An Introduction to Intermittency

Senior Honours MT5599 Project

University of
St Andrews

Tom Conti-Leslie
Supervisor: Dr Mike Todd

2020-2021

Acknowledgements

I would firstly like to thank my supervisor, Dr Mike Todd, for guiding me through my work on this project. Dr Todd is a treasure trove of references when it comes to dynamical systems and his insight has been invaluable. I complete this project with a million ideas of what to investigate next thanks to the perspective he is always able to provide during our meetings.

I would also like to thank Dr Viviane Baladi for her support of my interest in ergodic theory, and her kind invitation for me to attend the EPFL's spring school on transfer operators in March 2021, which provided me with some fascinating context that allowed me to complete some of the sections in this project.

Mathematical understanding is impossible without communication, and I am fortunate to have spent my undergraduate years in constant discussion with my close friends Sam Hatchell, Anna Wilson and Reinis Cirpons. Thank you for the endless corridor chats and all of your support. You have made me a better mathematician.

Finally I owe the biggest thanks to all my family and friends at home and in St Andrews, and especially to Sophie Purkis Charters—for standing with me through the highs and lows of university life and for believing in me.

Abstract

Chaotic dynamical systems are often not uniform, and may exhibit properties characteristic of intermittency, a qualitative phenomenon described by Pomeau and Manneville in 1980. Intermittent maps may alternate between laminar and chaotic phases, and this can quantifiably slow down their dynamical properties such as mixing speeds. In this project, we introduce two common tools that allow dynamical systems to be represented symbolically, in order to gain an understanding of the phenomena behind intermittency. We then apply these tools to the Manneville-Pomeau map, a non-uniformly expanding interval map commonly used in the literature as an example of intermittency. In the process, we point to some seminal papers and textbooks, in an effort to provide the reader with an in-depth introduction to the basics of this area of ergodic theory.

Declaration

I certify that this project report has been written by me, is a record of work carried out by me, and is essentially different from work undertaken for any other purpose or assessment.

Tom Conti-Leslie.

Contents

Chapter 1

Introduction and Preliminaries

In mathematics, dynamical systems are built in the image of physical processes that present some form of time evolution. This can either be in discrete time (modelled, for example, as iterates of a map) or continuous time (modelled perhaps using a smooth family of maps, or a system of differential equations). The general idea tends to be that dynamical systems have no memory of past states, so the state of a system at a given moment in time is enough to predict where the system could go next. Sometimes, systems may have a random component, in which case they are called *stochastic*. Otherwise, they are called *deterministic*. In this project we study discrete-time, deterministic dynamical systems.

The word *deterministic* is misleading, because in nature, physical processes are so complex that accurately measuring the state of a system at a given moment in time—including positions and velocities of particles, the effects of electric and magnetic forces, gravity, and so on—is impossible. The best we can hope to do is measure as much as possible, as accurately as we can. But after a large enough amount of time, small imprecisions in the initial measurement have consequential impacts on the accuracy of the predictions we have made about where the deterministic system will go next: this is sometimes called the *butterfly effect*. In mathematics, we speak of *chaos*.

When modelling chaotic dynamical systems via straightforward, abstract examples, it is easy to accidentally construct oversimplified models. Often, basic examples assume that these chaotic systems behave *uniformly*—that is, that they are in a sense "equally chaotic all of the time". However, this is not always a fair assumption. Pomeau and Manneville [PM80] noted that for certain well-chosen parameters, convective fluids could be observed to oscillate predictably most of the time, but occasionally exhibit turbulent, difficult-to-predict phases. In their paper, they went on to qualitatively describe this behaviour, calling it "intermittent".

The aim of this project, then, is to provide the reader with a gentle introduction to intermittency in dynamical systems from the point of view of ergodic theory, covering both the sorts of techniques and constructions that are useful, as well as an application of these techniques to one specific intermittent dynamical system.

We will spend the first chapter defining concepts and stating famous results that will assist us later, and we will also look at Pomeau and Manneville's paper in a bit more detail to determine exactly what phenomenon we are hoping to observe. In the second chapter, we will learn about two symbolic constructions that can be applied to an abstract dynamical system: Young towers and countable Markov shifts. These constructions are there to help us visualise how the dynamical system is acting on its state space, but they also give us technical information about the map if we construct them carefully. We will spend some time discussing under what conditions these symbolic representations can be faithful to the original system. Finally, in the third chapter, we will apply our knowledge to a canonical example of an intermittent discrete-time system acting on the unit interval: the *Liverani-Saussol-Vaienti map* [LSV99], usually referred to as the "Manneville-Pomeau map" for its intermittent properties. Though topologically similar to systems exhibiting uniform chaos (we might say *uniform expansion*), this map has properties which stand in stark contrast with its better-behaved cousins. It has been well-studied already, so we will be bringing together known facts and, where possible, taking the time to prove them in more detail—and sometimes with different techniques—than is usually offered in the literature.

The intended audience for this project is senior honours students with some background in analysis and measure theory, though I hope that mathematicians on either side of this level of preparation may find it useful—interpretation of the more technical concepts and derivations is offered where possible, in order to avoid the work being too tied to a specific level. In writing this report, I have drawn on my knowledge of some honours level modules taken at the University of St Andrews which have provided useful context. To give an idea of the prerequisites, and also to point a keen reader towards modules that may be of interest, here is a list of them.

- MT4508 Dynamical Systems (covers some interval maps, Lyapunov exponents, etc—though we will not be too interested in fixed points, bifurcations, and stable/unstable manifolds in this project);
- MT4528 Markov Chains and Processes (this module is good preparation for studying Markov shifts, but this is not essential);
- MT5862 Measure Theory (most results are stated in a measure-theoretic setting in this project, so it is useful to have seen measure theory beforehand, although we recap the basics in section 1.1);
- MT5877 Ergodic Theory and Dynamical Systems (we go over most of the necessary concepts from this module in the introduction, and the bulk of the project goes beyond what was covered in MT5877, but it is definitely useful preparation).

We adopt standard mathematical notation throughout the project. If referring to a mathematical object defined by a tuple e.g. (A, B, C), we may sometimes write "(A, C)", "(B, A)", "A", etc. depending on what is known and what is relevant. We use a vertical bar e.g. "$A|_B$" to denote the restriction of a function or collection A to a set B. For two functions $f : X \to X$ and $g : X \to \mathbb{Z}$, we denote by f^g the function given by $f^g(x) = f^{g(x)}(x)$. For partitions A and B of a set, their *join* (the smallest partition C such that every element of A and every element of B is partitioned by elements of C) is denoted $A \vee B$. In general when working with an underlying measure μ, statements hold μ-almost everywhere by default, and this includes equality of sets (though we usually try to be clear when this is the case). The symbol $\mathring{=}$ will occasionally be used to denote almost-everywhere equality for sets, partitions, and algebras.

