{Int}(A)$. If the class has two members, then each member must be contained in a cylinder. These cylinders will project to a left- and a right-neighbourhood of ζ , which implies, again that $\zeta \in \text{Int}(A)$. It follows that A is open in the metric topology, which shows that the two topologies coincide.

We can use this information to construct the map T . Note that if $x \sim y$ then $\sigma(x) \sim \sigma(y)$, so $\pi \circ \sigma(x) = \pi \circ \sigma(y)$. Using the universal property of quotients, this implies that there is a continuous map $T : S^1 \rightarrow S^1$ such that $T \circ \pi = \pi \circ \sigma$. Now, π is a measure-theoretic isomorphism between the pairs (σ, ν) and $(T, \pi^* \nu)$, where $\pi^* \nu(A) = \nu(\pi^{-1}(A))$. Note that $\pi^{-1}(\zeta)$ consists of at most two points. Write $\rho_+(\zeta)$ for $\max(\pi^{-1}(\zeta))$ and $\rho_-(\zeta)$ for $\min(\pi^{-1}(\zeta))$. Now, we have

$$\begin{aligned} \pi^* \nu([0, \zeta]) &= \nu(\pi^{-1}[0, \zeta]) = \nu([0, \rho_+(\zeta)]) \\ &= \pi(\rho_+(\zeta)) = \zeta = \lambda([0, \zeta]). \end{aligned}$$

It follows that $\pi^* \nu = \lambda$, so we have shown that π is a measure-theoretic isomorphism between (σ, ν) and (T, λ) . It remains to show that T is an expanding map. Note though

that $g \in \mathcal{G}^+$. This implies that $g \geq C$ for some constant $C > 0$. Since $g(0x) + g(1x) = 1$, this implies that $C \leq g \leq 1 - C$. Now pick x and y in the same 0-cylinder of Σ_2 . By (6), we see that $\nu(x, y) \leq (1 - C)\nu((\sigma(x), \sigma(y)))$, so $\nu((\sigma(x), \sigma(y))) \geq (1 - C)^{-1}\nu(x, y)$. Now given two points ζ and ξ in the same branch of T (thinking of T as a map of I), with $\zeta > \xi$, we can take x and y to be lifts of these points. Since $\nu(x, y) = \lambda(\pi(x), \pi(y))$, the above equation then implies that $|T(\zeta) - T(\xi)| \geq (1 - C)^{-1}|\zeta - \xi|$. It follows that the map T is expanding. Note we also have that $|T(\zeta) - T(\xi)| \leq C^{-1}|\zeta - \xi|$, so T is Lipschitz.

If $g \in \mathcal{G}^{\text{comp}}$, then we may once again appeal to (6), to get

$$T'(\pi(x)) = \lim_{y \rightarrow x} \frac{T(\pi(y)) - T(\pi(x))}{\pi(y) - \pi(x)}.$$

Since $T \circ \pi = \pi \circ \sigma$, this is equal to

$$\lim_{y \rightarrow x} \frac{\pi(\sigma(y)) - \pi(\sigma(x))}{\pi(y) - \pi(x)}.$$

If we now assume $y > x$ and that x and y lie in the same 0-cylinder, then the quotient is just $\nu(\sigma(x), \sigma(y))/\nu(x, y)$. By (6), this converges to $1/g(x)$ as $y \rightarrow x$ because of the continuity of g . The same analysis can be performed in the case that $y < x$. It is not hard to see that this implies $T'(\pi(x)) = 1/g(x)$. Note that the requirement (5) on g is needed to ensure that the left and right derivatives coincide at those points ζ of the form $\pi(a0111\dots) = \pi(a1000\dots)$.

We have that $1/g$ is continuous on Σ_2 and it collapses equivalence classes, so we can write $(1/g) = h \circ \pi$ for some continuous function $h : S^1 \rightarrow (1, \infty)$. The above shows that $T'(\zeta) = h(\zeta)$, which implies that T is C^1 as claimed. This completes the proof of the Lemma. \square

Note that the requirement that $g \in \mathcal{G}^+$ is stronger than we need. Most of the proof will still work if we have that the integral of g over any cylinder is positive. The only thing

which fails is that the map T will not in general be Lipschitz. Also the 2-shift used in the argument could be replaced by an l -shift for any l . We will need these facts in the next section.

The two proofs above show how we can translate between problems of expanding maps and g -measures. The general situation is that for expanding maps, one has a complicated map, but a straightforward measure (that is Lebesgue measure or some other measure absolutely continuous with respect to Lebesgue measure), while for g -measures, one has a straightforward map (the shift map), but a complicated measure.

This Lemma also gives a possible approach to Question 1. The approach would then be to construct a g satisfying (5), which has a non-ergodic g -measure, ν . The Lemma would then provide a C^1 expanding map T preserving Lebesgue measure which would have the property that λ is not ergodic for T by the isomorphism described in the Lemma.

We conclude this section by exhibiting the relationship between generalized baker's transformations and g -measures.

Lemma 4. *Suppose T is a generalized baker's transformation based on the expanding map ϕ . Then there is a continuous surjection $\pi : \Sigma_2 \rightarrow I$ and a shift-invariant measure ν such that*

- (i) *The pair (σ, ν) is measure-theoretically isomorphic to (ϕ, λ) , the vertical projection of $(T, \lambda \times \lambda)$.*
- (ii) *ν is a g -measure where $g(x) = 1/\phi'(\pi(x))$. This g is defined almost everywhere with respect to ν and we have $g \in \mathcal{G}^+$.*

Proof. We define π as in Proposition 2. The map ϕ is required to preserve Lebesgue measure (as described in the section defining generalized baker's transformations). π is

a bijection on a set of full Lebesgue measure. It follows that π is a measure-theoretic isomorphism between (ϕ, λ) and (σ, ν) for some shift-invariant measure ν . Following the proof of Proposition 2, we see that ν is a g -measure, where $g(x) = 1/\phi'(\pi(x))$. We know that ϕ is a Lipschitz map, so it follows that its derivative is defined on a set of Lebesgue measure 1. This definition therefore makes sense. By the conditions placed on ϕ , we see that g can be taken to be a member of \mathcal{G}^+ , by possibly redefining on a set of measure 0. \square

Note that we can be still more specific. If the first term of x is 0, then $\phi'(\pi(x)) = 1/f_0(x)$ almost everywhere (with respect to ν) and if the first term of x is 1, then $\phi'(\pi(x)) = 1/f_1(x)$ almost everywhere. This implies that $g(0x) = f_0(\pi(0x))$ and $g(1x) = f_1(\pi(1x))$. This means that the system (ϕ, λ) is isomorphic to (σ, ν) where ϕ is the vertical projection of a generalized baker's transformation T_f , g is just given by compositions of the f -functions with a semi-conjugacy and ν is a g -measure.

3. Construction of Examples of Expanding Maps

In this section, we use the results of the previous section to produce examples of expanding maps which preserve Lebesgue measure. We also prove some basic results about the non-existence of certain types of example. This section is in fact primarily motivated by finding an answer to Question 1. Throughout, we will be interested in full maps of the interval, which preserve Lebesgue measure. One well-known construction of invariant sets is that of 'Cookie Cutters'. These are degree 3 full maps of the interval, and one considers the Cantor set S of points whose forward orbit never enters the middle subinterval. This set satisfies $T(S) \subset S$ so if the map preserves Lebesgue measure and the set S has positive Lebesgue measure, then it follows that Lebesgue measure is not

ergodic. Bowen gives a similar construction in [Bow]. We now give a construction of such a Lebesgue measure-preserving Cookie Cutter with a positive measure Cantor set.

Consider the space Σ_3 . Let μ_1 be the Bernoulli measure on Σ_3 with probability vector $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and μ_2 be the Bernoulli measure with probability vector $(\frac{1}{2}, 0, \frac{1}{2})$. Let $\nu = \frac{1}{2}(\mu_1 + \mu_2)$. Let B be the subset of Σ_3 whose elements have no terms equal to 1. Then $\nu(B) = \frac{1}{2}$ since $\mu_1(B) = 0$ and $\mu_2(B) = 1$. We now define g by

$$g(ix) = \begin{cases} \frac{1}{3} & \text{if } x \notin B, \\ \frac{1}{2} & \text{if } x \in B \text{ and } i=0 \text{ or } 2, \\ 0 & \text{if } x \in B \text{ and } i=1. \end{cases}$$

Then $g \in \mathcal{G}$ and ν is a g -measure. Note that $g \notin \mathcal{G}^+$, but as indicated in the note following the proof of Lemma 3, we still get a Lebesgue measure-preserving degree 3 full map T of the interval such that (T, λ) is measure-theoretically isomorphic to (σ, ν) , because the ν -measure of any cylinder is positive. This map has the property that the set of those points which never enter the middle interval has Lebesgue measure $\frac{1}{2}$. The graph of this map is shown in Figure 2.3. The map however is certainly not C^1 or even Lipschitz since g is not continuous. This is so for a good reason.

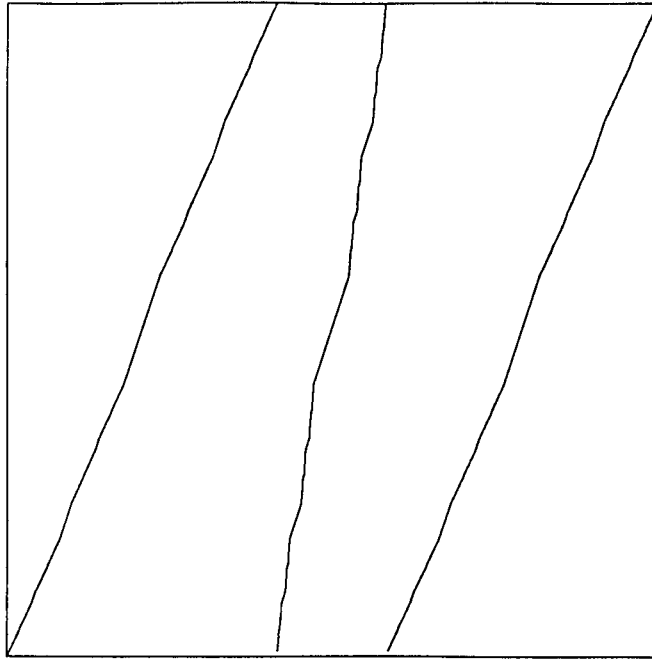


Figure 2.3 Graph of T .

Lemma 5. *Suppose T is a Lipschitz full map of the interval, preserving Lebesgue measure λ . Suppose further that C is a closed set such that $T(C) \subset C$. Then either $C = I$ or $\lambda(C) = 0$.*

Proof. Note that full maps of the interval may also be considered as expanding maps of the circle. This implies that, if J is any open interval, then $T^n(J) = I$ for some $n > 0$. Now suppose that C is a closed set, such that $C \neq I$ and $T(C) \subset C$. Let U be any open interval in $I \setminus C$. There is then an $n > 0$ such that $T^n(U) \supset C$.

Let $A = U \cap T^{-n}(C)$. It follows that $T^n(A) = C$. We have however that T is Lipschitz so there exists a constant K such that $T'(x) \leq K$, for almost all $x \in I$ (with respect to λ). It follows that $\lambda(T(A)) \leq K\lambda(A)$. In particular, $\lambda(C) \leq K^n\lambda(A)$. Note however that points of A are non-recurrent: Any point of A will never return to A after any time beyond n . It follows by the Poincaré recurrence theorem that the measure of A is 0 with respect

to any invariant measure, so we have $\lambda(A) = 0$ and hence $\lambda(C) = 0$ as claimed. \square

This proof in fact shows something stronger. Namely, if A is an invariant set of a Lipschitz map T preserving Lebesgue measure and if $\lambda(A) > 0$, then A is dense in I . We have already noted that if an invariant set contains an interval then it is all of I , so we have that if a Lipschitz map preserving λ is non-ergodic, then it has an invariant set of measure different from 0 and 1. Such a set would have to contain no intervals and would also have to be dense in I by the above.

In fact, we can use the same idea as the previous example to construct a Lipschitz map. For this example, work on Σ_2 . Let μ_1 be the Bernoulli measure with weight vector $(\frac{1}{2}, \frac{1}{2})$ and let μ_2 be any other Bernoulli measure. Suppose the weight vector is $(\alpha, 1 - \alpha)$. Then set $\nu = \frac{1}{2}(\mu_1 + \mu_2)$. Set $B = \{x \in \Sigma_2 : \lim_{n \rightarrow \infty} 1/n \sum_{i=0}^{n-1} x_i = 1/2\}$. Then $\mu_1(B) = 1$ and $\mu_2(B) = 0$ by the ergodic theorem (or the Strong Law of Large Numbers). Define g by

$$g(ix) = \begin{cases} \frac{1}{2} & \text{if } x \in B \\ \alpha & \text{if } i = 0 \text{ and } x \notin B \\ 1 - \alpha & \text{if } i = 1 \text{ and } x \notin B \end{cases}$$

Then $g \in \mathcal{G}^+$ and ν is a non-ergodic g -measure (since $\sigma(B) \subset B$ and $\nu(B) = \frac{1}{2}$). It follows, by Lemma 3, that there is a Lipschitz map T preserving Lebesgue measure, for which Lebesgue measure is non-ergodic. This answers Question 1 in the affirmative in the class of Lipschitz maps. This example has already been pointed out by Bose in [Bos1], although his proof relied on constructing a generalized baker's transformation.

4. Some Possible Approaches to Question 1

In the remainder of this chapter, I discuss two possible ways of attacking Question 1. The first is to construct C^1 maps with unbounded distortion. The idea behind this is that many of the proofs of uniqueness of ACIM rely on a property known as bounded distortion (see [Ma]). An example of a C^1 map preserving Lebesgue measure with unbounded distortion is therefore a possible candidate for a C^1 map with two ACIMs. This construction is based on the paper [GS]. The second approach, which seems more hopeful is to modify the proof of [BK] to make the g which is constructed satisfy (5). The proofs required to make this method work are likely to be even more difficult than the already technically advanced methods used in [BK].

Suppose T is a map of the interval. T is said to have *bounded distortion* if there exists a constant $C > 0$ such that when K_1 and K_2 are subintervals of an interval K , for which the restriction of T^n to K is a homeomorphism then

$$1/C < \frac{|T^n(K_1)|}{|T^n(K_2)|} / \frac{|K_1|}{|K_2|} < C.$$

It is shown in [Ma] that if T is piecewise $C^{1+\alpha}$, then T has bounded distortion, so it is of interest to find an example of a piecewise C^1 map which has unbounded distortion. Of course, it will follow that the derivative of such a map will not be Hölder continuous. We will construct a degree 3 full map of the interval with this property. Write T_0 , T_1 and T_2 for the three branches of the map. These are defined on the intervals $I_0 = [0, \frac{1}{3}]$, $I_1 = [\frac{1}{3}, \frac{2}{3}]$ and $I_2 = [\frac{2}{3}, 1]$. For each non-empty string s of 0s and 2s, define J_s to be an interval of length $3^{-(1+|s|)}$, where $|s|$ denotes the length of the string s . Specifically, J_s is the subinterval of $[0,1]$ which consists of those numbers where the first $|s|$ digits of their ternary expansions are given by s and whose $(|s| + 1)$ st digits are 1s. These intervals

are arranged as in Figure 2.4.

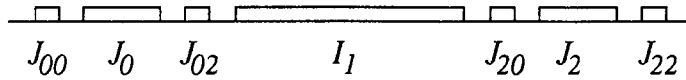


Figure 2.4 Configuration of J intervals.

Let J denote I_1 (that is the J interval indexed by the empty string). Then define a map τ acting on finite strings by truncation on the left: $\tau(a_0 a_1 \dots a_n) = (a_1 \dots a_n)$. Then define a map T on the intervals J_s as follows: For each non-empty string s , T is a diffeomorphism of J_s onto $J_{\tau(s)}$. The restriction of T to those J_s , for s starting with a 0 'bulge upwards', while the restrictions to J_s for s starting with a 2 'bulge downwards'. This is illustrated in Figure 2.5.

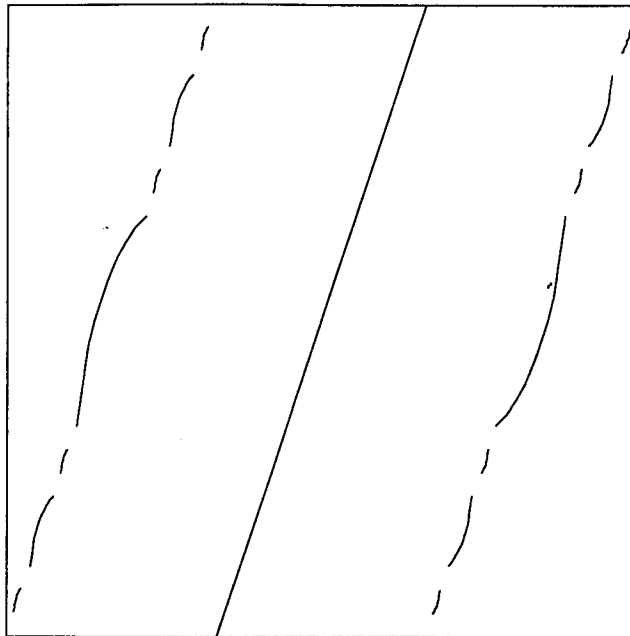


Figure 2.5 Construction of the map T .

Write T_s for the restriction of T to J_s . We will require the T_s to have derivative equal to 3 at the endpoints. The derivative must also be bounded above $\frac{3}{2}$, in order that T can

preserve Lebesgue measure. In order that T be differentiable, we will require that there exists a sequence a_n tending to 0 such that if $|s| = n$ then $3 - a_n \leq T'_s(x) \leq 3 + a_n$ for all $x \in J_s$. If one writes down the left-hand branches (that is the T_s for s starting with a 0), then the right-hand branches will be determined by the requirement that the map preserves Lebesgue measure. In order to have unbounded distortion, we will need to have $\sum a_n = \infty$.

We now construct the map T . Since the union of the intervals J_s is dense in I , it is clearly sufficient to specify the map on those intervals. Next note that if we define $T_0 : x \mapsto 3x \pmod{1}$, then T_0 maps J_s homeomorphically onto $J_{\tau(s)}$. To prove that T is differentiable, we will show that T is the limit in the C^1 topology of a sequence of differentiable maps T_n . To get T_n from T_{n-1} we modify the map T_{n-1} on the intervals J_s for which $|s| = n$. As described above, we will require the restrictions T_s to these intervals to have derivative 3 at their endpoints, and to have derivative bounded throughout the interval between $3 - a_n$ and $3 + a_n$ for some positive sequence a_n decreasing to 0. This will imply that the uniform distance between T_m and T_n is bounded above by $3^{-(m+1)}a_m$ for $m > n$. The uniform distance between their derivatives is bounded above by a_m . It follows that the sequence is Cauchy in the C^1 topology and hence converges to a differentiable map T as required.

Write α_s for the affine orientation-preserving map sending $[0, 1]$ onto J_s and write S_s for the composition $\alpha_{\tau(s)}^{-1} \circ T_s \circ \alpha_s$. The map S_s is a rescaled copy of T_s . We can therefore specify T_s by describing S_s . The requirements we have placed on T_s are equivalent to the requirements that S_s has derivative equal to 1 at its endpoints and that its derivative is bounded between $1 - \frac{a_n}{3}$ and $1 + \frac{a_n}{3}$. Now fix a sequence a_n : let $a_n = 1/n$. Now for $|s| = n$

and s starting with a 0, define S_s by

$$S_s(x) = x + \frac{a_n}{72}x^2(1-x)^2.$$

In particular, $S_s(x) > x$. We have for such an s , that $3 - \frac{a_n}{4} \leq T'_s \leq 3 + \frac{a_n}{4}$. For s starting with a 0, T_s therefore satisfies the conditions bounding the derivative. Also, we see that the derivative of S_s is 1 at the endpoints as required. We then define T_s for those s starting with a 2 by the requirement that the map as a whole must preserve Lebesgue measure. By the change of variables formula, this can be seen to be equivalent to the requirement that

$$\sum_{y \in T^{-1}(x)} \frac{1}{T'(y)} = 1,$$

for all x . Applying this to a point x of J_s , we get the requirement that

$$\frac{1}{3} + \frac{1}{T'_{0s}(y_0)} + \frac{1}{T'_{2s}(y_2)} = 1, \tag{7}$$

where y_0 and y_2 are the preimages of x in J_{0s} and J_{2s} respectively. Since we know that T'_{0s} is bounded within $a_n/4$ of 3, one can check that if we impose condition (7), in order that T preserves Lebesgue measure, then T'_{2s} is bounded within a_n of 3. Note that this relies on the fact that $a_n \leq 1$. It also follows from (7) that the derivative of T_{2s} is 3 at the endpoints as required. This completes the definition of T .

Note that the derivative of T is not of bounded variation. The variation is given by

$$\sum_s 2(\max(T'_s) - \min(T'_s)).$$

This is at least as big as $\sum_n \frac{1}{72}2^{n-1}a_n$ since the derivative of T_s is 3 at the endpoints and is at least as large as $3 + \frac{1}{144}a_n$ at some point of the interval J_s , providing that s starts with a 0. In particular, this sum is divergent.

We now show that T has unbounded distortion as claimed. We will take for the interval K , an interval of the form $J_{000\dots 0}$, set $K_2 = K$ and take K_1 to be an initial segment of K . In particular, write $J^{(m)}$ for the interval J_s where s is the string consisting of m 0s. Write $\alpha^{(m)}$ for the corresponding affine map and $T^{(m)}$ and $S^{(m)}$ for the T and S maps corresponding to $J^{(m)}$. Distortions are unaffected by composing T^n before or after with affine maps. To consider the distortion, it is therefore sufficient to work with the maps $S^{(m)}$, which are clearly also much more convenient. Now fix $\epsilon > 0$. In this picture, take $J = [0, 1]$, $K_2 = J$ and $K_1 = [0, \epsilon]$. We are then interested in

$$\frac{|S^{(1)} \circ \dots \circ S^{(n)}(K_1)|}{|S^{(1)} \circ \dots \circ S^{(n)}(K_2)|} \bigg/ \frac{|K_1|}{|K_2|}.$$

Since $S_s(K) = K$ for all s , this is equal to $S^{(1)} \circ \dots \circ S^{(n)}(\epsilon)/\epsilon$. Let $c_n = S^{(1)} \circ \dots \circ S^{(n)}(\epsilon)$.

The claim is that $c_n \rightarrow 1$ as $n \rightarrow \infty$.

Certainly, c_n is increasing, since $S^{(m)}(x) > x$ for all $x \in (0, 1)$ and $S^{(m)}$ is an increasing function, so we must have that c_n increases to some $\eta \in (\epsilon, 1]$. Suppose that $\eta \neq 1$. Write $f(x) = \frac{1}{72}x^2(1-x)^2$. Then $f(\eta) > 0$. By continuity of f , there exists a $k > 0$ such that $f \geq k$ on $[\epsilon, \eta]$. Now $S^{(m)}(x) = x + a_m f(x)$. It follows that $S^{(1)} \circ \dots \circ S^{(n)}(\epsilon) \geq \epsilon + k \sum_{i=1}^n a_i$. However this tends to infinity as $n \rightarrow \infty$, thus furnishing us with a contradiction. This proves that $c_n \rightarrow 1$ as $n \rightarrow \infty$. In particular, the distortion is at least as big as $1/\epsilon$ for all ϵ , so is infinite.

The second approach consists of modifying the g -function constructed in [BK] so that it satisfies (5). To describe the proof in [BK], we will need to introduce some further concepts. Bramson and Kalikow deal exclusively with the space Σ_2 , and construct a g -function there. If $x \in \Sigma_2$, write \bar{x} for the sequence obtained by reversing each term of x (that is $\bar{x}_i = 1 - x_i$, for all i). Say g is *symmetric* if $g(\bar{x}) = g(x)$ for all x . This says

that the whole system is symmetric under the involution $x \mapsto \bar{x}$. In particular, given a g -measure, there is a conjugate g -measure: $\bar{\mu}(A) = \mu(\bar{A})$, where $\bar{A} = \{\bar{x} : x \in A\}$. In [BK], Bramson and Kalikow introduce a $g \in \mathcal{G}^0$ which has the property that there is a g -measure ν for which $\nu([1]) > \frac{1}{2}$, where $[1]$ is the cylinder set of those sequences in Σ_2 which start with a 1. The conjugate g -measure $\bar{\nu}$ then has $\bar{\nu}([1]) < \frac{1}{2}$, which implies that there is more than one g -measure. There is therefore a non-ergodic g -measure. The situation of having more than one g -measure is known in statistical mechanics as phase transition, and the existence of non-symmetric solutions to symmetric equations (that is the non-symmetric ν being the solution to the symmetric equation $\mathcal{L}_g^* \mu = \mu$) is known as spontaneous symmetry breaking. If one could find a non-ergodic g -measure with $g \in \mathcal{G}^{\text{comp}}$, then by Lemma 3, one would have an answer to Question 1.

As already mentioned, the proof in [BK] relies heavily on the fact that the g which they construct is attractive. To mimic that proof, one would like to find a g which is continuous, compatible, attractive and symmetric. Unfortunately, we can show that such a g has a unique g -measure.

Lemma 6. *Suppose g is attractive, continuous and compatible. Then there is a unique g -measure.*

Proof. To show this, we will show that g is increasing with respect to the lexicographic ordering (\leq) on Σ_2 . Suppose that x and y are in Σ_2 and satisfy $x < y$. Then let n be such that $x_i = y_i$, for all $i < n$, but $x_n = 0$ and $y_n = 1$. Write a for the finite word $x_0 x_1 x_2 \dots x_{n-1}$. Then we have $x \preceq a0111\dots$ and $a1000\dots \preceq y$, so $g(x) \leq g(a0111\dots) = g(a1000\dots) \leq g(y)$. Now suppose ν is a g -measure. By Lemma 3, there exists a C^1 full map T which preserves Lebesgue measure such that the pairs (T, λ) and (σ, ν) are measure-

theoretically isomorphic. Further, we have that $1/T'(x) = g(\pi(x))$, from which it follows that $1/T'$ is monotonically increasing and hence of bounded variation. It follows by the result of Wong ([Wo]) that such maps preserve exactly one ACIM, so Lebesgue measure is ergodic for T and hence ν is ergodic, proving that ν is the unique g -measure. \square

One attempt to get around this introduces a different notion of compatibility. We define a second equivalence relation \approx on Σ_2 .

$$x \approx y \text{ if } \begin{cases} x = y; \\ x = a01000\dots \text{ and } y = a11000\dots \text{ where } a \text{ is a finite word, or} \\ x = a11000\dots \text{ and } y = a01000\dots \text{ where } a \text{ is a finite word.} \end{cases}$$

Note that by finite words, we are allowing the possibility that they are empty.

Lemma 7. *Suppose $h \in \mathcal{G}^0$ has the property that there is a non-ergodic h -measure and*

$$x \approx y \Rightarrow h(x) = h(y) \tag{8}$$

Then there is a $g \in \mathcal{G}^{\text{comp}}$, such that there is a non-ergodic g -measure.

Proof. Define the 2-1 map $P : \Sigma_2 \rightarrow \Sigma_2$ by $P(x)_n = x_n + x_{n+1} \pmod{2}$. This map is certainly continuous. It has two inverse branches τ_0 and τ_1 given by $\tau_0(x)_n = x_0 + x_1 + \dots + x_{n-1} \pmod{2}$ and $\tau_1(x)_n = 1 + x_0 + x_1 + \dots + x_{n-1} \pmod{2}$. Note that $(\tau_i(x))_0 = i$. Let \mathcal{M} denote the probability measures on Σ_2 and define the map $\tau^* : \mathcal{M} \rightarrow \mathcal{M}$ by $\tau^*\mu(A) = \frac{1}{2}\mu(\tau_0^{-1}A) + \frac{1}{2}\mu(\tau_1^{-1}A)$. This is equal to $\frac{1}{2}\mu(P(A \cap [0])) + \frac{1}{2}\mu(P(A \cap [1]))$. We will use (1) to show that if μ is an h -measure, then $\tau^*\mu$ is an $h \circ P$ -measure. We have for $n \geq 1$,

$$\begin{aligned} \frac{\tau^*\mu([x]^n)}{\tau^*\mu([\sigma(x)]^{n-1})} &= \frac{\frac{1}{2}\mu(P([x]^n \cap [0])) + \frac{1}{2}\mu(P([x]^n \cap [1]))}{\frac{1}{2}\mu(P([\sigma(x)]^{n-1} \cap [0])) + \frac{1}{2}\mu(P([\sigma(x)]^{n-1} \cap [1]))} \\ &= \frac{\mu(P([x]^n))}{\mu(P([\sigma(x)]^{n-1})}. \end{aligned}$$

Since P and σ commute and $P([x]^n) = [P(x)]^{n-1}$, this is equal to $\mu([P(x)]^{n-1})/\mu([\sigma(Px)]^{n-2})$. Since μ is an h -measure, we see that

$$\lim_{n \rightarrow \infty} \frac{\tau^* \mu([x]^n)}{\tau^* \mu([\sigma(x)]^{n-1})} = h \circ P(x).$$

It follows that $\tau^* \mu$ is a g -measure, where $g = h \circ P$ as claimed. However, we see that $x \sim y$ implies that $P(x) \approx P(y)$, so by the conditions on h , we have $g(x) = g(y)$. This means that g satisfies (5). It remains to check that if μ is non-ergodic, then $\tau^* \mu$ is non-ergodic. Suppose then that μ is non-ergodic. There exists a Borel set B such that $\sigma^{-1}B = B$, with $\mu(B)$ different from 0 and 1. Since P and σ commute, it follows that $\sigma^{-1}(P^{-1}B) = P^{-1}B$. Now, we have

$$\begin{aligned} \tau^* \mu(P^{-1}B) &= \frac{1}{2} \mu(P(B_0 \cap P^{-1}B)) + \frac{1}{2} \mu(P(B_1 \cap P^{-1}B)) \\ &= \frac{1}{2} \mu(B) + \frac{1}{2} \mu(B) = \mu(B), \end{aligned}$$

so $\tau^* \mu$ has a shift-invariant set of measure distinct from 0 and 1. It follows that $\tau^* \mu$ is also non-ergodic as required. \square

This proof works by finding a recoding of Σ_2 to a (hopefully) more useful form. At first sight, the equivalence relation (8) seems as if it might be more easy to satisfy, whilst maintaining attractiveness, than (5). This implies that to answer Question 1, it would be sufficient to exhibit an $h \in \mathcal{G}^0$ satisfying (8) such that there is a non-ergodic h -measure. Unfortunately this equivalence relation is even worse than (5) in its interaction with the property of attractiveness.