Most plots in this project are straightforward in the sense that the reader shouldn't have any trouble replicating the figure using a standard graphing package (such as Desmos). However, in a few cases, some additional steps have been carried out in the background in order to computationally generate an image. In the interest of transparency, the Python scripts used to do this are available on GitHub at `tomcontileslie/manneville-pomeau`. Wherever some computation has happened behind the scenes, the margin will be marked with a calculator symbol 🖩.

Where one calculation or part of a proof requires some careful thought to be justified, I have marked the margin with a Bourbaki dangerous bend sign ☡.

With all this in mind, let's get started!

1.1 Measure theory

In the interest of having all of our definitions in one place, let's quickly give an overview of measure theory. Much more detailed introductions are readily available—for example, in [Bar95]. The content of this section is adapted from MT5862.

Definition 1.1 (Measurable Space, Measure, Measure Space)**.** Let X be a set and $\mathcal{E}$ be a set of subsets of X. We call $\mathcal{E}$ a *σ-algebra* if it contains $\emptyset$ and X and is closed under complements and countable unions. We refer to $(X, \mathcal{E})$ as a *measurable space.*

Let $\mu : \mathcal{E} \to [0, \infty]$ be a function that we think of as mapping a subset of X to its "size" or "mass". Then we call μ a *measure* if $\mu(\emptyset) = 0$ and if μ is countably additive on disjoint measurable sets (i.e. if $E_1, E_2, \dots \in \mathcal{E}$ pairwise disjoint, then $\mu(\cup_n E_n) = \sum_n \mu(E_n)$). If $\mu(X) = 1$, then we call μ a *probability measure.*

We then refer to $(X, \mathcal{E}, \mu)$ as a *measure space*, or a *probability space* if μ is a probability measure.

Measure theory is the rigorous underpinning for the modern theory of integration, and is a robust model since we are allowed to take countable combinations of things. Usually properties that hold countably are referred to using a "σ", one important example being that we call a measure *σ-finite* if X can be partitioned into a countable union of sets with finite measure.

We can also use this theory to help us decide what subsets of X "matter". Usually, if a set has measure 0, it is not significant enough that its properties will have an effect on the overall set X from the point of view of μ. We call such sets *null sets.*

If a property is true for every point in X except for those lying inside some null set with respect to μ, we say that the property holds *μ-almost everywhere* or *μ-a.e.*; simply *a.e.* if the underlying measure is clear. Note, however, that a.e. statements are highly dependent on which measure we are using, and measures may have wildly different null sets (in the most extreme case, this means they are *mutually singular*). We will need to be careful about this when claiming that a statement is true almost everywhere.

In terms of measures, many weird and wonderful measures can be defined on a σ-algebra. We will typically only use Lebesgue measure on Borel sets on an interval. In this case, X is taken to be an interval $[a, b] \subseteq \mathbb{R}$; the σ-algebra $\mathcal{E}$ equals the Borel algebra $\mathcal{B}(X)$ (the algebra generated from all the open subsets of $[a, b]$, or equivalently all open intervals); and the measure μ equals Lebesgue measure λ, the measure which assigns to an interval (a, b) its natural notion of "width", $\lambda((a, b)) = |b - a|$.[1]

One other measure that we should be aware of, at least for the sake of counterexamples, is the *Dirac measure*: for a measurable space $(X, \mathcal{E})$ and $x \in X$, define $\forall E \in \mathcal{E} : \delta_x(E) = \chi_E(x)$ where χ_E is the characteristic function on E. This is a probability measure.

[1] Defining λ on intervals then uniquely determines its value on all Borel sets.

For the often-used Borel algebra $\mathcal{B}(X)$, we *generate* a σ-algebra based on all combinations of open sets in X. In general, if we have a collection of sets $\mathcal{D}$, the σ-algebra that it generates is $\sigma(\mathcal{D})$ (the smallest σ-algebra containing all the sets in $\mathcal{D}$).

The work in later chapters often involves shifting our frame of reference to a subset of the space, so we should also note that σ-algebras can easily be restricted:

Definition 1.2 (Restriction of a Measurable Space)**.** Let $(X, \mathcal{E})$ be a measurable space and $A \in \mathcal{E}$. Then the *restriction of* $(X, \mathcal{E})$ *to* A is the measurable space $(A, \mathcal{E}|_A)$ where

$$\mathcal{E}|_A = \{A \cap E : E \in \mathcal{E}\}.$$ [2]

The presence of sets lying outside $\mathcal{E}$—*non-measurable sets*—means that we must, at least in our theoretical setup, be careful not to fall outside of the measurable σ-algebra at any point. Therefore, in the literature, it usually goes without saying that the functions we study must be compatible with the structures they map between.

Definition 1.3 (Measurable Function)**.** Let $(X, \mathcal{E})$ and $(Y, \mathcal{F})$ be measurable spaces. A function $f : X \to Y$ is said to be *$\mathcal{E}$-$\mathcal{F}$-measurable* or simply *measurable* if $\forall F \in \mathcal{F} : f^{-1}F \in \mathcal{E}$.

If Y is a subset of $\mathbb{R}$ and $\mathcal{F}$ is not specified, then we call f measurable if it is $\mathcal{E}$-$\mathcal{B}(Y)$-measurable (i.e. measurable with respect to the Borel algebra).

In practice, non-measurability is hardly ever an issue. In the Borel algebra, for example, one has to try very hard to find a set that is not measurable (a common example is the Vitali set, but its construction requires the axiom of choice). Certainly when our function is defined in standard calculus terms, we may continue our work undisturbed by the definition above.

Our next result allows us to create new measures from pre-existing ones.

Proposition 1.4 (Standard Constructions of Measures)**.**

1. *(Linear Combinations of Measures). Let $(X, \mathcal{E})$ be a measurable space with measures $\mu_1, \mu_2, \ldots$. Let $a_1, a_2, \cdots \geq 0$. Then $\sum_n a_n \mu_n$ given by $\forall E \in \mathcal{E} : \left(\sum_n a_n \mu_n\right)(E) = \sum_n a_n \mu_n(E)$ is a measure.*

2. *(Restriction of a Measure). Let $(X, \mathcal{E}, \mu)$ be a measure space and let $A \in \mathcal{E}$. Then $\mu|_A$ given by $\forall E \in \mathcal{E} : \mu|_A(E) = \mu(E \cap A)$ is a measure.*

3. *(Pre-Composition of a Measure). Let $(X, \mathcal{E}, \mu)$ be a measure space and let $(Y, \mathcal{F})$ be a measurable space. Let $f : X \to Y$ be a measurable function. Then $\mu \circ f^{-1}$ given by $\forall F \in \mathcal{F} : (\mu \circ f^{-1})(F) = \mu(f^{-1}F)$ is a measure on $(Y, \mathcal{F})$.*[3]

These constructions will all prove useful in the next chapter.

Remark 1.5. In some definitions of a restriction, the measure $\mu|_E$ is "normalised" by dividing through by $\mu(E)$, giving a probability measure. This will not be hugely useful to us, since we will typically be restricting Lebesgue measure—in which case normalising would mean that the restricted measure of an interval does not equal its size. So we will opt to only normalise when necessary, and do so explicitly. This will be done notationally with a bar, e.g. $\bar{\mu} = \mu/\mu(X)$ where X is the set whose algebra μ is defined on.

---

[2] It is easy to check that $\mathcal{E}|_A$ is a σ-algebra, and that the restriction can alternatively be expressed as $\mathcal{E}|_A = \mathcal{E} \cap \mathscr{P}(A)$.

[3] This pre-composition is sometimes written $f_*\mu$.

Finally let us end this brief introduction with a reminder on Lebesgue integration.

Definition 1.6 (Lebesgue Integral)**.** Let $(X, \mathcal{E}, \mu)$ be a measure space. The *Lebesgue integral* with respect to μ of a measurable function $f : X \to \mathbb{R}$ is a notion built up from integrals of characteristic functions. For any $E \in \mathcal{E}$, we impose

$$\int \chi_E \, d\mu = \mu(E),$$

and we then extend the notion so that the integral of a finite linear combination of measurable characteristic functions is the linear combination of the integrals (these functions are called *simple functions*). Finally, the Lebesgue integral of an arbitrary measurable function f is the supremum over all simple functions less than f of the Lebesgue integrals of each of these simple functions.

The Lebesgue integral is implicitly a definite integral over the entire space X. This space can be chosen to be e.g. a subset of $\mathbb{R}^n$ with Lebesgue measure, which allows us to recover a notion of definite integration as seen in calculus. If we want to integrate over a set A smaller than X, we write $\int_A f \, d\mu := \int f \cdot \chi_A \, d\mu$.

1.2 Ergodic theory and dynamical systems

The content of this section is mostly taken from [Tod20].

Definition 1.7 (Dynamical System)**.** Let $(X, \mathcal{E}, \mu)$ be a measure space and let $f : X \to X$ be a measurable function.[4] Then $(X, \mathcal{E}, \mu, f)$ is referred to as a *dynamical system*. The set X is referred to as the *state space* of the system. The (discrete) time evolution of the system is given by the repeated application of f to a given point $x \in X$. The values taken by the iterates of f applied to x give the *orbit* of x: $O(x) = \{x, f(x), f^2(x), \dots\}$. We may also refer to the dynamical system as (X, f), depending on what is known.

In some sources, dynamical systems are referred to as *transformations*.

We always use powers of a function to denote its composition with itself, rather than the value of that function raised to a power. For $k \geq 0$, we write:

$$f^k(x) = \underbrace{f(f \dots f(f}_{k \text{ times}}(x)) \dots); \quad f^{-k}(x) = \underbrace{f^{-1}(f^{-1} \dots f^{-1}(f^{-1}}_{k \text{ times}}(x)) \dots); \quad f^0(x) = x.$$

In dynamical systems, the hope is usually that the function f plays nicely with the properties of the measure space it acts on. One fairly weak condition we might impose is that f in some sense "disregards" the null sets of μ. Namely, f should map null sets to null sets.

Definition 1.8 (Nonsingularlity)**.** A dynamical system $(X, \mathcal{E}, \mu, f)$ is said to be *nonsingular* if:

$$\forall E \in \mathcal{E} : \mu(E) > 0 \implies \mu(f^{-1}E) > 0.$$

(It is a common pattern that dynamical properties are phrased using preimages rather than images).

Stronger properties of compatibility between the measure space $(X, \mathcal{E}, \mu)$ and the dynamics f are also desirable. One strengthened version of the above condition is the standard starting point when proving properties of a system. It states that the underlying measure, μ, assigns the same mass to each set as it does to the set of all points whose f-images lie in that set. Formally, we have:

[4] We sometimes write $f : X \circlearrowleft$ since the map is from the space X back into itself.

Definition 1.9 (Invariance)**.** Given a dynamical system $f : X \to X$, we say f is *μ-invariant* if

$$\forall E \in \mathcal{E} : \mu(f^{-1}E) = \mu(E).$$

In this case, the tuple $(X, \mathcal{E}, \mu, f)$ is called a *measure preserving transformation* or *mpt*. If $\mu(X) = 1$, then we call it a *probability preserving transformation* or *ppt*.

Example 1.10 (Full Shift)**.** To illustrate dynamical systems and invariance, and to arm ourselves with an example whose generalisation will be useful in the next section, let us study the *full shift*, with the notation of [Tod20].

Let $N \geq 2$, and consider the set Σ_N^+ of all infinite sequences of non-negative integers $< N$:

$$\Sigma_N^+ = \{\boldsymbol{x} = (x_0, x_1, x_2, \dots) : 0 \leq x_i < N\}.$$

It's not immediately clear how to find reasonable subsets of Σ_N^+ that will give us a good σ-algebra, but note that we might consider two sequences $\boldsymbol{x}$ and $\boldsymbol{y}$ to be "close together" if their first few terms are the same. This is loosely what leads us to the following definition.

For any finite sequence $\boldsymbol{w} = (w_0, w_1, \dots, w_{n-1})$ with $0 \leq w_i < N$, the *cylinder set* containing $\boldsymbol{w} = (w_0, w_1, \dots, w_{n-1})$ contains all sequences starting with $\boldsymbol{w}$.[5] That is,

$$[\boldsymbol{w}] = [w_0, w_1, \dots w_{n-1}] = \{(x_0, x_1, x_2, \dots) \in \Sigma_N^+ : x_i = w_i \ \forall i = 0, 1, \dots, n-1\}.$$

Cylinder sets exist for each $n \geq 1$; we often call n the *depth*. For fixed n, the n-cylinder sets partition Σ_N^+; e.g. at level 1, we have $\Sigma_N^+ = [0] \cup [1] \cup \dots \cup [N-1]$ (because every sequence must start with *some* number, so must lie in one of these cylinders). We place ourselves in the σ-algebra generated by the cylinders for the rest of this example (usually we don't need to think too hard about the underlying algebra, but it's good to know that there is one and that the sets we might want to measure—the cylinder sets—are measurable).

We now define the "shift" dynamics σ on this space:[6]

$$\begin{array}{rccc} \sigma : & \Sigma_N^+ & \longrightarrow & \Sigma_N^+ \\ & (x_0, x_1, x_2, \dots) & \longmapsto & (x_1, x_2, x_3, \dots). \end{array}$$

We can show that for any probability vector $p = (p_0, p_1, \dots, p_{N-1})$ (i.e. the p_i are non-negative and sum to 1), a reasonable measure for Σ_N^+ could be to assign the following mass to cylinders:

$$\mu_p([w_0, w_1, \dots, w_{n-1}]) = p_{w_0} p_{w_1} \dots p_{w_{n-1}}.$$

This uniquely determines a measure called a *Bernoulli measure* that is shift-invariant (this was proved in [Tod20]). We'll do something similar for a generalised case in section 1.3.