Lemma 8. *Suppose $h \in \mathcal{G}^0$ is attractive and satisfies (8). Then $h \equiv \frac{1}{2}$.*

Proof. Suppose h is as in the statement of the Lemma. Then we will show that if a is any finite word (possibly empty), then $h(a1000\dots) = h(a0000\dots)$.

We have for any finite word a that $a10000\dots \preceq a11000\dots \approx a01000\dots$. Repeating this, we have $a10000\dots \preceq a01000\dots \preceq a00100\dots$ etc. Since h is attractive, we have $h(a10000\dots) \leq h(a01000\dots) \leq h(a00100\dots)\dots$. But since h is continuous, if we write $z^{(n)}$ for the word consisting of n 0s, it follows that $h(az^{(n)}1000\dots) \rightarrow h(a0000\dots)$ as $n \rightarrow \infty$. It follows that $h(a1000\dots) \leq h(a0000\dots)$. But we also have $a0000\dots \preceq a1000\dots$ so $h(a0000\dots) \leq h(a1000\dots)$. Taking these together, we see that $h(a0000\dots) = h(a1000\dots)$, as we wanted.

Now define a function C on the finite words by $C(a) = h(a0000\dots)$. Define a map ϕ on the finite words which truncates them on the right, so for example $\phi(101000) = 10100$. If a ends in a 0, then $C(a) = C(\phi(a))$. But if a ends in a 1 then by the previous argument $C(a) = C(\phi(a)0)$ (that is we can replace the last 1 by a 0), but $C(\phi(a)0) = C(\phi(a))$ so for all a , we have $C(a) = C(\phi(a))$. It now follows that C is independent of a . We have however that the elements $a000\dots$ are dense in Σ_2 , so using continuity, we see that h is constant, and therefore equal to $\frac{1}{2}$ everywhere. \square

This approach is based on an attempt to find a shift-commuting recoding of Σ_2 to one where it is easy to copy the argument of [BK]. I conjecture that whatever recoding one uses, one will always find problems of incompatibility with attractiveness. As a result of this, one is led to attempt to find a proof which does not rely upon the property of attractiveness.

In order to do this, it is important to be clear about the rôle played by attractiveness in [BK]. As described above, attractiveness allows one to get hold of g -measures as weak*-limits of $\mathcal{L}_g^{*n}\delta_i$ where δ_i is the measure concentrated on the point of Σ_2 , all of whose terms are i s. A second aspect of this property is used in ‘coupling’ arguments. Given two

measures μ_1 and μ_2 on Σ_2 , a *coupling* is a measure μ on the product $\Sigma_2 \times \Sigma_2$ which projects under the coordinate projections to μ_1 and μ_2 respectively (that is $\mu(A \times \Sigma_2) = \mu_1(A)$ and $\mu(\Sigma_2 \times A) = \mu_2(A)$). Note that the product of the measures μ_1 and μ_2 is always a coupling, but we are in general interested in more complicated couplings.

The following Lemma is implicitly used at several important points in the proof of [BK].

Lemma 9. *Suppose $g \in \mathcal{G}^0$ is attractive, $h \in \mathcal{G}^0$ and $h(1x) \geq g(1x)$ for all $x \in \Sigma_2$. Let w be a point of Σ_2 . Then we consider random variables $(X_n)_{n \in \mathbb{Z}}$ and $(Y_n)_{n \in \mathbb{Z}}$ taking values in $\{0, 1\}$, which have $X_n = Y_n = w_{-(n+1)}$ for $n \leq -1$. Suppose the evolution of (X_n) is governed by g for $n \geq 0$ as in (3), while that of (Y_n) is governed by h for $n \geq 0$. Then there is a coupling of X_n and Y_n such that $Y_n \geq X_n$ with probability 1.*

Proof. We have $2^{\mathbb{Z}} \times 2^{\mathbb{Z}} \simeq 4^{\mathbb{Z}}$ by the shift-commuting homeomorphism $\Theta : 2^{\mathbb{Z}} \times 2^{\mathbb{Z}} \rightarrow 4^{\mathbb{Z}}$, $(x, y) \mapsto z$, where $z_i = 2y_i + x_i$. Write π_1 and π_2 for the two coordinate projections $4^{\mathbb{Z}} \rightarrow 2^{\mathbb{Z}}$, and p_1 and p_2 for their truncations $4^{\mathbb{Z}^+} \rightarrow 2^{\mathbb{Z}^+}$. Consider the function f defined on $\{0, 2, 3\}^{\mathbb{Z}^+}$ by

$$f(3z) = g(p_1(z))$$

$$f(2z) = h(p_2(z)) - g(p_1(z))$$

$$f(0z) = 1 - h(p_2(z)).$$

Note that on $\{0, 2, 3\}^{\mathbb{Z}^+}$, $p_2(z) \succeq p_1(z)$, so $h(p_2(z)) \geq g(p_2(z)) \geq g(p_1(z))$ so the above function f is non-negative, and lives in \mathcal{G} . Now consider random variables given by

$$Z_n = \begin{cases} 0 & \text{if } w_{-(n+1)} = 0 \\ 3 & \text{if } w_{-(n+1)} = 1, \end{cases}$$

for $n \leq -1$ and evolving under f for $n \geq 0$. Then certainly $\pi_2(Z) \succeq \pi_1(Z)$ with probability

1. It remains to show that the random variables $\pi_1(Z)$ and $\pi_2(Z)$ are evolving under g and h respectively.

The probability that $\pi_1(Z)_n$ is 1 given the values of $(\pi_1(Z)_m)_{m < n}$ is the same as the probability that Z_n is 3 given the values of $(\pi_1(Z)_m)_{m < n}$, but this is just $g((\pi_1(Z)_m)_{m < n})$ as required. A similar argument works for $\pi_2(Z)$ proving the Lemma. \square

The importance of this Lemma is that it gives a way of relating g - and h -measures for attractive g -functions.

In the remainder of this chapter, we present a class of examples of g -functions, based on those in [BK], which are in \mathcal{G}^0 , are symmetric and have some compatibility properties. To describe these, we will need a third equivalence relation, which is essentially the same as (8), but which contains the symmetry and requires that the finite words a be non-empty.

$$x \asymp y \text{ if } \begin{cases} x = y \\ x = a11000\dots, y = a01000\dots & \text{for some finite non-empty word } a, \\ x = a01000\dots, y = a11000\dots & \text{for some finite non-empty word } a, \\ x = a10111\dots, y = a00111\dots & \text{for some finite non-empty word } a, \text{ or} \\ x = a00111\dots, y = a10111\dots & \text{for some finite non-empty word } a. \end{cases} \quad (9)$$

Note that if we find a g satisfying $x \asymp y \rightarrow g(x) = g(y)$ and use Lemmas 7 and 3 to get a map T preserving Lebesgue measure, then the derivative of T will have discontinuities at $\pi(001111\dots)$ and $\pi(101111\dots)$, where π is the semi-conjugacy described in Lemma 3. (These are the cases we miss out by assuming a is non-empty). If we consider T^2 , though, these points will be endpoints of intervals, so T^2 will be a degree 4 C^1 full map. If λ is not ergodic for T , then λ is not ergodic for T^2 .

The construction is in several steps. First, we show that Σ_2 / \asymp is metrizable and that the metric κ can be chosen in such a way that the map $(x, y) \mapsto \kappa(x, y)$ is Hölder continuous. Note that a metric on Σ_2 / \asymp is the same as a pseudometric \tilde{d} on Σ_2 such that

$\tilde{d}(x, y) = 0 \Leftrightarrow x \asymp y$. Such a metric can be constructed by defining $c(x, y) = \beta^n$ where the n th is the first place where x and y differ, for $\beta < 1$ sufficiently large, and letting

$$\tilde{d}(x, y) = \inf_{x=\tilde{x}_0, x_1 \asymp \tilde{x}_1, x_2 \asymp \tilde{x}_2, \dots, x_{n-1} \asymp \tilde{x}_{n-1}, x_n=y} c(\tilde{x}_0, x_1) + c(\tilde{x}_1, x_2) + \dots + c(\tilde{x}_{n-1}, x_n).$$

This \tilde{d} can then be checked by a lengthy calculation to be a pseudometric with the required properties. Note that we may easily assume that \tilde{d} is symmetric (that is $\tilde{d}(\tilde{x}, \tilde{y}) = \tilde{d}(x, y)$).

Next, let $C^{(2n)} = \{x \in \Sigma_2 : x_1, x_2, \dots, x_{2n} \text{ have more than } n \text{ } x_0\text{s}\}$ and define

$$W^{(n)}(ix) = \frac{\epsilon \tilde{d}(\tilde{ix}, C^{(2n)}) + (1 - \epsilon) \tilde{d}(ix, C^{(2n)})}{\tilde{d}(\tilde{ix}, C^{(2n)}) + \tilde{d}(ix, C^{(2n)})}$$

This function $W^{(n)}$ is symmetric (that is $W^{(n)}(x) = W^{(n)}(\bar{x})$) because $C^{(2n)} = \overline{C^{(2n)}}$. It satisfies $W^{(n)}(0x) + W^{(n)}(1x) = 1$. The functions $W^{(n)}$ are also Hölder continuous. The desired function g is then constructed by taking a very rapidly increasing sequence n_i , taking $p_i = \frac{1}{2}(\frac{2}{3})^i$ for $i \geq 1$ so that $\sum_i p_i = 1$ and letting $g = \sum_i p_i W^{(n_i)}$. This g is symmetric, continuous and of a similar construction to that in [BK]. It also satisfies $x \asymp y \Rightarrow g(x) = g(y)$. In particular, if it is truncated by replacing all the W functions after the k th by $\frac{1}{2}$, then it becomes Hölder. I conjecture that for a careful choice of n_i , the above example has a non-ergodic g -measure, and hence, as described above gives rise to a full map preserving Lebesgue measure, but for which Lebesgue measure is non-ergodic.

Chapter 3. Representation of Markov Chains on Manifolds

1. Introduction

In this chapter, we consider the problem of representation of Markov chains. By a Markov chain, one usually means a process described by a sequence of random variables $(X_n)_{n \in \mathbb{N}}$, where each random variable takes values in some finite or countable set S and the sequence satisfies the Markov property:

$$\mathbb{P}(X_n = s \mid X_{n-k} = s_{n-k}, \dots, X_{n-1} = s_{n-1}) = \mathbb{P}(X_n = s \mid X_{n-1} = s_{n-1})$$

where $k \in \mathbb{N}$, $s \in S$ and $s_{n-i} \in S$ for $1 \leq i \leq k$.

Here, we will take a generalized definition. A *Markov chain* will be a random process described by a sequence of random variables $(X_n)_{n \in \mathbb{N}}$, each taking values in a measure space M with σ -algebra \mathcal{B} . This time, the above statement of the Markov property is insufficient as we will be considering spaces M which are not countable. We will require instead

$$\mathbb{P}(X_n \in A \mid X_{n-k} \in A_{n-k}, \dots, X_{n-1} \in A_{n-1}) = \mathbb{P}(X_n \in A \mid X_{n-1} \in A_{n-1})$$

where $k \in \mathbb{N}$, $A \in \mathcal{B}$, $A_{n-i} \in \mathcal{B}$ for $1 \leq i \leq k$ and $\mathbb{P}(C|D)$ is the probability of C given D .

Such a process may be described by a map $P : M \times \mathcal{B} \rightarrow [0, 1]$ such that for fixed x , the map $P_x : A \mapsto P(x, A)$ is a probability measure on (M, \mathcal{B}) , and for fixed $A \in \mathcal{B}$, the map $x \mapsto P(x, A)$ is measurable. This latter condition is required for the above statement of the Markov property to make sense. The map P is called the *transition map* of the

Markov chain \mathcal{M} . The quantity $P(x, A)$ is to be thought of as the probability of moving from the point x into the set A .

The idea behind representation of Markov chains was first introduced in [Kak]. There, he considers a sequence of independent identically distributed random variables $(Y_n)_{n \in \mathbb{N}}$ (with the distribution being given by a probability measure m) taking values in a collection \mathcal{F} of maps from a space M to itself. He shows how a Markov chain \mathcal{M} is induced by defining

$$P(x, A) = m(\{f : f(x) \in A\}), \quad \forall x \in M, A \in \mathcal{B}. \quad (1)$$

A *measured collection* of maps from M to itself is a collection \mathcal{F} of maps from M to itself and a probability measure m on a compatible σ -algebra on \mathcal{F} , where a σ -algebra is *compatible* if $\{f : f(x) \in A\}$ is a measurable subset of \mathcal{F} for any measurable subset A of M and any point x in M . We ask for conditions on a Markov chain that it can be induced by a measured collection of maps as in (1). Given a Markov chain \mathcal{M} on M , we will say a collection \mathcal{F} of maps from M to itself and a probability measure m on \mathcal{F} is a *representation* of \mathcal{M} if the transition map of the Markov chain is induced by the collection \mathcal{F} and the probability measure m as in (1). The rest of this chapter asks under what conditions on a Markov chain \mathcal{M} may we find a representation of \mathcal{M} for which the \mathcal{F} lies in a specific collection of maps. In this case, we will say that \mathcal{M} may be represented by maps of this type. This allows us to consider Markov chains as examples of Random Dynamical Systems (RDS), (see [Ki] and [AC]).

We consider Markov chains for which the space M is a smooth manifold and \mathcal{B} is the σ -algebra of Borel sets. In §3–§6, we find conditions under which \mathcal{M} may be represented by smooth maps. This answers a question in [Ki] and has appeared in the literature ([Q1]).

In §7, we present an aside showing that all the Markov chains in question appearing

in §3–§6 have a unique invariant probability distribution, and all other probability distributions converge to this one exponentially fast. This material turns out to be well-known, but I did it independently, and even my proofs are very similar to those which appear in the literature (see for example [LM]).

The material contained in this chapter is all original except §2, §3 and those parts of §5 which are explicitly credited to other authors.

2. Previously Known Results

By way of an introduction, we first consider the case of finite state Markov chains. In this case, some of the questions take on particularly simple forms relating to the properties of matrices. An n -state Markov chain may be described by an $n \times n$ matrix A , where A_{ij} is the probability of going from state i to state j . In the more general transition map notation, this would have been written $P(i, \{j\}) = A_{ij}$. Clearly, we require that for each i , $\sum_{j=1}^n A_{ij} = 1$, and $A_{ij} \geq 0$. In this case, we call the matrix A *stochastic*. In this set-up, the version of the question about representation, with which this chapter concerns itself is: Does there exist a collection C of maps from $S = \{1, 2, \dots, n\}$ to itself and a probability measure m on them such that A_{ij} is equal to $m\{\phi \in C : \phi(i) = j\}$?

Note that in this case, there are finitely many maps from S to itself. Each map may be described by a matrix of 0s and 1s such that each row contains exactly one 1. Such matrices will be called *basic*. Such a matrix B corresponds to a map ϕ by $B_{ij} = 1 \Rightarrow \phi(i) = j$. Since there are finitely many possible maps, a probability measure on them is just a collection of weights, one for each possible map, which add up to 1. Suppose the maps are denoted by $(\phi^{(k)})_{1 \leq k \leq N}$ and correspond to matrices $(B^{(k)})_{1 \leq k \leq N}$. These could then have weights $w^{(k)}$. The condition for the collection of weights and maps to be a representation of the

Markov chain described by the matrix A_{ij} is that

$$A_{ij} = \sum_{\{k:\phi^{(k)}(i)=j\}} w^{(k)} = \sum_{\{k:B_{ij}^{(k)}=1\}} w^{(k)} = \sum_{k=1}^N w^{(k)} B_{ij}^{(k)}.$$

This is the same as $A = \sum w^{(k)} B^{(k)}$, and so to show any finite state Markov chain has a representation by maps, it is simply necessary to show that any stochastic matrix A is a convex linear combination of basic matrices. This fact admits a simple proof by induction on the number of zeros in the matrix. Before we start on the proof, define a matrix A to be β -stochastic if $A_{ij} \geq 0$ and $\sum_j A_{ij} = \beta$ for all i .

The inductive step is then as follows: Suppose A is β -stochastic and has r zero entries. Then if $\beta = 0$, we are done. Otherwise, each row contains some non-zero entry. Define ϕ to be a map taking a row to the number of a column with a non-zero entry in that row (that is $A_{i,\phi(i)} > 0$ for all i). Then let $\alpha = \min_i A_{i,\phi(i)}$ and B be the basic matrix corresponding to ϕ . Then $A - \alpha B$ is $\beta - \alpha$ -stochastic and has more than r zero entries.

Since the matrices are finite, this procedure will terminate (in less than n^2 steps), at which the remainder will become 0. Since the matrices subtracted are basic, it follows that the original matrix A was a convex linear combination of basic matrices as required.

Since one is interested in finding 'good' representations of Markov chains, it is natural to ask whether it is possible to find a representation by permutations. These maps correspond to permutation matrices (that is there is exactly one 1 in each row and each column and all other entries are 0), all of which are *bistochastic* (the columns and rows all sum to 1). Any convex linear combination of such matrices will clearly also be bistochastic, so a necessary condition for a finite state Markov chain to have a representation by permutations is that its transition matrix is bistochastic. In fact, this turns out to be sufficient.

This result is originally due to Birkhoff ([Bi]), and was later reproved by Hammersley and Mauldon ([HM]). The proof I include here is based on a sketch of Birkhoff's proof which was included in the book of Bollobàs, [Bol].

The proof relies on the Hall 'Marriage Theorem' (see [Bol]) which may be stated as follows. Let S and T be two finite sets of equal cardinality and suppose there is a relation $\mathcal{R} \subset S \times T$ between members of S and T . If M is a subset of S , write $[M]$ for $\{t \in T : (m, t) \in \mathcal{R} \text{ for some } m \in M\}$, and write $n(M)$ for the cardinality of $[M]$, $|[M]|$. If we have for each subset M of S , that $n(M) \geq |M|$, then there is a pairing of S and T (i.e. a bijection $S \rightarrow T$ such that $(s, \phi(s)) \in \mathcal{R}$ for all s). (Note that in the traditional statement of this result, S and T are sets of men and women, and \mathcal{R} is the relation of being acquainted with, then the conclusion of the theorem says that if for each set of k men, they collectively know at least k women, then there is an arrangement by which they can all get married to someone they previously knew).

The proof that all bistochastic matrices are convex linear combinations of permutation matrices is similar in style to the proof that stochastic matrices are convex linear combinations of basic matrices, the difference lying in the inductive step.

As before, we say a matrix A is β -bistochastic if $A_{ij} \geq 0$ and all rows and columns sum to β . In the inductive step, it is necessary only to show that if A is β -bistochastic, then there exists a permutation matrix B and an α such that $A - \alpha B$ is $(\beta - \alpha)$ -bistochastic and has more zero entries than A . To show this, let $S = T = \{1, 2, \dots, n\}$ and $\mathcal{R} = \{(i, j) :$

$A_{ij} > 0$ }. Then let M be a non-empty subset of S . In this case, we have

$$\begin{aligned} \beta \cdot n(M) &= \sum_{j \in [M]} \sum_{i=1}^n A_{ij} \geq \sum_{j \in [M]} \sum_{i \in M} A_{ij} \\ &= \sum_{i \in M} \sum_{j \in [M]} A_{ij} = \sum_{i \in M} \sum_{j=1}^n A_{ij} = \beta \cdot |M|. \end{aligned}$$

We see that the condition for the Hall Marriage Theorem is satisfied unless $\beta = 0$, in which case we are already done, so we get a pairing ϕ between the rows and the columns of the matrix in such a way that $A_{i, \phi(i)} > 0$. Let α be the minimum of these entries and B be the permutation matrix corresponding to ϕ . Then $A - \alpha B$ is $(\beta - \alpha)$ -bistochastic. This completes the inductive step of the proof.

These results have been extended by Kendall [Ken] and from our point of view more usefully by Révész [Re], where it is shown that every Markov chain with countably infinitely many states and a bistochastic transition matrix may be represented by a measured collection of permutations.

Some further known results about representation of Markov chains are to be found in the book [Ki], where he shows the following.

Theorem 1. *If M is a Borel subset of a complete metric space, then any Markov chain on M can be represented by a collection of measurable maps.*

With the notation that $P_x(A) \equiv P(x, A)$, [Ki] then reproduces the following result of [BC].

Theorem 2. *Let M be a connected and locally connected compact metric space. Let \mathcal{M} be a Markov chain on M with transition map P such that P_x depends weak*-continuously*

on x (that is continuously with respect to the weak*-topology on the set of measures on M) and such that P_x has full support for each $x \in M$. Then \mathcal{M} may be represented by a collection of continuous maps on M .

Sudakov [S] has a result which extends Theorem 1 by finding conditions under which a Markov chain may be represented by maps preserving an invariant measure of the Markov chain (see §7).

3. Background for Smooth Representation

The next sections consider the problem of representation of Markov chains on manifolds by maps in the smooth category, where by smooth, we will always mean C^∞ . All manifolds which we consider will be assumed to be connected. The motivation for looking at the problem comes from results like Theorems 1 and 2 about representation of Markov chains on measure spaces and metric spaces. These are described in detail in the book [Ki], and several questions are raised, including the one which is answered below by Theorem 3.

We consider the case where M is a smooth manifold, and \mathcal{M} a Markov chain on M . Under certain further conditions, \mathcal{M} may be represented by a measured collection of smooth maps on M . Specifically, we take M to be a smooth, compact, orientable Riemannian manifold, with metric g say. This induces a natural volume element ω , with associated Riemannian volume measure V , say. Let \mathcal{B} be the σ -algebra of Borel sets on M and let P be the transition map of the Markov chain as described above. We will consider the collection of transition maps P satisfying the following properties:

- (i) P_x is absolutely continuous with respect to V for all $x \in M$,
- (ii) $h(x, y) \equiv dP_x(y)/dV(y)$ is smooth in x and y , for all $x, y \in M$.

(iii) $h(x, y) > 0$, for all $x, y \in M$.

Such a map is called a *smooth transition map with density h* and a Markov chain with a smooth transition map is called a *smooth Markov chain*. Note that it is a crucial part of the definition that the density h is strictly positive. We can now state the main theorem of §3–§6.

Theorem 3. *Suppose M is a smooth, compact, orientable Riemannian manifold. If \mathcal{M} is a smooth Markov chain on M , then \mathcal{M} may be represented by a collection of smooth maps on M .*

4. Physical Motivation for Theorem 3

We first present an outline of the proof of Theorem 3, showing the fluid dynamical motivation. This is not essential for what follows.

The Markov chain is to be represented by a collection of smooth maps. We regard the function $h(x, y)$ as giving the density of maps taking x into a neighbourhood of y (that is the measure of the maps taking x into a neighbourhood U of small diameter about y is approximately $h(x, y)V(U)$). The problem is then to find a collection of maps, and a measure on them such that the density of the images of x under the maps is $h(x, y)$. We are thus seeing the images of x for the varying maps as part of a continuum, and we are seeing how the points of the continuum move as we vary x along smooth paths. Since $h(x, y) > 0$ for all $x, y \in M$, we expect to find at least one map taking any given $x \in M$ to any given $y \in M$. Further, when x moves along any smooth curve (to x' say), we expect the images of the maps to move along curves of the flow, so that if two maps agree at x , then they should agree at x' , and hence everywhere. With this in mind, we impose that there should be exactly one map taking each $x \in M$ to each $y \in M$. Fixing $x_0 \in M$,

each map on the manifold may thus be labelled by the image of x_0 under that map. The maps are then smooth maps f_y with the property that $f_y(x_0) = y$. We then define the map $\alpha_x : y \mapsto f_y(x)$. By the fluid analogy again, we expect the map α_x to be a smooth diffeomorphism, since $\alpha_x(y)$ is the point to which $y = \alpha_{x_0}(y)$ flows as x moves along a path from x_0 to x . (See Figure 3.1.)

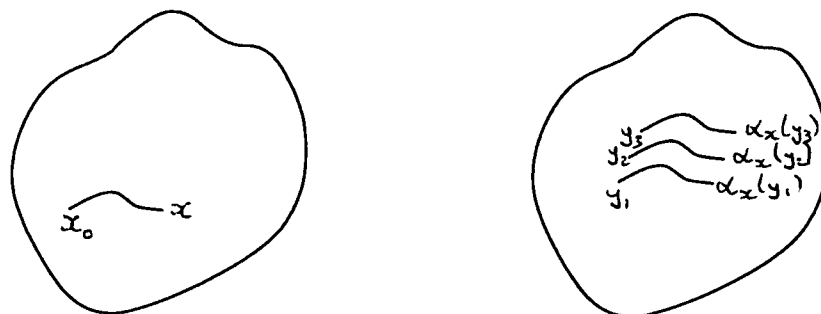


Figure 3.1 Possible paths in the domain and image.

Take \mathcal{P} to be the space of smooth positive density distributions on the manifold (that is smooth functions with $\int f(x)dV(x) = 1; f > 0$), then the diffeomorphisms α on the manifold act naturally on \mathcal{P} as

$$\alpha^* : \mathcal{P} \rightarrow \mathcal{P}; (\alpha^*(\rho))(\alpha(x)) = \rho(x)/\text{Expansion Coefficient}$$

where the expansion coefficient is the limiting ratio of the volume of the image of a neighbourhood (of small diameter) of x to the volume of the neighbourhood (that is the Jacobian of the map α with respect to a set of "locally orthonormal" coordinates). This is just an expression of conservation of mass. For each $x, y \in M$, write $\rho_x(y) = h(x, y)$. Then $\rho_x \in \mathcal{P}$. Further, let ρ_0 be the distinguished density ρ_{x_0} .

We then define the corresponding measures μ_x by $\mu_x(A) = P(x, A)$. Specifically, the correspondence is $d\mu_x/dV = \rho_x$. Set $\mu(A) = P(x_0, A)$. We are then forced to define

$m(\{f_y : y \in A\}) = \mu(A)$ by considering equation (1) in the case that $x = x_0$. Further, by considering equation (1), we see

$$\begin{aligned} P(x, A) &= m(\{f_y : f_y(x) \in A\}) = \mu(\{y : f_y(x) \in A\}) \\ &= \mu(\{y : \alpha_x(y) \in A\}) = \mu(\alpha_x^{-1}(A)). \end{aligned}$$

Since α_x is a homeomorphism, however, we get $P(x, \alpha_x(A)) = \mu(A)$. That is

$$\int_{\alpha_x(A)} \rho_x(y) dV(y) = \int_A \rho_0(y) dV(y)$$

This is equivalent to saying that $\alpha_x^*(\rho_0) = \rho_x$. The problem is then reduced to finding a smoothly parameterized collection of diffeomorphisms α_x such that $\alpha_x^*(\rho_0) = \rho_x$.

It is clearly sufficient to find a collection of diffeomorphisms α_ρ such that $\alpha_\rho^*(\rho_0) = \rho$ with enough smoothness that α_{ρ_x} is smoothly parameterized by x . Given a $\rho \in \mathcal{P}$, define a path in \mathcal{P} by $\rho(t) = \rho_0 + t\eta$ where η is given by $\rho - \rho_0$. We then seek a collection $\alpha_{\rho(t)}$ of diffeomorphisms associated to densities $\rho(t)$ (that is such that $\alpha_{\rho(t)}^*(\rho_0) = \rho(t)$). Moving along this path, there is a constant rate of change of density at each point on the manifold, such as could arise from a constant flux (by comparison with the fluid dynamics equation $\nabla \cdot \Phi + \dot{\rho} = 0$, where $\Phi = \rho v$ is the flux). We therefore seek a flux vector field whose divergence is $-\eta$, and which depends with sufficient smoothness on η . This gives an expression for the velocity of each point in the continuum which gives rise to the required flux (at a specific time, the velocity is given by $\Phi/\rho(t)$). We then let $\alpha_\rho(x)$ be the position of the point x after unit time flow along the parameterized velocity field. We will then find that $\alpha_\rho^*(\rho_0) = \rho$, as required, and it will remain to check that we have the required smoothness. This is shown by the theory of elliptic partial differential equations completing the proof.

5. Differential Equations Background for Theorem 3

For the proof of Theorem 3, we need to use a lemma, which relies on the following theorems from the theory of Green functions for the Laplacian on compact manifolds. The Laplacian is defined by $\Delta f = \nabla_i \nabla^i f$ in local coordinates, where ∇ is the covariant derivative operator on M (with the Riemannian connection).

Theorem 4. *Let M be a smooth, compact, orientable Riemannian manifold. Then if f is a smooth function on M , with $\int f(x)dV(x) = 0$, then there exists a smooth function u with $\Delta u = f$. Further, u is unique up to an additive constant.*

Proof. See [A], §4.1.2

Theorem 5. *Let M be a smooth, compact Riemannian manifold. There exists a function $G : M \times M \rightarrow \mathbb{R}$ such that if ϕ is a smooth function on M , we get*

$$\begin{aligned} \phi(x) &= V(M)^{-1} \int_M \phi(y)dV(y) + \int_M G(x,y)\Delta\phi(y)dV(y) \\ G(x,y) &\geq 0, \quad \forall x,y \in M \\ \int_M G(x,y)dV(y) &= C, \text{ where } C \text{ is a finite constant.} \end{aligned} \tag{2}$$

Proof. See [A], §4.2.3

Define

$$\begin{aligned} \mathcal{P} &= \left\{ f : M \rightarrow (0, \infty) \text{ smooth with } \int f(x)dV(x) = 1 \right\} \\ \mathcal{Z} &= \left\{ f : M \rightarrow \mathbb{R} \text{ smooth with } \int f(x)dV(x) = 0 \right\} \\ \mathcal{V} &= \left\{ \text{Smooth vector fields on } M \right\}. \end{aligned}$$

Lemma 6. *Given a smooth, compact, orientable Riemannian manifold M , and a collection $\{\eta_\beta\}_{\beta \in \mathcal{M}}$ of smoothly parameterized functions in \mathcal{Z} (that is the map $(\beta, x) \mapsto \eta_\beta(x)$)*

is a smooth map $M \times M \rightarrow M$), then there is a map $\Phi : \mathcal{Z} \rightarrow \mathcal{V}$ satisfying

(i) $\operatorname{div}(\Phi(\eta_\beta)) = -\eta_\beta$,

(ii) the map $(\beta, x) \mapsto \Phi(\eta_\beta)(x)$ is smooth.