---

[5] In general, the word *cylinder* is used in dynamical systems to denote a set whose elements "agree" for the first n steps of a process. Here the process is the left shift; later in this project we will also talk about cylinders with respect to other sorts of maps.

[6] This is different to the "σ" of "σ-algebra", but little confusion is possible.

Let's return to the general setting. When we have invariance as defined in Definition 1.9, we are able to make strong claims about the orbit of a "typical" starting point in the state space X (and by "typical", I mean we will be searching for results valid μ-almost everywhere). This is a point where ergodic theory differs from other areas of the study of dynamical systems. To wit, some mathematicians may be concerned with the rare (usually finite) set of points that have very specific orbits under f (perhaps they are constant, or periodic, orbits); on the other hand, the ergodic theorist is mostly interested in making statistical claims about what would likely happen if we took the orbit $O(x)$ of a point x picked at random in the space X (at random, that is, with respect to μ). Poincaré's Recurrence Theorem is an example of such a claim.

Theorem 1.11 (Poincaré's Recurrence Theorem)**.** *Let $(X, \mathcal{E}, \mu, f)$ be a ppt. Then for all $E \in \mathcal{E}$, we have that μ-almost every point $x \in X$ gives an orbit $O(x)$ which returns to E infinitely many times (i.e. there is a sequence $0 \leq n_1 < n_2 < \dots$ such that $f^{n_k}(x) \in E$ for all k).*

Remark 1.12. Ergodic theorems will often begin by assuming that we have a mpt or ppt. However, if we are trying to apply theoretical results to a concrete example of a dynamical system (X, f), we may be armed with a standard measure space—say, $(X, \mathcal{B}(X), \lambda)$—which has a non-invariant measure with respect to f. One of the initial challenges, then, is to find an invariant measure. Invariant measures can usually be found (for example, if $x \in X$ is a fixed point, i.e. $f(x) = x$, then δ_x is invariant). However, the "almost everywhere" statements we get for these measures using theorems such as Poincaré's Recurrence Theorem may in general mean nothing to us, if the invariant measures have drastically different null sets to our reference measure (usually λ). For example, if a statement is valid for δ_x-almost every $y \in X$, it could in fact be valid only for $y = x$, which tells us nothing much at all. So, the harder challenge is not only finding an invariant measure, but finding a meaningful one. This is the subject of the first few sections of the third chapter, when we are faced with a concrete map.

For the time being we may continue our illusion of everything being nice and already existing. There are two further refinements of the notion of invariance which imply famous and meaningful results that we should hope to harness when working with concrete maps.

Definition 1.13 (Ergodicity)**.** Let $(X, \mathcal{E}, \mu, f)$ be a ppt. We say μ is *ergodic* with respect to f if

$$\forall E \in \mathcal{E} : \ f^{-1}E = E \implies \mu(E) = 0 \text{ or } 1.$$

This condition says that if a set is invariant under the dynamics, then it is either very large or very small. Informally, we can interpret this as saying that f moves the space around a lot. Indeed, X (the whole space) is always f-invariant; if we can find a smaller set $A \subseteq X$ with measure between 0 and 1 that is also invariant, we would have found a part of the state space that acts independently of the rest.

Provided we have ergodicity, we can apply the following powerful theorem.

Theorem 1.14 (Birkhoff's Ergodic Theorem)**.** *Suppose $(X, \mathcal{E}, \mu, f)$ is an ergodic ppt. Let $\varphi \in L(\mu)$ (i.e. $\varphi : X \to \mathbb{R}$ such that $\int |\varphi| \, d\mu < \infty$). Then for μ-a.e. $x \in X$:*

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x)) \longrightarrow \int \varphi \, d\mu$$

as $n \to \infty$.

A common interpretation of this theorem is to say that "time averages converge to space averages": if we sample the value of φ at points along the orbit of x (time average), in the long run, we expect to get a representative sample of φ across the whole space (space average)—all of this, of course, with respect to the ergodic measure μ.

We spoke about ergodicity implying that f moves the space around a lot. A stronger description still would be to say that f mixes points very well (so that applying f many times to points removes any initial correlation they may have had). This is the motivation for the final definition we will need.

Definition 1.15 (Mixing)**.** Let $(X, \mathcal{E}, \mu, f)$ be a ppt. We say that the system is *(strongly) mixing* if:

$$\forall A, B \in \mathcal{E} : \ \mu(f^{-n} A \cap B) \longrightarrow \mu(A)\mu(B)$$

as $n \to \infty$. We say that the system is *weakly mixing* if:

$$\forall A, B \in \mathcal{E} : \ \frac{1}{n} \sum_{k=0}^{n-1} \left| \mu(f^{-n} A \cap B) - \mu(A)\mu(B) \right| \longrightarrow 0$$

as $n \to \infty$.

We have progressively increased the restrictiveness of the ergodic properties in this section, so it would only take a small amount of additional work to prove the following chain of implications:

$$\begin{array}{c} \text{Strong mixing} \\ \Downarrow \\ \text{Weak mixing} \\ \Downarrow \\ \text{Ergodicity} \\ \Downarrow \\ \text{Invariance} \\ \Downarrow \\ \text{Nonsingularity} \end{array}$$

1.3 Countable Markov shifts

The shift map we saw in the previous section generalises nicely to types of systems reminiscent of Markov chains. A (discrete-time) Markov chain, broadly speaking, is a sequence of random variables $\{X_0, X_1, X_2, \dots\}$ all taking values in a certain set S, such that the conditional probability distribution of X_{i+1} given X_i is the same as given $X_i, X_{i-1}, X_{i-2}, \dots, X_0$. In other words, the system changes state at every timestep i based on a probability distribution informed only by its state at the previous timestep: the system has no memory of its past states except its most recent one.

We often represent Markov chains using a *transition graph* which displays the probability of transition from a state s to another state t, for all $s, t \in S$.

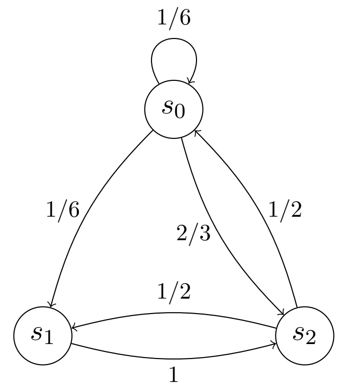


Figure 1.1: A transition graph.

In the graph above, if we impose $X_0 = s_2$, we can calculate that the probability of seeing the sequence $(X_0, X_1, X_2, X_3, X_4) = (s_2, s_1, s_2, s_0, s_0)$ is the product of the probabilities along the path $s_2 s_1 s_2 s_0 s_0$, i.e.

$$\frac{1}{2} \cdot 1 \cdot \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{24}.$$

To study Markov *shifts*, we change our perspective from looking at probabilistic transitions on a graph to looking at infinite walks on this same graph.