Proof. Suppose U is an open set in \mathbb{R}^k and $\{\theta_\alpha\}_{\alpha \in U}$ is a smoothly parameterized collection of functions in \mathcal{Z} , then let $H_\alpha(x)$ be the solution of the equation $\Delta H_\alpha = \theta_\alpha$ such that $\int H_\alpha dV = 0$. This exists by Theorem 4, is unique, and is smooth in x . Then by Theorem 5, we see

$$H_\alpha(x) = \int_M G(x, y) \theta_\alpha(y) dV(y).$$

We now have to show that the function $H(\alpha, x) \equiv H_\alpha(x)$ is smooth. The i th parametric partial derivative of H is given by

$$\begin{aligned} \frac{\partial}{\partial \alpha^i} H(\alpha, x) &= \lim_{t \rightarrow 0} \int_M \frac{1}{t} G(x, y) (\theta_{\alpha + t e_i}(y) - \theta_\alpha(y)) dV(y) \\ &= \lim_{t \rightarrow 0} \int_M G(x, y) \left(\frac{\partial}{\partial \alpha^i} \theta_\alpha(y) + \zeta(t, x) \right) dV(y) \end{aligned}$$

where e_i is the i th coordinate vector field in U and $\zeta(t, x)$, the remainder term is smooth in x and $\zeta(t, x) \rightarrow 0$ as $t \rightarrow 0$ for all $x \in M$. It follows that $\zeta(t, x) \rightarrow 0$ uniformly on M (by compactness) as $t \rightarrow 0$, and hence by (2), it follows that

$$\frac{\partial}{\partial \alpha^i} H(\alpha, x) = \int_M G(x, y) \frac{\partial}{\partial \alpha^i} \theta_\alpha(y) dV(y).$$

But the i th partial derivative of θ_α remains a smoothly parameterized collection of functions, and the i th partial derivative of H is clearly continuously dependent on α , and is a smooth function of x by the argument above, so replacing θ_α by its partial derivative in the above procedure shows inductively that $H(\alpha, x)$ depends smoothly on α and x .

Finally, take a chart (U, ψ) of M and use the above, with $\theta_\alpha = -\eta_{\psi^{-1}(\alpha)}$, to get a smooth function $H(\alpha, x)$ such that $\Delta H_\alpha(x) = -\eta_{\psi^{-1}(\alpha)}$. Then set $F_\beta(x) = H_{\psi(\beta)}(x)$ (This is chart-independent by the uniqueness mentioned above.) Patch these together using the independence to get a smooth function $F : M \times M \rightarrow \mathbb{R}$ such that $\Delta F_\beta(x) = -\eta_\beta(x)$ and take (in local coordinates)

$$\Phi(\eta_\beta)^i(x) = \nabla^i F_\beta(x).$$

This gives the map Φ , as required. \square

6. Proof of Theorem 3

Proof of Theorem 3. First note that \mathcal{P} is a convex set, and that there is a canonical map from M to \mathcal{P} given by $x \mapsto \rho_x$ where ρ_x is defined by the equation $\rho_x(y) = h(x, y)$. Let Φ be as defined in Lemma 6 and then define $\gamma_x(y, t)$ by

$$\begin{aligned} \gamma_x(y, 0) &= y, \quad \forall x, y \in M \\ \frac{d\gamma_x(y, t)}{dt} &= \left(\frac{\Phi(\rho_x - \rho_0)}{(1-t)\rho_0 + t\rho_x} \right) (\gamma_x(y, t)). \end{aligned}$$

Write $\gamma_{x,t}(y) \equiv \gamma_x(y, t)$, and using the lemma, we see the vector field above depends smoothly on the parameters x and t , so that we can use the parameterized flow theorem (see [AMR], §21.4) to show that the $\gamma_{\bar{x},t}$ form a smoothly parameterized collection of diffeomorphisms. In particular, define smoothly parameterized diffeomorphisms $\theta_x(y) = \gamma_{x,1}(y)$. We will then show that

$$\int_{\theta_x(A)} \rho_x(y) dV(y) = \int_A \rho_0(y) dV(y) \tag{3}$$

for Borel sets A and $\rho \in \mathcal{P}$. Assuming this for now, we complete the claim by setting

$$f_y(x) = \theta_x(y)$$

$$\mathcal{F} = \{f_y : y \in M\}$$

$$m\{f_y : y \in A\} = P(x_0, A) \text{ for } A \in \mathcal{B}.$$

We then check

$$\begin{aligned} P(x, A) &= \int_A \rho_x(y) dV(y) = \int_{\theta_x^{-1}(A)} \rho_0(y) dV(y) = P(x_0, \theta_x^{-1}(A)) \\ &= P(x_0, \{y : \theta_x(y) \in A\}) = P(x_0, \{y : f_y(x) \in A\}) \\ &= m(\{f_y : f_y(x) \in A\}). \end{aligned}$$

In this, we used for the second equality, (3) and the fact that θ_x is a homeomorphism. This statement is then the required condition, completing the proof subject to the proof of the claim (3) made above.

To prove (3), note that it is sufficient to prove it for sets A which are open subsets of M with piecewise smooth boundary. So fix U open in M with piecewise smooth boundary, take $x \in M$ and set $\rho(t) = (1 - t)\rho_0 + t\rho_x$. We then show the following equation holds.

$$\frac{d}{dt} \left(\int_{\gamma_{x,t}(U)} (\rho(t))(y) dV(y) \right) = 0$$

Equation (3) then follows from this.

Now, set $\eta = \rho_x - \rho_0$ and $X = \Phi(\eta)$. Using the Transport Theorem with mass density (see [AR, §8.2.1]), the left hand side of the above is equal to

$$\int_{\gamma_{x,t}(U)} \frac{d\rho(t)}{dt} \omega + \mathcal{L}_{X/\rho(t)}(\rho(t)\omega)$$

where \mathcal{L}_Y is the Lie derivative in direction Y . The integrand is then equal to

$$\begin{aligned}
& \eta\omega + \omega\mathcal{L}_{X/\rho(t)}(\rho(t)) + \rho(t)\mathcal{L}_{X/\rho(t)}(\omega) \\
&= \eta\omega + \frac{\omega}{\rho(t)}\left(X^i\frac{\partial}{\partial x^i}\rho(t)\right) + \rho(t)\operatorname{div}\left(\frac{X}{\rho(t)}\right)\omega \\
&= \eta\omega + \frac{\omega}{\rho(t)}\left(X^i\frac{\partial}{\partial x^i}\rho(t)\right) + (\operatorname{div}(X))\omega + X^i\rho(t)\omega\nabla_i\left(\frac{1}{\rho(t)}\right) = 0.
\end{aligned}$$

This completes the proof. \square

7. Invariant Measures for Smooth Markov Chains

In this section, \mathcal{M} will continue to be a smooth Markov chain on a compact Riemannian manifold. We will assume that the volume measure V is normalized: $V(M) = 1$.

Write $\mathcal{P}(M)$ for the collection of probability measures on M . Define $\mathcal{P}_S(M) = \{\mu \in \mathcal{P}(M) : d\mu(x) = g(x)dV(x) \text{ for a smooth } g\}$. These are the *smooth measures*. We then define the *transition operator* $P^* : \mathcal{P}(M) \rightarrow \mathcal{P}(M)$ by $P^*[\mu](A) = \int P(x, A)d\mu(x)$. We say a measure μ is *invariant* if $P^*\mu = \mu$.

Theorem 7. *Let \mathcal{M} be a smooth Markov chain. Then the following hold.*

- (i) \mathcal{M} has a unique invariant probability measure ν .
- (ii) ν is smooth and has everywhere positive density.
- (iii) If μ is any probability measure, then $P^{*n}\mu \rightarrow \nu$ as $n \rightarrow \infty$ in the strong topology.
- (iv) ν depends continuously on the transition density h .

In fact, in (iii), we have $\operatorname{diam}(P^{*n}[\mathcal{P}(M)]) \rightarrow 0$ as $n \rightarrow \infty$. Before we prove the theorem, it will be necessary to establish a preliminary lemma.

Lemma 8. $\operatorname{Im} P^* \subset \mathcal{P}_S(M)$.

Proof.

$$\begin{aligned} P^*\mu(A) &= \int P(x, A) d\mu(x) = \int \left(\int_A h(x, y) dV(y) \right) d\mu(x) \\ &= \int_A \left(\int h(x, y) d\mu(x) \right) dV(y). \end{aligned}$$

From this, we see that $P^*\mu$ is absolutely continuous with respect to V and $dP^*\mu(y)/dV(y) = \int h(x, y) d\mu(x)$ which is smooth. \square

Note that this means that any P^* -invariant measure must be smooth. Using [Ki], which asserts the existence of an invariant measure for such a system, we have therefore guaranteed the existence of a smooth invariant measure, ν say. We now turn to the question of uniqueness.

Note that $\mathcal{P}_S(M)$ is in bijection with $\{g \in C^\infty(M) : g(x) \geq 0, \forall x \in M; \int g(x) dV(x) = 1\}$, so we may consider the restriction of P^* to $\mathcal{P}_S(M)$ by looking at the map $L : C^\infty(M) \rightarrow C^\infty(M)$ given by the equation below, since if μ is absolutely continuous with respect to V with Radon-Nikodym derivative g , then $P^*\mu$ is absolutely continuous with respect to V with derivative $L[g]$, where

$$L[g](y) = \int g(x) h(x, y) dV(x).$$

We in fact consider L as a map $C(M) \rightarrow C(M)$ (noting however that by the lemma, $\text{Im } L \subset C^\infty(M)$). Equip $C(M)$ with its usual norm $\|g\| = \sup_{x \in M} |g(x)|$ and set $\|g\|_1 = \int |g(x)| dV(x)$. Note that $\|g\|_1 \leq \|g\|$. The operator L is positive (i.e. $f \geq 0 \Rightarrow L[f] \geq 0$) and

$$\begin{aligned} \int L[g](x) dV(x) &= \int \left(\int g(x) h(x, y) dV(x) \right) dV(y) \\ &= \int g(x) \left(\int h(x, y) dV(y) \right) dV(x) = \int g(x) dV(x). \end{aligned} \tag{4}$$

Define

$$S_0 = \{f \in C(M) : \int f(x)dV(x) = 0\}$$

$$S_1 = \{f \in C(M) : f(x) \geq 0, \forall x; \int f(x)dV(x) = 1\}.$$

By the above, $L(S_0) \subset S_0$ and $L(S_1) \subset S_1$.

We now proceed to the proof of the Theorem 7.

Proof of Theorem 7.

Since we have shown any invariant measure is smooth, we are looking for fixed points of L lying in S_1 . For (i) and (ii), we therefore need to show that there is exactly one function $g_0 \in S_1$ such that $L[g_0] = g_0$. This function must also have $g_0(x) > 0, \forall x$. Since we are already guaranteed the existence of a smooth invariant measure ν , we write g_0 for its derivative $d\nu/dV$. Then $L[g_0] = g_0$.

Suppose we have an independent function g_1 such that $L[g_1] = g_1$. We may then assume by scaling and adding that there is a function $g \in S_0 - \{0\}$ with $L[g] = g$. Write $g_+ = \max(g, 0)$, $g_- = \min(g, 0)$, $V_+ = \{x \in M : g(x) \geq 0\}$ and $V_- = \{x \in M : g(x) \leq 0\}$. Then $g = g_+ + g_-$ and $L[g_+] \geq L[g_+ + g_-] = g_+ + g_-$. But $x \in V_+ \Rightarrow L[g_+](x) \geq g_+(x) + g_-(x) = g_+(x)$ and $x \in V_- \Rightarrow L[g_+](x) \geq 0 = g_+(x)$, so we see that $L[g_+] \geq g_+$. By (4) then, we see that $L[g_+] = g_+$. Now, pick $y \in V_-$. We have $g_+(y) = L[g_+](y) = \int g_+(x)h(x, y)dV(x) > 0$ and this a contradiction. This proves part (i). This last argument can be applied to g to prove part (ii) also.

We now move on to part (iii). Let μ be a probability measure. By Lemma 8, $P^*\mu$ is a smooth probability measure. Suppose its derivative with respect to V is g . Then $g \in S_1$. It follows that $g_0 - g \in S_0$. We therefore have to show that $L^n[g_0 - g]$ converges uniformly to 0. It is clearly sufficient to show that $f \in S_0 \Rightarrow \|L^n f\| \rightarrow 0$.

Define $\sigma = \min_{x,y} h(x,y)$ and $\tau = \max_{x,y} h(x,y)$. Note that $0 < \sigma \leq 1 \leq \tau$. We have

$$\begin{aligned} \|L[f]\| &= \sup_y \left| \int h(x,y)f(x)dV(x) \right| \leq \sup_y \int h(x,y)|f(x)|dV(x) \\ &\leq \tau \|f\|_1 \end{aligned}$$

Write $f = f_+ + f_-$ as before. Note $\|f_+\|_1 = \|f_-\|_1$ and $\|f\|_1 = \|f_+\|_1 + \|f_-\|_1$. Then $L[f_+](y) = \int f_+(x)h(x,y)dV(x) \geq \sigma \|f_+\|_1$. Similarly, $L[f_-](y) \leq -\sigma \|f_-\|_1$. We then have

$$\begin{aligned} \|L[f]\|_1 &= \|L[f_+] + L[f_-]\|_1 \\ &= \left\| L[f_+] - \sigma \|f_+\|_1 + L[f_-] + \sigma \|f_-\|_1 \right\|_1 \\ &\leq \left\| L[f_+] - \sigma \|f_+\|_1 \right\|_1 + \left\| L[f_-] + \sigma \|f_-\|_1 \right\|_1 \\ &= \|L[f_+]\|_1 - \sigma \|f_+\|_1 + \|L[f_-]\|_1 - \sigma \|f_-\|_1 \\ &= \|f_+\|_1 + \|f_-\|_1 - \sigma (\|f_+\|_1 + \|f_-\|_1) \\ &= (1 - \sigma) \|f\|_1. \end{aligned}$$

From this, we see $\|L^n[f]\| \leq \tau \|L^{n-1}[f]\|_1 \leq \tau(1 - \sigma)^{n-1} \|f\|_1 \leq \tau(1 - \sigma)^{n-1} \|f\|$. So $L^n[f]$ converges uniformly to 0 as required. This completes the proof of part (iii).

We now prove part (iv). By the above, we see that $\|L^n|_{S_0}\| \leq \tau(1 - \sigma)^{n-1}$. We can therefore define $\Phi : S_0 \rightarrow S_0$ by $\Phi = 1 + L + L^2 + \dots$. Then note that $1 + L \circ \Phi = \Phi$.

Suppose g is the invariant member of S_1 for L . Then suppose δL is a perturbation of L with $\|\delta L\| < 1/\|\Phi\|$. Form $\Psi = \Phi \circ \delta L$. Then $\delta L + L \circ \Psi = \Psi$.

Now

$$\begin{aligned} &(L + \delta L)[g + \Psi[g] + \Psi^2[g] \dots] \\ &= g + (L \circ \Psi + \delta L)[g] + (L \circ \Psi + \delta L)[\Psi[g]] \dots \\ &= g + \Psi[g] + \Psi^2[g] + \dots \end{aligned}$$

It follows that $g + \Psi[g] + \Psi^2[g] \dots$ is the invariant member of S_1 for $L + \delta L$. Note the invariant functions differ by norm at most $\|g\| \|\Phi\| \|\delta L\| / (1 - \|\Phi\| \|\delta L\|)$, so that as $\|\delta L\| \rightarrow 0$, $\delta g \rightarrow 0$ also. \square

Chapter 4. Representation of Markov Chains on Tori

In this chapter, we continue the investigation started in Chapter 3 of representation of Markov chains on manifolds. In particular, we look at the special case where the base space of the Markov chain is a torus. This allows us to insist that the maps used in the representation have special properties.

In §1, we consider the case where the base space M of the Markov chain is an n -torus (T^n). In this case any smooth Markov chain may be represented by homotopic N -to-1 local diffeomorphisms for some sufficiently large value of N . The material of this chapter has appeared as [Q3].

It is natural to ask which values of N can occur in the results of §1. In §2, we give an answer to this question in the case that the underlying manifold is the circle. In particular, we exhibit a Markov chain which cannot be represented by degree 1 homeomorphisms.

This partly answers a second question in [Ki], where he asks: Can every smooth Markov chain on a manifold be represented by diffeomorphisms? We give in §3 an example of a Markov chain on S^1 which cannot be represented by diffeomorphisms (that is which cannot be represented by a combination of degree 1 and degree -1 diffeomorphisms.)

The material contained in this chapter is all original except that I received some advice on the proof of Theorem 1, where my original version was somewhat more complicated than the current version.

1. Representation of Markov Chains on Tori

In what follows, we will frequently refer to ‘regarding the torus \mathbb{T}^n as $[0, 1)^n \bmod 1$ ’. By this, we mean that we identify \mathbb{T}^n with $\mathbb{R}^n/\mathbb{Z}^n$ and denote an equivalence class by its unique representative in the set $[0, 1)^n$. The additive structure of the torus is then referred to as ‘addition mod 1’. This will be denoted in the normal way by $+$ and it should be clear from the context whether addition is taking place in \mathbb{R}^n or \mathbb{T}^n .

We also frequently refer to the lift of a map of the torus. Suppose the map $\phi : \mathbb{T}^n \rightarrow \mathbb{T}^n$ is continuous. Then a lift of ϕ is a continuous map $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $\pi \circ \Phi = \phi \circ \pi$ where π is the standard projection $\mathbb{R}^n \rightarrow \mathbb{T}^n$. Such a lift is unique up to an additive integer vector constant. The map Φ may then be uniquely decomposed into a linear part and a periodic part as $\Phi(x) = p(x) + Ax$ where A is an integer matrix and $p(x + m) = p(x)$ for all $m \in \mathbb{Z}^n$. Note that two maps of \mathbb{T}^n are homotopic if and only if their lifts have equal linear parts. In the case that the map is a map of the circle, the matrix of the linear part is just a number and is known as the *degree* of a map.

Theorem 1. *Let \mathcal{M} be a Markov Chain on \mathbb{T}^n with a smooth transition map. Then \mathcal{M} may be represented by a collection of homotopic N -to-1 surjections for some N .*

Proof. By Theorem 3.3, there exists a smoothly parameterized collection $\{f_y\}_{y \in \mathbb{T}^n}$ of maps with the property that for each pair $x, z \in \mathbb{T}^n$, there is exactly one y such that $f_y(x) = z$. In this case, write $y = \phi(x, z)$. For fixed x , the map $z \mapsto \phi(x, z)$ is a diffeomorphism of \mathbb{T}^n . We also have that $P(x, A) = \mu\{y : f_y(x) \in A\}$ where there exists an $x_0 \in \mathbb{T}^n$ such that for all $B \in \mathcal{B}$, $\mu(B) = P(x_0, B)$. As such, μ is a smooth volume form on \mathbb{T}^n , so by Moser’s Theorem ([Mo]), there exists a smooth diffeomorphism $\alpha : \mathbb{T}^n \rightarrow \mathbb{T}^n$ such that $\mu(B) = \lambda(\alpha^{-1}B)$ for Borel sets B , where λ is Haar measure.

Now define $e_y(x) = f_{\alpha(y)}(x)$. Then

$$\begin{aligned} P(x, A) &= \mu\{y : f_y(x) \in A\} = \lambda\{\alpha^{-1}(y) : f_y(x) \in A\} \\ &= \lambda\{y : f_{\alpha(y)}(x) \in A\} = \lambda\{y : e_y(x) \in A\}. \end{aligned}$$

Note also that $e_{\alpha^{-1}(\phi(x,z))}(x) = f_{\phi(x,z)}(x) = z$, so the collection e_y has exactly the properties of the collection f_y except that the measure on the parameters is just Haar measure. We may therefore assume without loss of generality that the original measure μ was in fact Haar measure.

Next, write $\phi_z(x) = \phi(x, z)$. Then ϕ_z has lift $\Phi_z : \mathbb{R}^n \rightarrow \mathbb{R}^n$ say. As usual, we write $\Phi_z(x) = A_z x + p_z(x)$ where A_z is an integer matrix and p_z is periodic. Since the collection ϕ_z is smoothly parameterized, it follows that the linear part A_z is continuously dependent on z , so since A_z is an integer matrix, A_z must be constant, say $A_z = A$.

Now choose a norm $\|\cdot\|$ on \mathbb{R}^n . This induces an operator norm (which we will also denote by $\|\cdot\|$) on $M_n(\mathbb{R})$, the $n \times n$ matrices over \mathbb{R} satisfying $\|Ax\| \leq \|A\|\|x\|$. Consider \mathbb{T}^n as $[0, 1)^n \bmod 1$, pick $M \in \mathbb{N}$ such that $M > \|A\| + \sup_{x,z \in \mathbb{T}^n} \|D_x p_z\|$ and set $\theta(x, z) = \phi(x, z) + Mx$. For fixed x , the map $z \mapsto \theta(x, z)$ remains a diffeomorphism of \mathbb{T}^n . Set $g_y(x) = \theta_x^{-1}(y)$. Clearly $g_y(x)$ is continuously dependent on y for fixed x , and so the maps g_y are certainly homotopic. Note also that

$$\begin{aligned} P(x, A) &= \lambda\{y : f_y(x) \in A\} = \lambda\{\phi(x, z) : z \in A\} \\ &= \lambda\{\theta(x, z) : z \in A\} = \lambda\{y : g_y(x) \in A\}, \end{aligned}$$

so the collection $\{g_y\}_{y \in \mathbb{T}^n}$ represents \mathcal{M} . It therefore remains to show that all the maps g_y are N -to-1 surjections for some uniform N .

To prove this, consider $g_y^{-1}\{z\} = \{x : \theta_x(z) = y\}$. Setting $\gamma_z(x) = \theta(x, z)$, we see

$g_y^{-1}\{z\} = \gamma_z^{-1}\{y\}$, so it is sufficient to show that γ_z is an N -to-1 surjection for the some N which is independent of z . But $\gamma_z(x) = \phi(x, z) + Mx$, which has lift $\Gamma_z(x) = Ax + Mx + p_z(x)$. Write L for the matrix $(A + MI)$ where I is the identity matrix and suppose $x \neq y$. Then

$$\begin{aligned} \|\Gamma_z(x) - \Gamma_z(y)\| &= \|M(x - y) + A(x - y) + p_z(x) - p_z(y)\| \\ &\geq M\|x - y\| - (\|A\| + \sup_{z, x \in T^n} \|D_x p_z\|)\|x - y\| > 0. \end{aligned}$$

so Γ_z is injective.

We now show Γ_z is surjective. Since $\|A\| < M$, we see that the matrix L is invertible, so given $y \in \mathbb{R}^n$, define the map

$$F : \mathbb{R}^n \rightarrow \mathbb{R}^n ; x \mapsto L^{-1}(y - p_z(x)).$$

The image of F is a bounded subset of \mathbb{R}^n and so is contained in some closed ball $B(0, R)$. Now consider F as a map from $B(0, R)$ into itself. By the Brouwer fixed point theorem, there exists a point $x_0 \in B(0, R)$ such that $F(x_0) = x_0$. Then $x_0 = L^{-1}(y - p_z(x_0))$, so we see that $\Gamma_z(x_0) = y$. It follows then that Γ_z is surjective. We then show that this implies that γ_z is a $|\det L|$ -to-1 surjection. Note that \mathbb{Z}^n is the disjoint union of cosets $L\mathbb{Z}^n + x_i$, $1 \leq i \leq m$ where $m = |\det L|$ by standard theory of maps on tori. Denote by π the standard projection from \mathbb{R}^n to \mathbb{T}^n and pick $\zeta \in \mathbb{T}^n$. Then $\pi^{-1}(\zeta) = \mathbb{Z}^n + x$ for some $x \in \mathbb{R}^n$. Let $\rho_i = \pi(\Gamma_z^{-1}(x + x_i))$. These are distinct, for if $\rho_i = \rho_j$, then

$$\Gamma_z^{-1}(x + x_i) = \Gamma_z^{-1}(x + x_j) + \mathbf{m}, \text{ where } \mathbf{m} \in \mathbb{Z}^n.$$

Applying Γ_z , we get $x_i = x_j + L\mathbf{m}$ which implies $i = j$. It therefore follows that the points ρ_1, \dots, ρ_m are distinct and $\gamma_z^{-1}\{\zeta\} \supset \{\rho_1, \dots, \rho_m\}$.

Conversely, suppose $\gamma_z(\rho) = \zeta$, then pick $w \in \pi^{-1}(\rho)$. So, $\pi(\Gamma_z(w)) = \zeta$ and $\Gamma_z(w) \in \mathbb{Z}^n + x$, so in particular $\Gamma_z(w) = L\mathbf{m} + x + x_i$ for some $\mathbf{m} \in \mathbb{Z}^n$ and some i . So

$$\Gamma_z(w) = \Gamma_z(\mathbf{m} + \Gamma_z^{-1}(x + x_i)).$$

From this, we deduce $w = \mathbf{m} + \Gamma_z^{-1}(x + x_i)$ and $\rho = \rho_i$. Then $\gamma_z^{-1}\{\zeta\} = \{\rho_1, \dots, \rho_m\}$ and γ_z is $|\det L|$ -to-1 as required. This number is clearly independent of z . \square

Further, we may characterize the homotopy class of the maps as follows. By standard results on the theory of maps of the torus, the map θ has a lift $\Theta : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$. The lift Θ may then, as usual, be split up into linear and periodic parts: $\Theta(x, z) = Ax + Bz + C(x, z)$, where A and B are integer matrices, and C is periodic in x and z . Note that $|\det B| = 1$ as the map $z \mapsto \theta(x, z)$ is a diffeomorphism for fixed x . We have, however, that $g_y(x) = \theta_x^{-1}(y)$. Let $G_y(x)$ be the lift of g_y . By definition, we see $\theta(x, g_y(x)) = y$. Lifting this, we get $\Theta(x, G_y(x)) = Y$, where Y is a preimage of y under the natural projection. Substitution gives $Ax + BG_y(x) + C(x, G_y(x)) = Y$. As x varies, the right hand side must remain a preimage of y , so by continuity, we have that the right hand side is constant. As x moves through an integer displacement $\mathbf{m} \in \mathbb{Z}^n$, Ax moves through $A\mathbf{m}$ and $C(x, G_y(x))$ remains constant as C is periodic. It therefore follows that $BG_y(x + \mathbf{m}) = BG_y(x) - A\mathbf{m}$, so in particular

$$G_y(x + \mathbf{m}) = G_y(x) - B^{-1}A\mathbf{m}.$$

The linear part of G_y therefore has matrix $-B^{-1}A$, where A is the matrix of the linear part of θ for fixed z (considered as a map of x) and B is the corresponding matrix for fixed x .

2. Degree of a Markov Chain on the Circle

Definition. A smooth Markov chain \mathcal{M} on a smooth Riemannian manifold M is said to be nicely represented by a collection $\{f_y\}_{y \in M}$ of maps and a volume form μ on M if the following properties hold:

- (i) For all points x and z in M , there exists a unique y in M such that $f_y(x) = z$. In this case, write $y = Y(x, z)$.
- (ii) $Y(x, z)$ as defined above is smooth in both variables and for fixed x , the map $z \mapsto Y(x, z)$ is a diffeomorphism of M .
- (iii) For each Borel set A and each point $x \in M$, $P(x, A) = \mu\{y : f_y(x) \in A\}$.

By the proof of Theorem 3.3, every smooth Markov chain has a nice representation by smooth maps. We now restrict ourselves to the case where $M = S^1$.

Definition. The positive and negative degrees of \mathcal{M} are given by

$$\begin{aligned}\delta_+ &= \int \sup_z \frac{\partial}{\partial x} P(x, [z_0, z]) dx \\ \delta_- &= \int \inf_z \frac{\partial}{\partial x} P(x, [z_0, z]) dx.\end{aligned}$$

Note that these degrees are independent of the point z_0 as

$$\frac{\partial}{\partial x} P(x, [z_0, z]) = \frac{\partial}{\partial x} P(x, [z_0, z']) + \frac{\partial}{\partial x} P(x, [z', z]).$$

Taking suprema over z , the first term of the right hand side is unaffected, and then integrating, this term gives no contribution, so we see the degree δ_+ is independent of the point z_0 . The same obviously holds for δ_- . We therefore fix a point $z_0 \in S^1$ for the rest of this section. Note that $\delta_- \leq 0 \leq \delta_+$, as can be seen by taking $z \equiv z_0$.

Theorem 2. *Let $N > 0$. Suppose \mathcal{M} is a smooth Markov chain on the circle. \mathcal{M} may be nicely represented by degree N local diffeomorphisms if and only if $N > \delta_+$.*

Theorem 3. *Let $N > 0$. Suppose \mathcal{M} is a smooth Markov chain on the circle which may be represented by degree N local homeomorphisms. Then $N \geq \delta_+$.*

Corollary. *Let $N > 0$. There exist smooth Markov chains on the circle which cannot be represented by degree N local homeomorphisms.*

Note these have corresponding versions with $N < 0$ involving the quantity δ_- .

Theorem 2'. *Let $N < 0$. Suppose \mathcal{M} is a smooth Markov chain on the circle. \mathcal{M} may be nicely represented by degree N local diffeomorphisms if and only if $N < \delta_-$.*

Theorem 3'. *Let $N < 0$. Suppose \mathcal{M} is a smooth Markov chain on the circle which may be represented by degree N local homeomorphisms. Then $N \leq \delta_-$.*

Corollary'. *Let $N < 0$. There exist smooth Markov chains on the circle which cannot be represented by degree N local homeomorphisms.*

Proof of Theorem 2. Suppose \mathcal{M} is nicely represented by degree N local diffeomorphisms. We may then assume the measure on the parameter space to be Haar measure as in §1. Write $Y(x, z)$ for the parameter value of the unique map taking x into z and write $h(x, z)$ for the probability density of going from x to z . It then follows that

$$\frac{\partial}{\partial z} Y(x, z) = h(x, z).$$

In what follows, we treat the circle as the interval $[0, 1) \bmod 1$. We then see that $Y(x, z) = P(x, [z_0, z]) - X(x)$ where X is some map $S^1 \rightarrow S^1$. For \mathcal{M} to be represented by local diffeomorphisms, by the implicit function theorem and the condition that the maps are

locally orientation-preserving, we require

$$\frac{d}{dx}X(x) > \sup_{z \in S^1} \frac{\partial}{\partial x}P(x, [z_0, z]).$$

We therefore see that X has degree greater than δ_+ . But we have

$$N = |\{x : Y(x, z) = y\}| = |\{x : P(x, [z_0, z]) - X(x) = y\}|.$$

By the conditions placed on X , we have $P(x, [z_0, z]) - X(x)$ is monotonic in x . The cardinality is then the modulus of the degree of the expression as a function of x . This is precisely the degree of X , so we see that $N > \delta_+$.