Each possible event for the above Markov chain is an infinite sequence of values taken by the Markov variables: $(X_0, X_1, X_2, \dots) = (x_0, x_1, x_2, \dots)$ where the transition from each x_i to x_{i+1} happened with a certain non-zero probability. Initially, we don't mind what the probability of each transition was: we are simply interested in characterising the set of all sequences that could have occurred.

The following definitions are taken from Sarig's introduction in [Sar99].

Definition 1.16 (Topological Transition Matrix)**.** Let $S = \{s_0, s_1, \dots\}$ be a finite or countable set of *states*, and let $A = (a_{ij})$ be an $S \times S$ matrix of zeroes and ones. (Interpreting A using the context above, the link is that a transition from state i to state j is possible if $a_{ij} = 1$). A is called a *topological transition matrix* if transition from and to every state is possible:

$$\forall s \in S : \exists i, j \in S : a_{si} = a_{js} = 1.$$

The matrix A determines which states can come directly before other states, and is equal to the adjacency matrix of the (unweighted) transition graph we saw above. We say that A is *topological* because rather than giving the exact transition probabilities, it only has $a_{ij} \in \{0, 1\}$, simply determining if a transition from i to j has non-zero probability, regardless of the exact probability of transition. Two systems with the same possible transitions, but different probabilities, will still be topologically equivalent.

We can now construct the set of all possible infinite paths:

Definition 1.17 (Countable Markov Shift)**.** For a topological transition matrix A, we define the associated *shift space* to be:

$$\Sigma_A^+ = \{\boldsymbol{x} = (x_0, x_1, x_2, \dots) \in S^{\mathbb{N}_0} : a_{x_i x_{i+1}} = 1 \, \forall i \geq 0\}.$$

We apply the following *one-sided shift* or *left shift* to the space:

$$\begin{array}{rccc}\sigma: & \Sigma_A^+ & \longrightarrow & \Sigma_A^+ \\ & (x_0, x_1, x_2, \dots) & \longmapsto & (x_1, x_2, x_3, \dots).\end{array}$$

The dynamical system (Σ_A^+, σ) is then referred to as the (one-sided) *countable Markov shift* (or *CMS*) generated by A.[7]

Applying the shift map to an infinite path in Σ_A^+ returns the same path, but starting one step later. We will usually wish to endow this CMS with the following metric: pick a $\theta \in (0,1)$, then for any $\boldsymbol{x} = (x_0, x_1, x_2, \dots), \boldsymbol{y} = (y_0, y_1, y_2, \dots) \in \Sigma_A^+$, define

$$d_\theta(\boldsymbol{x}, \boldsymbol{y}) = \theta^{\min\{i:\, x_i \neq y_i\}}.$$

One can show that the open sets in Σ_A^+ are precisely the cylinder sets we are about to define.

Definition 1.18. For a finite word $\boldsymbol{w} = (w_0, w_1, \dots, w_{n-1})$ over the alphabet S, the *cylinder set* containing $\boldsymbol{w}$ is

$$[\boldsymbol{w}] = [w_0, w_1, \dots, w_{n-1}] = \{(x_0, x_1, x_2, \dots) \in \Sigma_A^+ :\ x_i = w_i\ \forall i = 0, 1, \dots, n-1\}.$$

If a cylinder $[\boldsymbol{w}]$ is empty for some word $\boldsymbol{w}$, then that means one of the state transitions in $\boldsymbol{w}$ is impossible. In this case we say that $\boldsymbol{w}$ is *not admissible*. Otherwise, we say $\boldsymbol{w}$ is an *admissible* word.

The Borel σ-algebra generated by the open sets in Σ_A^+ will be denoted $\mathcal{B}$ and $(\Sigma_A^+, \mathcal{B})$ will be our measurable space of choice. We now want to characterise the σ-invariant measures on $(\Sigma_A^+, \mathcal{B})$, and will exhibit a class of these measures below. The direction taken for the rest of this section is inspired by [Aar97, Chapter 4], but I have combined the statements with some notions from Markov chains to reach the invariant measures faster while also giving more detailed proofs.

Definition 1.19 (Stochastic Matrix)**.** Let $A = (a_{ij})$ be a topological transition matrix on the set of states S. Let $P = (p_{ij})$ be a matrix with the same dimensions as A, satisfying:

1. $p_{ij} = 0 \iff a_{ij} = 0$;
2. For each row i, $\sum_j p_{ij} = 1$.

Then P is called a *stochastic matrix* for A. As a transition matrix, it gives a Markov chain with the same graph as the one associated with A.

In certain cases, the Markov chain associated with P will converge in probability to a stable distribution, as defined below.

Definition 1.20 (Stationary Probability Vector)**.** A (row) vector $\pi = (\pi_i)_{i \in S}$ is called a *stationary probability vector* for the stochastic matrix P if:

1. $\sum_i \pi_i = 1$;

[7] Every point in Σ_A^+ can also be seen as a point in the full shift space $\Sigma_{|S|}^+$ from example 1.10, leading some sources to refer to CMS as *subshifts*.

2. $\pi P = \pi$.

Given a stochastic matrix and its associated stationary probability vector (if it exists), we can now define a shift-invariant measure.

Definition 1.21 (Markov Measure)**.** Let (Σ_A^+, σ) be a CMS. Let $P = (p_{ij})$ be a stochastic matrix for A, and suppose P has a stationary probability vector $\pi = (\pi_i)$. We define a measure $\mu_{P,\pi}$ on the cylinder sets as follows:

$$\mu_{P,\pi}([w_0, w_1, \dots, w_{n-1}]) = \pi_{w_0} p_{w_0 w_1} p_{w_1 w_2} \dots p_{w_{n-2} w_{n-1}}.$$

Measures constructed in this way are referred to as *Markov measures*. Note that they are a generalisation of the Bernoulli measures we briefly mentioned in example 1.10.

The measure $\mu_{P,\pi}$ is indeed a probability measure on the algebra of all finite unions of cylinders. Once we've shown this, and once we've convinced ourselves that the set of cylinders generates $\mathcal{B}$, the fact that it extends to a unique measure on $(\Sigma_A^+, \mathcal{B})$ follows from Carathéodory's Extension Theorem and Hahn's Extension Theorem (for more information see, for example, [Bar95, Chapter 9]).

We need to show that our measure is well-defined and that it gives full measure to the whole space. Suppose $S = \{s_0, s_1, \dots\}$.

- Take any cylinder $[w_0, w_1, \dots, w_{n-1}]$. To show well-definedness it is sufficient to show that the measure is appropriately additive on the natural partition of this cylinder:

$$\begin{aligned}
&\mu_{P,\pi}([w_0, w_1, \dots, w_{n-1}, s_0] \cup [w_0, w_1, \dots, w_{n-1}, s_1] \cup \dots) \\
&= \mu_{P,\pi}([w_0, w_1, \dots, w_{n-1}]) \\
&= \pi_{w_0} p_{w_0 w_1} \dots p_{w_{n-2} w_{n-1}} \\
&= \pi_{w_0} p_{w_0 w_1} \dots p_{w_{n-2} w_{n-1}} (p_{w_{n-1} s_0} + p_{w_{n-1} s_1} + \dots) \\
&= \mu_{P,\pi}([w_0, w_1, \dots, w_{n-1}, s_0]) + \mu_{P,\pi}([w_0, w_1, \dots, w_{n-1}, s_1]) + \dots,
\end{aligned}$$

  summing to 1 along row w_{n-1} since P is stochastic. So $\mu_{P,\pi}$ does indeed break down appropriately into sub-cylinders.

- To check $\mu_{P,\pi}$ is a probability measure, consider:

$$\begin{aligned}
\mu_{P,\pi}(\Sigma_A^+) &= \mu_{P,\pi}([s_0] \cup [s_1] \cup \dots) \\
&= \mu_{P,\pi}([s_0]) + \mu_{P,\pi}([s_1]) + \dots \\
&= \sum_i \pi_i \\
&= 1,
\end{aligned}$$

  as required.

So far we have not used the fact that π_i is stationary (in fact we can define Markov measures without the stationary property on the probability vector π, in which case π is called the *initial distribution*). However in this case, because of the stationary property, $\mu_{P,\pi}$ is invariant under the left shift.

Proposition 1.22. *Let (Σ_A^+, σ) be a CMS. Let P be a stochastic matrix for A with a stationary probability vector π. Then $\mu_{P,\pi}$ is σ-invariant.*

Proof. Since the cylinders generate $\mathcal{B}$, it is sufficient to show $\mu_{P,\pi}(\sigma^{-1}[\boldsymbol{w}]) = \mu_{P,\pi}([\boldsymbol{w}])$ for any cylinder $[\boldsymbol{w}] = [w_0, w_1, \ldots, w_{n-1}]$. We have

$$\begin{aligned}
\mu_{P,\pi}(\sigma^{-1}[\boldsymbol{w}]) &= \mu_{P,\pi}([s_0, w_0, w_1, \ldots, w_{n-1}] \cup [s_1, w_0, w_1, \ldots, w_{n-1}] \cup \ldots) \\
&= \sum_i \mu_{P,\pi}([s_i, w_0, w_1, \ldots, w_{n-1}]) \\
&= \sum_i \pi_{s_i} p_{s_i w_0} p_{w_0 w_1} \ldots p_{w_{n-2} w_{n-1}} \\
&= \left(\sum_i \pi_{s_i} p_{s_i w_0} \right) p_{w_0 w_1} \ldots p_{w_{n-2} w_{n-1}} \\
&= \pi_{w_0} p_{w_0 w_1} \ldots p_{w_{n-1} w_{n-1}} \\
&= \mu_{P,\pi}([\boldsymbol{w}]).
\end{aligned}$$

□

This proposition shows that there are potentially many measures which are invariant under a countable Markov shift; indeed, the Perron-Frobenius theorem tells us that there exists a (unique) stationary probability vector for every *finite, topologically transitive* stochastic matrix P (see, for example, [SC97]). Furthermore, Markov measures are not necessarily the only types of measures that exist on a CMS.

Typically, amongst all these invariant measures, only a few will be of interest to us: recall the need for an invariant measure to be "meaningful", as in Remark 1.12. In the next section we return to the general case and settle on a notion of "meaningful" which will allow us to single out the important measures.

1.4 Absolute continuity

Suppose we have a dynamical system $(X, \mathcal{E}, \mu, f)$. We noted in Remark 1.12 that in order to apply ergodic theorems, we need an invariant (ideally ergodic) measure. However, amongst many possible invariant measures, we need to find one whose "almost everywhere" statements translate back into our frame of reference, with our measure μ. One way of doing this is to impose that the null sets of μ are also null sets for the invariant measure.

Definition 1.23 (Absolute Continuity)**.** Given two measures μ, ν on a measurable space $(X, \mathcal{E})$, we say that ν is *absolutely continuous* (or *a.c.*) with respect to μ, and write $\nu \ll \mu$, if

$$\forall E \in \mathcal{E} : \nu(E) > 0 \implies \mu(E) > 0.$$

Searching for a.c. invariant measures can be difficult, but finding one ensures that the conclusions we can draw are meaningful with respect to the reference measure μ.

In the case of interval maps, μ will often be taken to be Lebesgue measure λ, and any corresponding probability measure ν will be called an *absolutely continuous invariant probability measure*, or *acip*.

An a.c. measure can be seen as a distortion of the original measure, as this next theorem (mentioned in most ergodic theory textbooks, for example [BG97, Chapter 2]) shows.

Theorem 1.24 (Radon-Nikodym [Nik30])**.** *Suppose μ, ν are two finite*[8] *measures on $(X, \mathcal{E})$ such that $\nu \ll \mu$. Then there exists an $L^1(\mu)$ density function $\psi : X \to [0, \infty)$ such that*

$$\forall E \in \mathcal{E} : \nu(E) = \int_E \psi \, d\mu.$$

This function is called the Radon-Nikodym derivative *of ν with respect to μ and is sometimes denoted*

$$\psi = \frac{d\nu}{d\mu}.$$

Conversely, a measure ν can be defined via its Radon-Nikodym derivative ψ with respect to μ, and in some sources this is written as $\nu := \psi\mu$, or $d\nu := \psi \, d\mu$.

This derivative then behaves exactly as we would expect when it comes to linear combinations of measures. Furthermore, integrals with respect to ν can be converted into integrals with respect to the reference measure μ using a sort of measure-theoretic chain rule:

$$\int \varphi \, d\nu = \int \varphi \cdot \frac{d\nu}{d\mu} \, d\mu. \tag{1.1}$$

There are other ways to single out particular types of meaningful invariant measures which we will not look at here. One modern field of study called *thermodynamic formalism* involves the hunt for, and description of, invariant measures that satisfy some relationship with a function $\phi : X \to \mathbb{R}$ called the *potential*. These meaningful measures include *Gibbs measures*, *conformal measures* and *equilibrium states* (all invariant measures satisfying some condition involving ϕ). I refer the interested reader to the introduction of [Sar99], which defines the key concepts and applies them to countable Markov shifts (which we defined in the previous section).

1.5 Push-forwards and the transfer operator

A dynamical system moves one timestep ahead by applying a function f to X. We can define similar "one-timestep" functions that can be applied to ergodic objects other than the state space—namely, measures and Radon-Nikodym derivatives.

Definition 1.25 (Push-Forward Measure)**.** Let $(X, \mathcal{E}, \mu, f)$ be a dynamical system. Then the *push-forward measure* of the system is the pre-composition (see Proposition 1.4)

$$f_*\mu := \mu \circ f^{-1},$$

i.e. $f_*\mu(E) = \mu(f^{-1}E)$.