Conversely, suppose $N > \delta_+$. Set

$$\alpha(x) = \sup_z \frac{\partial}{\partial x}P(x, [z_0, z]).$$

Then we have $\int \alpha(x)dx = \delta_+$, so we can find an $\epsilon > 0$ and a smooth function $\beta(x)$ such that

$$(i) \beta(x) \geq \alpha(x) + \epsilon$$

$$(ii) \int \beta(x)dx = N.$$

Then set $X(x) = \int_{z_0}^x \beta(x)dx \pmod{1}$, and finally, let $Y(x, z) = \pi(P(x, [z_0, z]) - X(x))$, where π is the standard projection of the real line onto the circle. For fixed $x \in S^1$, the map $z \mapsto Y(x, z)$ is a diffeomorphism $S^1 \rightarrow S^1$. Y is also smooth in x . Write $Y_x(z) = Y(x, z)$ and define $f_y(x) = Y_x^{-1}(y)$. By the implicit function theorem, we see that $D_x f_y \neq 0$ since $D_x Y \neq 0$ and $D_z Y \neq 0$. The map f_y is a smooth local diffeomorphism, and

$$f_y^{-1}\{z\} = \{x : f_y(x) = z\} = \{x : y = Y_x(z)\}.$$

From this, we see that f_y is N -to-1 and so f_y has degree N . Finally, we check that with μ taken to be Haar measure, this does indeed provide a nice representation of the Markov chain \mathcal{M} .

$$\begin{aligned}\mu\{y : f_y(x) \in A\} &= \mu\{y : Y_x^{-1}(y) \in A\} = \mu\{Y_x(z) : z \in A\} \\ &= \mu\left\{P(x, [z_0, z]) : z \in A\right\} = P(x, A).\end{aligned}$$

Note that the third equality uses the translation invariance of Haar measure. It follows that \mathcal{M} is nicely represented by degree N local diffeomorphisms. \square

Before embarking on the proof of Theorem 3, we need some lemmas and definitions.

Definition. A map $p : S^1 \times \mathcal{B} \rightarrow [0, 1]$ is an *S-map* if

- (i) For each $x \in S^1$, the map $A \mapsto p(x, A)$ is a measure,
- (ii) $p(x, S^1)$ is independent of x and
- (iii) For fixed $A \in \mathcal{B}$, the map $x \mapsto p(x, A)$ is measurable.

Note, an S-map is just a constant multiple of a transition map. An S-map is said to be *smooth* if it is a constant multiple of a smooth transition map. In particular, it is required to have strictly positive transition densities or all densities identically zero.

Further, if p_1 and p_2 are S-maps, then we say p_1 is *subordinate* to p_2 if $p_1(x, A) \leq p_2(x, A)$ for each $x \in S^1$ and $A \in \mathcal{B}$.

The *weight* of an S-map p is denoted by $w(p)$ and is defined to be $p(x, S^1)$ (which is independent of x).

Definition. Suppose the Markov chain \mathcal{M} is represented by the collection of maps \mathcal{F} and a measure ν on them. A *subrepresentation* of this is defined by a measurable subset \mathcal{F}' of \mathcal{F} and the restriction of the measure ν to a measure ν' defined on the measurable

subsets of \mathcal{F}' .

Note that in this case the induced S-map (defined by $p(x, A) = \nu'\{f \in \mathcal{F}' : f(x) \in A\}$) is subordinate to P .

Let Σ be the collection of all S-maps. Then define the maps

$$V_+ : \Sigma \rightarrow \mathbb{R}^+; p \mapsto \lim_{m \rightarrow \infty} \sup_{z_1 \dots z_m} \sum_{i=1}^m \left[p\left(\pi\left(\frac{i}{m}\right), [z_0, z_i]\right) - p\left(\pi\left(\frac{i-1}{m}\right), [z_0, z_i]\right) \right]$$

$$V_- : \Sigma \rightarrow \mathbb{R}^-; p \mapsto \lim_{m \rightarrow \infty} \inf_{z_1 \dots z_m} \sum_{i=1}^m \left[p\left(\pi\left(\frac{i}{m}\right), [z_0, z_i]\right) - p\left(\pi\left(\frac{i-1}{m}\right), [z_0, z_i]\right) \right].$$

These quantities satisfy for all $p \in \Sigma$, $V_-(p) \leq 0 \leq V_+(p)$.

Lemma 4. *Let p be a smooth S-map. Then we have*

$$\int \sup_z \frac{\partial}{\partial x} p(x, [z_0, z]) dx = V_+(p).$$

Clearly a similar relation will hold for V_- .

Proof. Set $A(x, z) = p(x, [z_0, z])$. This is smooth in x and z . Write A_x for $\frac{\partial}{\partial x} A$. Then

$$\begin{aligned} p\left(\pi\left(\frac{i}{m}\right), [z_0, z_i]\right) - p\left(\pi\left(\frac{i-1}{m}\right), [z_0, z_i]\right) &= \int_{\pi(i-1/m)}^{\pi(i/m)} \frac{\partial}{\partial x} p(x, [z_0, z_i]) dx \\ &\leq \int_{\pi(i-1/m)}^{\pi(i/m)} \sup_z \frac{\partial}{\partial x} p(x, [z_0, z]) dx. \end{aligned}$$

From this, we see

$$\sup_{z_1 \dots z_m} \sum_{i=1}^m \left[p\left(\pi\left(\frac{i}{m}\right), [z_0, z_i]\right) - p\left(\pi\left(\frac{i-1}{m}\right), [z_0, z_i]\right) \right] \leq \int \sup_z \frac{\partial}{\partial x} p(x, [z_0, z]) dx.$$

This shows the right hand side is bounded above by the left hand side.

Now we show they are equal. Pick $\epsilon > 0$. By uniform continuity, there exists a $\delta > 0$ such that $|x_1 - x_2| < \delta \Rightarrow \forall z, |A_x(x_1, z) - A_x(x_2, z)| < \epsilon$. Now pick $m \in \mathbb{N}$ such that $m > \delta^{-1}$, and choose i with $1 \leq i \leq m$. Then let $y \in S^1$ be such that $A_x(\frac{i-1}{m}, y) = \sup_z A_x(\frac{i-1}{m}, z)$. Then

$$x \in [\frac{i-1}{m}, \frac{i}{m}] \Rightarrow A_x(x, y) \geq A_x(\frac{i-1}{m}, y) - \epsilon.$$

Also,

$$A_x(x, z) \leq A_x(\frac{i-1}{m}, z) + \epsilon \leq A_x(\frac{i-1}{m}, y) + \epsilon, \forall z.$$

It follows that $A_x(x, y) \geq \sup_z A_x(x, z) - 2\epsilon$. Integrating between $\pi(\frac{i-1}{m})$ and $\pi(\frac{i}{m})$, we get

$$A(\frac{i}{m}, y) - A(\frac{i-1}{m}, y) \geq \int_{\frac{i-1}{m}}^{\frac{i}{m}} \left[\sup_z \frac{\partial}{\partial x} p(x, [z_0, z]) - 2\epsilon \right] dx.$$

Finally, adding gives

$$\sup_{z_1, \dots, z_m} \left[\sum_{i=1}^m p(\pi(\frac{i}{m}), [z_0, z_i]) - p(\pi(\frac{i-1}{m}), [z_0, z_i]) \right] \geq \left(\int \sup_z \frac{\partial}{\partial x} p(x, [z_0, z]) dx \right) - 2\epsilon.$$

This completes the proof of the lemma. \square

Lemma 5. Suppose p is an S -map arising from some measure ν on some collection \mathcal{F} of degree N local homeomorphisms, with $N > 0$ (possibly with $\nu(\mathcal{F}) \neq 1$), then $V_+(p) \leq Nw(p)$.

Again, there will be a similar version of this lemma which operates for $N < 0$ and using V_- .

Proof.

$$\begin{aligned}
V_+(p) &= \lim_{m \rightarrow \infty} \sup_{z_1 \dots z_m} \sum_{i=1}^m \left[p\left(\pi\left(\frac{i}{m}\right), [z_0, z_i]\right) - p\left(\pi\left(\frac{i-1}{m}\right), [z_0, z_i]\right) \right] \\
&= \lim_{m \rightarrow \infty} \sup_{z_1 \dots z_m} \int \sum_{i=1}^m \left[\chi_{[z_0, z_i]} \left(f\left(\frac{i}{m}\right)\right) - \chi_{[z_0, z_i]} \left(f\left(\frac{i-1}{m}\right)\right) \right] d\nu(f) \\
&\leq \lim_{m \rightarrow \infty} \int \sum_{i=1}^m \sup_{z_i} \left[\chi_{[z_0, z_i]} \left(f\left(\frac{i}{m}\right)\right) - \chi_{[z_0, z_i]} \left(f\left(\frac{i-1}{m}\right)\right) \right] d\nu(f)
\end{aligned}$$

where χ_A is the characteristic function of the set A . But we have

$$\begin{aligned}
&\sum_{i=1}^m \sup_{z_i} \left[\chi_{[z_0, z_i]} \left(f\left(\frac{i}{m}\right)\right) - \chi_{[z_0, z_i]} \left(f\left(\frac{i-1}{m}\right)\right) \right] \\
&= \left| \left\{ i : 1 \leq i \leq m, z_0 \leq f\left(\frac{i}{m}\right) < f\left(\frac{i-1}{m}\right) \right\} \right|.
\end{aligned}$$

where by inequalities on the circle, we mean that there is a continuous choice of argument on a connected subset of the circle including the specified points on which the order of the values of the argument is that specified. The cardinality above is however bounded above by N as there can be at most one such i between any adjacent pair of preimages of z_0 under f . It therefore follows that $V_+(p) \leq N\nu(\mathcal{F}) = Nw(p)$. \square

Theorem 3 then follows as a straightforward application of these lemmas.

Proof of Theorem 3. Applying Lemmas 4 and 5 to the transition map P of the Markov chain, we see $\delta_+ = V_+(P) \leq Nw(P)$. But, we also have that $w(P) = 1$, so the theorem is proved. \square

Proof of the Corollary. To prove the corollary, it is sufficient to construct a Markov chain with $\delta_+ > N$. As an example of such a Markov chain, consider the following:

Considering the circle as the interval $[0, 1) \bmod 1$, and given positive constants α and β , such that $\alpha < 1$ and $\beta < \frac{1}{2}$, pick a smooth function f such that

$$(i) f(x) > 0, \forall x \in S^1$$

$$(ii) \int f(x)dx = 1$$

$$(iii) \int_{1/2-\beta}^{1/2+\beta} f(x)dx = \alpha$$

$$(iv) f(x) = \frac{1-\alpha}{1-2\beta} \text{ for } x \notin (\frac{1}{2}-\beta, \frac{1}{2}+\beta).$$

Next, set $h(x, z) = f(z + rx)$, where $r \in \mathbb{N}$, and using this, define P to be the transition map with probability density h : For a Borel set A and a point $x \in S^1$, $P(x, A)$ is defined to be $\int_A h(x, z)dz$. We then estimate the value of $\sup_z \frac{\partial}{\partial x} P(x, [z_0, z])$ as follows:

$$\begin{aligned} \frac{\partial}{\partial x} P(x, [z_0, z]) &= \frac{\partial}{\partial x} \int_{z_0}^z h(x, y)dy = \frac{\partial}{\partial x} \int_{z_0+rx}^{z+rx} f(y)dy \\ &= r(f(z+rx) - f(z_0+rx)) \end{aligned}$$

Note that defining $\zeta(x) = \frac{1}{2} - rx$, $|z_0 - \zeta(x)| \geq \beta \Rightarrow h(x, z_0) = \frac{1-\alpha}{1-2\beta}$. Clearly, however, we have $\sup_z h(x, z) > \frac{\alpha}{2\beta}$, so we have

$$|z_0 - \zeta(x)| \geq \beta \Rightarrow \sup_z \frac{\partial}{\partial x} P(x, [z_0, z]) > r \left(\frac{\alpha}{2\beta} - \frac{1-\alpha}{1-2\beta} \right).$$

We also have, however, that $\sup_z \frac{\partial}{\partial x} P(x, [z_0, z]) \geq 0$. But $|z_0 - \zeta(x)| \geq \beta$ on a set of measure $1 - 2\beta$, so we deduce

$$\delta_+ > r \left(\frac{\alpha}{2\beta} - \frac{1-\alpha}{1-2\beta} \right) (1 - 2\beta) = r \left(\frac{\alpha}{2\beta} - 1 \right).$$

Then taking $\alpha = \frac{1}{2}$, $\beta = \frac{1}{8}$ and $r = N$, we get $\delta_+ > N$. This proves the corollary. \square

The positive and negative degree are very easy to calculate, but have the drawback that they do not contain all the information which we might want. In particular, it is not clear that if a Markov chain has positive and negative degrees greater than 1 in modulus, how to see if it has a representation by a combination of degree 1 and -1 maps. The following definition remedies this at the cost of extra complexity involved in the calculation.

Definition. Let $\bar{\delta}(p) = \frac{1}{w(p)} \max(V_+(p), |V_-(p)|)$. The degree of a Markov chain \mathcal{M} on the circle with transition probability P is then defined by

$$\delta = \inf_{p_1+p_2=P} \max(\bar{\delta}(p_1), \bar{\delta}(p_2)),$$

where the infimum is taken over S -maps p_1 and p_2 of non-negative weight.

The smooth degree of a Markov chain \mathcal{M} on the circle with transition probability P is defined by

$$\delta_S = \inf_{p_1+p_2=P} \max(\bar{\delta}(p_1), \bar{\delta}(p_2)),$$

where the infimum this time is taken over smooth S -maps p_1 and p_2 of non-negative weight.