Just like we can iterate f on X, we can iterate f_* on the space of all measures on X. Clearly $(f_*)^n\mu = (f^n)_*\mu$. One other nice property of push-forwards is that for $\varphi \in L(\mu)$,

$$\int \varphi \, d(f_*\mu) = \int \varphi \circ f \, d\mu. \tag{1.2}$$

This notion of moving objects through time is also available in a functional-theoretic version. The rest of this section is from [BG97, Chapter 4] and [Aar97, Chapter 1].

[8] This can be loosened to *σ-finite*, but we lose the L^1 result, and we won't need a statement stronger than finite measures here.

Definition 1.26 (Transfer Operator)**.** Let $(X, \mathcal{E}, \mu, f)$ be a nonsingular dynamical system on a finite measure space. The *transfer operator*, or *Frobenius-Perron operator*, or *dual operator* associated with f is the operator $\mathcal{L}_f : L^1(\mu) \to L^1(\mu)$ which maps $\varphi \in L^1(\mu)$ to the unique (up to μ-a.e. equality)[9] function $\mathcal{L}_f\varphi$ such that

$$\forall \psi \in L^\infty(\mu) : \int \mathcal{L}_f\varphi \cdot \psi \, d\mu = \int \varphi \cdot \psi \circ f \, d\mu.$$

(We will use $\mathcal{L}$ for transfer operators in this project, and use L^p or $\mathfrak{L}^p$ instead to denote classical function spaces.) Sources will use wildly different notations for the transfer operator, and these may include P_f or $\hat{f}$. If it is clear what the function f is, we can just write $\mathcal{L}$. Note that there is an implicit dependence on the measure μ, which is in this case taken to be a reference measure such as Lebesgue (there is no invariance assumption on μ).

The idea is to think of the function φ we are applying the operator to as a probability density function of some random variable $\boldsymbol{Y}$. Then, $\mathcal{L}_f\varphi$ can be thought of as the probability density function of $\boldsymbol{Y} \circ f$.

The transfer operator is a fashionable concept in dynamical systems, as it allows us to convert problems in ergodic theory into problems in functional analysis, where many theorems (particularly relating to eigenvalues) are available. The standard modern reference for a textbook exploring transfer operator techniques in dynamical systems is Viviane Baladi's book [Bal00].

Proposition 1.27 (Properties of the Transfer Operator)**.** *Let $(X, \mathcal{E}, \mu, f)$ be a nonsingular dynamical system on a finite measure space and let $\mathcal{L}_f$ be the transfer operator of f. Then the operator has the following properties.*

1. *(Linearity). $\mathcal{L}_f$ is linear, i.e. $\mathcal{L}_f(a_1\varphi_1 + a_2\varphi_2) = a_1\mathcal{L}_f(\varphi_1) + a_2\mathcal{L}_f(\varphi_2)$.*
2. *(Composition). Let g be another nonsingular transformation of $(X, \mathcal{E}, \mu)$. Then $\mathcal{L}_{g\circ f} = \mathcal{L}_g \circ \mathcal{L}_f$. Namely, $\mathcal{L}_{f^n} = (\mathcal{L}_f)^n$.*
3. *(Invariance). For $\varphi \in L^1(\mu)$, we have $\mathcal{L}_f\varphi = \varphi$ if and only if the measure $\nu \ll \mu$ given by $d\nu = \varphi \, d\mu$ is f-invariant.*
4. *(Relationship with Push-Forwards). If $\nu \ll \mu$ has Radon-Nikodym derivative φ, then the Radon-Nikodym derivative of $f_*\nu$ is $\mathcal{L}_f\varphi$.*

These properties are all from [BG97] where their proofs are also available, except for the final point which follows from footnote 9.

It is possible to phrase ergodic properties of dynamical systems in terms of the transfer operator, but we will not do this here. Note, though, that finding an absolutely continuous invariant measure is equivalent to finding a fixed point of the transfer operator.

Later we will need to apply the transfer operator to a function on the interval. There happens to be a direct formula for the transfer operator associated with a dynamical system on the interval, provided this system is nice enough. This is proved in [BG97, Section 4.3].

[9] Note that existence and uniqueness of a function $\mathcal{L}_f\varphi$ satisfying this equation is not immediately obvious, but justification is given in [Aar97]. It uses the following explicit construction: let φ be the R-N derivative for $\nu_\varphi := \varphi\mu$. Then $\mathcal{L}_f\varphi$ is the R-N derivative of $f_*\nu_\varphi$.

Proposition 1.28 (Pointwise Definition of the Transfer Operator)**.** *Let $(I, \mathcal{B}(I), \lambda)$ be the standard measure space on an interval $I = [a, b] \subseteq \mathbb{R}$. Let $f : I \circlearrowleft$ be a* piecewise monotonic transformation *on I, i.e. there exists an $r \geq 1$ such that I can be partitioned into a finite number of intervals on which f is C^r and $|f'| > 0$. Then for all $\varphi \in L^1$:*

$$\mathcal{L}_f \varphi(x) = \sum_{z \in f^{-1}(\{x\})} \frac{\varphi(z)}{|f'(z)|}.$$

This final characterisation is the main one we will use in Chapter 3.

1.6 Intermittency

With some dynamical systems background under our belt, we can now define the phenomenon we are seeking to observe in this project. The choice of the term "intermittency" goes back to Pomeau and Manneville's paper [PM80] mentioned in the introduction, although the phenomenon had been observed in many physical experiments before then. The setting of Pomeau and Manneville's paper is parametrised families of dynamical systems, and their claim is that "intermittency" can be observed near the transition from a stable parameter to a turbulent one (these are often referred to as *bifurcations*). Pomeau and Manneville distinguish between three types of bifurcations where intermittency can be observed, and call these type 1, type 2, and type 3. However, the motion observed at the bifurcation seems to have the same kind of behaviour regardless of the type.

The intuition for intermittency appearing at bifurcations seems to be due to phase transitions. On one side of the critical parameter, motion is smooth and predictable, while on the other, motion is chaotic and unpredictable. Only fine choices of parameters allow the system to be balanced enough that a bit of both can appear.

Rather than looking for bifurcations where intermittency can be found, the aim of this project is to study the intermittency itself, in a qualitative sense. For reference, let's introduce two quotes from [PM80] where the intermittency they observe is described. The examples giving these descriptions are rather unrelated to the direction of this project, but the motion we uncover will nonetheless be similar.

This first quote is an observation on convective fluids, which are continuous-time dynamical systems.[10]

> "[At the bifurcation,] the fluctuations remain apparently periodic during long time intervals (which we shall call "laminar phases") but this regular behavior seems to be randomly and abruptly disrupted by a "burst" on the time record. This "burst" has a finite duration, it stops and a new laminar phase starts and so on." [PM80]

This second quote is from a discrete dynamical system on the torus designed to have a tangency at the origin, not unlike the map we will study in the third chapter.