Claim. We have in fact

$$\delta = \inf_{p_1+p_2+\dots+p_n=P} \max_{1 \leq i \leq n} (\bar{\delta}(p_i)).$$

Proof. Let p_1 and p_2 be any two S -maps with non-zero weights w_1 and w_2 . Then we have

$$\begin{aligned} V_+(p_1 + p_2) &= \lim_{m \rightarrow \infty} \sup_{z_1 \dots z_m} \sum_{i=1}^m \left[(p_1 + p_2)\left(\pi\left(\frac{i}{m}\right), [z_0, z_i]\right) - (p_1 + p_2)\left(\pi\left(\frac{i-1}{m}\right), [z_0, z_i]\right) \right] \\ &\leq \lim_{m \rightarrow \infty} \sup_{z_1 \dots z_m} \sum_{i=1}^m \left[p_1\left(\pi\left(\frac{i}{m}\right), [z_0, z_i]\right) - p_1\left(\pi\left(\frac{i-1}{m}\right), [z_0, z_i]\right) \right] \\ &\quad + \lim_{m \rightarrow \infty} \sup_{z_1 \dots z_m} \sum_{i=1}^m \left[p_2\left(\pi\left(\frac{i}{m}\right), [z_0, z_i]\right) - p_2\left(\pi\left(\frac{i-1}{m}\right), [z_0, z_i]\right) \right] \\ &= V_+(p_1) + V_+(p_2). \end{aligned}$$

From this, it follows that $V_+(p_1 + p_2)/(w_1 + w_2)$ is bounded above by $(V_+(p_1) + V_+(p_2))/(w_1 + w_2)$ which is a weighted average of $V_+(p_1)/w_1$ and $V_+(p_2)/w_2$. In particular, we have $V_+(p_1 + p_2)/(w_1 + w_2) \leq \max(V_+(p_1)/w_1, V_+(p_2)/w_2)$. The corresponding result also holds for V_- : $|V_-(p_1 + p_2)|/(w_1 + w_2) \leq \max(|V_-(p_1)|/w_1, |V_-(p_2)|/w_2)$. Let $\mathcal{P}_+ = \{\text{S-maps } p : \bar{\delta}(p) = V_+(p)/w(p)\}$ and $\mathcal{P}_- = \{\text{S-maps } p : \bar{\delta}(p) = |V_-(p)|/w(p)\}$. Now suppose $p_1 + p_2 + \dots + p_n = P$. Then we may assume p_1, p_2, \dots, p_k are in \mathcal{P}_+ and p_{k+1}, \dots, p_n are in \mathcal{P}_- . In this case, we have

$$\bar{\delta}(p_1 + p_2 + \dots + p_k) \leq \max_{1 \leq i \leq k} (\bar{\delta}(p_i)),$$

and

$$\bar{\delta}(p_{k+1} + \dots + p_n) \leq \max_{k+1 \leq i \leq n} (\bar{\delta}(p_i)).$$

Thus, the claim is proved. \square

A similar statement holds for δ_S . We can then use the degree to determine whether a given Markov chain may be represented by homeomorphisms by the following.

Theorem 6. *Suppose a Markov chain \mathcal{M} on the circle has $\delta_S < N$. Then \mathcal{M} may be represented by a combination of degree N and degree $-N$ local diffeomorphisms.*

Theorem 7. *Suppose a Markov chain \mathcal{M} on the circle has $\delta > N$. Then \mathcal{M} cannot be represented by a combination of degree N and degree $-N$ local homeomorphisms.*

Proof of Theorem 6. Let \mathcal{M} be as in the statement of the theorem. Suppose \mathcal{M} has transition map P . Then by the definition of δ_S , there exist smooth S-maps p_1 and p_2 such that $\bar{\delta}(p_1) < N$ and $\bar{\delta}(p_2) < N$. Now, applying Lemma 4, and Theorem 2, we get the required result. \square

Proof of Theorem 7. Suppose a Markov chain \mathcal{M} has a representation by degree N and degree $-N$ local homeomorphisms. Let p_1 and p_2 be the S-maps associated to the subrepresentations of degree N and degree $-N$ maps. By Theorem 3, we see that $\tilde{\delta}(p_1) \leq N$ and $\tilde{\delta}(p_2) \leq N$, and hence $\delta \leq N$. This is a proof of the theorem by contradiction. \square

I conjecture that in the case of smooth Markov chains, one has that $\delta = \delta_S$.

3. A Markov Chain which cannot be Represented by Homeomorphisms

The above gives a criterion for maps to be represented by diffeomorphisms. The degree and smooth degree are however extremely unwieldy objects. In the following, we check one of its most basic properties: that there exist Markov chains with $\delta_S \geq 1$. By Theorem 6, it is sufficient to show, as we do here by ad hoc means, that there is a Markov chain which cannot be represented by a combination of orientation-preserving and orientation-reversing homeomorphisms. In this section, we modify the example of the previous section to show this.

The strategy will be to construct a Markov chain with transition map P and to show that there can be no S-map induced by a collection of degree 1 homeomorphisms of weight $\frac{1}{2}$ which is subordinate to P and the same thing for degree -1 homeomorphisms. This will then complete the proof of the theorem as, if the result did not hold, there would be a representation of the Markov chain which would be composed of degree 1 and -1 homeomorphisms. In particular, the measure of one of these subsets would have to be at least $\frac{1}{2}$, and taking the S-map induced by a subrepresentation of this would contradict the above.

In the course of the proof, we will take λ to be Haar measure on the circle. The Markov chain which we will use is that which we constructed in the Corollary above. The

parameters α, β and r are to be determined. Write P for the transition map of this Markov chain. Let $\zeta(x)$ be given by $\frac{1}{2} - rx$.

Suppose then that \mathcal{F}_+ is a collection of orientation-preserving homeomorphisms and ν_+ is a measure on them such that $\nu_+(\mathcal{F}_+) = \frac{1}{2}$ and such that the induced S-map is subordinate to P . We perform two estimates: First, fix $f \in \mathcal{F}_+$ and consider the set $\{x : |f(x) - \zeta(x)| < \beta\}$. The function $f(x) - \zeta(x)$ is monotonic of degree $r + 1$, so the set above has $r + 1$ components. Take a lift G of the function $f(x) - \zeta(x)$ and suppose $G(y) = n - \beta$ where $n \in \mathbb{N}$. Then $G(y + \frac{2\beta}{r}) > n + \beta$, so the measure of each component of the set is less than $\frac{2\beta}{r}$, so we get

$$\lambda(\{x : |f(x) - \zeta(x)| > \beta\}) > 1 - \frac{2\beta}{r}(r + 1).$$

By construction of P however, we have $P(x, S^1 \setminus [\zeta(x) - \beta, \zeta(x) + \beta]) = 1 - \alpha$, so in order for the induced S-map to be subordinate to P , we must also have

$$\nu_+(\{f \in \mathcal{F}_+ : |f(x) - \zeta(x)| > \beta\}) \leq 1 - \alpha.$$

Integrating these inequalities with respect to f and x respectively and applying Fubini's theorem, we see that for consistency, we are forced to have

$$1 - \alpha > \frac{1}{2} \left(1 - \frac{2\beta}{r}(r + 1) \right).$$

Suppose instead we have a collection \mathcal{F}_- of orientation-reversing homeomorphisms and that ν_- is a measure on them such that $\nu_-(\mathcal{F}_-) = \frac{1}{2}$ and such that the induced S-map is subordinate to P . Then, as before, we have

$$\nu_-(\{f \in \mathcal{F}_- : |f(x) - \zeta(x)| > \beta\}) \leq 1 - \alpha.$$

We will also require an estimate of the measure of the set $\{x : |f(x) - \zeta(x)| > \beta\}$ for fixed $f \in \mathcal{F}_-$. This time, $f(x) - \zeta(x)$ has degree $r - 1$, but we can no longer say that the function is monotonic. We consider a lift G of the function $f(x) - \zeta(x)$. As we noted above, this has degree $r - 1$. Then pick a point y such that $G(y) = \frac{1}{2} + \beta$ and for $1 \leq i \leq r - 1$ set

$$a_i = \sup\{z \in [y, y + 1) : G(z) = \frac{1}{2} + (i - 1) + \beta\}$$

$$b_i = \inf\{z \in [a_i, y + 1) : G(z) = \frac{1}{2} + i - \beta\}.$$

Note that $G(y + 1) = \frac{1}{2} + \beta + (r - 1)$, so that each of the above exists. We have also however that $\pi(a_i, b_i) \subset \{x : |f(x) - \zeta(x)| > \beta\}$, but $b_i - a_i > \frac{1 - 2\beta}{r}$ as $G(a_i) = \frac{1}{2} + \beta + (i - 1)$ and $G(a_i + \sigma) < G(a_i) + r\sigma$. So since the sets $\pi(a_i, b_i)$ are disjoint, we get

$$\lambda(\{x : |f(x) - \zeta(x)| > \beta\}) > (r - 1) \frac{1 - 2\beta}{r}.$$

Integrating and using Fubini's theorem as before, we find that we require for consistency that

$$1 - \alpha > \frac{1}{2}(r - 1) \frac{1 - 2\beta}{r}.$$

We may then choose α , β and r , so taking $r = 2$, $\beta = \frac{1}{4}$ and $\alpha > \frac{7}{8}$, we find that the above inequalities are not satisfied, and so we have a Markov chain which cannot be represented by homeomorphisms.

Bibliography

- [A] T. Aubin, *Nonlinear Analysis on Manifolds. Monge-Ampère Equations*, (Springer, 1982).
- [AC] L. Arnold and H. Crauel, 'Random dynamical systems', *Lyapunov Exponents, Proc. Oberwolfach 1990. Lecture Notes in Mathematics 1486*, 1–22, (Springer, 1990).
- [AMR] R. Abraham, J. E. Marsden and T. Ratiù, *Manifolds, Tensor Analysis and Applications*, (Addison-Wesley, 1983).
- [AR] R. Abraham and J. Robbin, *Transversal Mappings and Flows*, (W. A. Benjamin, 1967).
- [BB] A. Boyarsky and W. Byers, 'A graph-theoretic bound on the number of independent absolutely continuous invariant measures', *J. Math. Anal. Appl.* **139**(1989), 139–151.
- [BC] R. M. Blumenthal and H. H. Corson, 'On continuous collections of measures', *Proc. 6th Berkeley Symp. on Math. Stat. and Prob.*, v. 2 (1972), 33–40.
- [Be] H. Berbee, 'Chains with infinite connections: uniqueness and Markov representation', *Probab. Th. Rel. Fields* **76**(1987), 243–253.
- [Bi] G. Birkhoff, 'Three observations on linear algebra', *Rev. Univ. Nac. Tucuman (A)* **5**(1946), 147–151.
- [BK] M. Bramson and S. Kalikow, 'Nonuniqueness in g -functions', *Israeli J. Math.* (to appear).
- [Bol] B. Bollobás, *Graph Theory*, (Springer, 1979).
- [Bos1] C. J. Bose, 'Generalized baker's transformations', *Ergodic Theory Dynamical Systems* **9**(1989), 1–17.
- [Bos2] C. J. Bose, 'Mixing examples in the class of piecewise monotone and continuous maps

of the unit interval', *Israeli J. Math.* (to appear).

- [Bow] R. Bowen, 'A horseshoe with positive measure', *Invent. Math.* **29**(1975), 203–204.
- [BS] A. Boyarsky and M. Scarowsky, 'On a class of transformations which have unique absolutely continuous invariant measures', *Trans. Amer. Math. Soc.* **255**(1979), 243–262.
- [CFS] I. P. Cornfeld, S. V. Fomin and Ya. G. Sinai, *Ergodic Theory*, (Springer, 1982).
- [DF] W. Doeblin and R. Fortet, 'Sur les chaînes à liaisons complètes', *Bull. Soc. Math. France* **65**(1937), 132–148.
- [GS] P. Gora and B. Schmitt, 'Un exemple de transformation dilatante et C^1 par morceaux de l'intervalle, sans probabilité absolument continue invariante', *Ergodic Theory Dynamical Systems* **9**(1989), 101–113.
- [He] M. R. Herman, 'Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations', *Inst. Hautes Études Sci. Publ. Math.* **49**(1979), 5–233.
- [HM] J. M. Hammersley and J. G. Mauldon, 'General principles of antithetic variates', *Math. Proc. Camb. Phil. Soc.* **52**(1956), 476–481.
- [Hu] P. Hulse, 'Uniqueness and ergodic properties of attractive g -measures', *Ergodic Theory Dynamical Systems* **11**(1991), 65–77.
- [Kak] S. Kakutani, 'Random ergodic theorems and Markoff processes with a stable distribution', *Proc. 2nd Berkeley Symp. on Prob. and Stat.* (1950), 247–261.
- [Kal] S. Kalikow, 'Random Markov processes and uniform martingales', *Israeli J. Math.* **71**(1990), 33–54.
- [Kar] S. Karlin, 'Some random walks arising in learning models, I', *Pac. J. Math.* **3**(1953), 725–756.

- [Kea] M. Keane, 'Strongly mixing g -measures', *Invent. Math.* **16**(1974), 309–324.
- [Ken] D. G. Kendall, 'On infinite doubly-stochastic matrices and Birkhoff's Problem 111', *J. London Math. Soc. (1)* **35**(1960), 81–84.
- [Ki] Y. Kifer, *Ergodic Theory of Random Transformations*, (Birkhäuser, 1986).
- [Ko] Z. S. Kowalski, 'Invariant measures for piecewise monotonic transformations', *Proc. Winter School of Probability, Karpacz 1975. Lecture Notes in Mathematics* **472**, 77–94, (Springer, 1975).
- [Kr1] K. Krzyżewski, 'On expanding mappings', *Bull. Acad. Polon. Sci. Sér. Sci. Math.* **XIX**(1971), 23–24.
- [Kr2] K. Krzyżewski, 'A remark on expanding mappings', *Colloq. Math.* **XLI**(1979), 291–295.
- [KS] K. Krzyżewski and W. Szlenk, 'On invariant measures for expanding differentiable mappings', *Studia Math.* **XXXIII**(1969), 83–92.
- [LM] A. Lasota and M. C. Mackey, *Probabilistic Properties of Deterministic Systems*, (Cambridge University Press 1985).
- [LaY] A. Lasota and J. A. Yorke, 'On the existence of invariant measures for piecewise monotonic transformations', *Trans. Amer. Math. Soc.* **186**(1973), 481–488.
- [LiY] T. Y. Li and J. A. Yorke, 'Ergodic transformations from an interval into itself', *Trans. Amer. Math. Soc.* **235**(1978), 183–192.
- [Ma] R. Mañé, *Ergodic Theory and Differentiable Dynamics*, (Springer, 1988).
- [Mo] J. Moser, 'On the volume elements on a manifold', *Trans. Amer. Math. Soc.* **120**(1965), 286–294.
- [PP] W. Parry and M. Pollicott, *Zeta Functions and the Periodic Orbit Structure of Hy-*

perbolic Dynamics, Astérisque 187–188, (Soc. Math. de France, 1990).

- [PPW] M. R. Palmer, W. Parry and P. Walters, 'Large sets of endomorphisms and of g -measures', *The Structure of Attractors in Dynamical Systems, Lecture Notes in Mathematics 668*, 191–210, (Springer, 1978).
- [Q1] A. N. Quas, 'On representations of Markov chains by random smooth maps', *Bull. London Math. Soc.* **23**(1991), 487–492.
- [Q2] A. N. Quas, 'Invariant measures for families of circle maps', *Math. Proc. Camb. Phil. Soc.* **111**(1992), 585–597.
- [Q3] A. N. Quas, 'Representations of Markov chains on tori', *Random and Computational Dynamics* **1**(1993), 261–276.
- [Ra] M. Rahe, 'On a class of generalized baker's transformations', *Can. J. Math.* **45**(1993), 638–649.
- [Re] P. Révész, 'A probabilistic solution of Problem 111 of G. Birkhoff', *Acta Math. Acad. Sci. Hungar.* **13**(1962), 187–198.
- [S] V. N. Sudakov, 'Geometrical problems of infinite-dimensional probability distributions', *Tr. Mosk. Inst. Akad. Nauk* **141** (1976), 1–178.
- [Wa1] P. Walters, 'Ruelle's operator theorem and g -measures', *Trans. Amer. Math. Soc.* **214**(1975), 375–387.
- [Wa2] P. Walters, 'Invariant measures and equilibrium states for some mappings which expand distances', *Trans. Amer. Math. Soc.* **236**(1978), 121–153.
- [Wa3] P. Walters, *An Introduction to Ergodic Theory*, (Springer, 1981).
- [Wo] S. Wong, 'Some metric properties of piecewise monotonic mappings of the unit interval', *Trans. Amer. Math. Soc.* **246**(1978), 493–500.