[10] It is possible to turn a continuous-time system into a discrete-time one, and in fact this is done in the paper for type 1 intermittency. For the Lorenz model in 3D, the authors consider a map which sends a point $\boldsymbol{x}$ in the plane $x = 0$ to the point of next crossing of $O(\boldsymbol{x})$ with the plane $x = 0$.

> "Once an iterate falls near $z = 0$, it enters a laminar phase and a large number of further iterations are needed to expell it towards the "bursting region" (where correlations are broken)" [PM80]

These quotes lead us to a general definition of intermittency which might look something like:

Definition 1.29 (Intermittency)**.** A dynamical system is called *intermittent* if its orbits alternate unpredictably between long, laminar, predictable phases and chaotic, unpredictable phases.

We will mostly be looking at interval maps (from the point of view of ergodic theory) in this project, and the literature in this area tends to refer to "intermittency" when a map is *non-uniformly expanding*, specifically in the case of there being a *neutral fixed point*. We will clarify what this means, and how it relates to intermittency, in the third chapter. However, we note in the conclusion of this project that interval maps can be intermittent outside of this context too.

For the time being, our goal should be to build up some tools that will allow us to study examples in such a way that we will able to see, and distinguish between, the laminar and chaotic phases.

Chapter 2

Visualising Dynamical Systems

Consider a dynamical system $f : X \circlearrowleft$. In general, X could be a complicated set that we wish to simplify in some way to make the dynamics easier to understand. This is often done in the literature, regardless of whether we are expecting f to have intermittent properties or not.

However, in the case of intermittent maps, this simplification task is of particular importance. It will, for example, allow us to encode systems in a way which will make their laminar and chaotic phases appear. There are several ways of doing this, and we will first study the "tower" description of a dynamical system before looking at how we may encode one as a Markov shift.

2.1 Young towers

2.1.1 Some motivation

In classical ergodic theory research going back as early as the 1940s, there appears to have been a desire to describe dynamical systems (X, f) through some form of upward motion followed by an eventual return to the ground.

This usually means partitioning the state space into a finite or countable number of "floors", each one possibly subdivided into different sections, in such a way that applying f moves from one floor to the next one up. This has led to multiple different sorts of dynamical "towers" earning a name, each with their own applications and assumptions.

A *Rokhlin tower* partitions all but ε of the space X into n floors, each one an iterate of a base set B.[1] This is a somewhat low-resolution representation, as some points in B might map back to B in exactly n iterates, while others may wander around in the remaining ε for many more iterates before returning. There is therefore a need to describe the space X completely, rather than up to an arbitrarily small set, if we want to make any inferences about the long-term behaviour of orbits.

[1] The existence of such a B for any n and any $\varepsilon > 0$ is implied (provided we have an aperiodic ppt) by a theorem of Rokhlin in 1948. This is discussed in [Wei89].

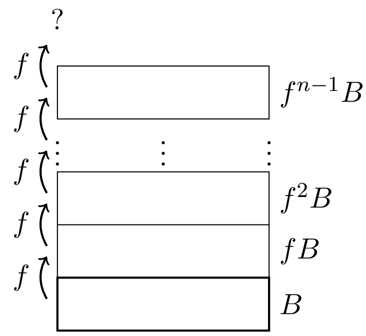


Figure 2.1: A Rokhlin tower. Movement up the tower is "laminar" in a sense, but we know too little about the remaining part of the space with measure ε.

To get more information out of the Rokhlin tower, we could simply add more levels on the top in an effort to cover the remaining ε. Progressively, as points in B see their orbits return to B after $n, n+1, \ldots$ iterations of f, the size of each additional level will whittle down until we eventually cover all of X. This is a special case of a *Kakutani skyscraper* (see [Pet83, Section 1.3]). It certainly covers all of X, but doesn't really provide an interesting framework for the study of intermittent systems; we require something more specialised.

Returning to our Rokhlin tower, then, we know that the different "return times" of points in B are causing a mess at the top of the tower. One thing we could try is to further subdivide the base set B in such a way that if two points lie in the same subdivision, they take the same number of applications of f to return to B. Say this number is some integer r; then we will be able to extend the subdivision containing the two points into r levels, the top of which will then map back into B under f. This will be the intuition necessary for the tower model I am seeking to introduce.

First, let's formalise this notion of return time that we will use to partition the base set.

Definition 2.1 (Return Time)**.** Let $f : X \circlearrowleft$ for some measure space $(X, \mathcal{E}, \mu)$. Let $B \in \mathcal{E}$. The *return time function* of B is

$$\begin{aligned} R_B : B &\longrightarrow \mathbb{Z}^+ \\ x &\longmapsto \min\{k \geq 1 : f^k(x) \in B\}. \end{aligned}$$

If the orbit of some $x \in B$ never re-enters B, we set $R_B(x) = \infty$.

In general, there is no reason for the orbit of $x \in B$ to ever re-enter B. However, if the system $(X, \mathcal{E}, \mu, f)$ satisfies the conditions of Poincaré's Recurrence Theorem (Theorem 1.11) and $\mu(B) > 0$, then in fact R_B is finite μ-a.e. By this point, notice that we have freed ourselves from the setting of the Rokhlin tower, and can take B to be whatever positive-measure set we want (rather than one whose iterates cover all but ε of the space).

When the return time is finite μ-a.e., we can break down our base set B into subsets on which every point has the same return time. In the most complete case, this could give a partition $B_1, B_2, \dots \subseteq B$, where $R_B|_{B_i}$ is the constant function i. However, in some cases, not every return time is possible, so in general we refer to R_i as the constant integer return time on the i^{th} partition set.

2.1.2 Defining a Young tower

We are now ready to define a *Young tower*, which has some technical conditions attached that immediately prove a number of mixing and recurrence properties. This definition is from Lai-Sang Young's paper [You99], although some additional interpretation is available in [Bal00, Section 3.5]. The initial definition is abstract, but we will soon also see how to construct a Young tower from a pre-existing system.

Definition 2.2 (Young Tower)**.** Let $(\Delta_0, \mathcal{B}|_{\Delta_0}, m)$ be a finite measure space (the idea is that Δ_0 will be our base set for a larger space Δ).[2] Partition the base set Δ_0 into $\{\Delta_{0,i}\}_{i=1,2,\dots}$ (these will be our sets of constant return time). Let $R : \Delta_0 \to \mathbb{Z}^+$ be a return time function which is constant on each $\Delta_{0,i}$, and let R_i denote the value of R on $\Delta_{0,i}$. Finally, for each $i \geq 1$, let $\zeta_i : \Delta_{0,i} \to \Delta_0$ be a bijection. This is all we are free to choose, and the rest of this definition introduces notation and conditions on these parameters.

First, this setup uniquely determines a larger set Δ which we call the *tower*, and which forms the state space. Each element of Δ is a pair (z, n), where z denotes base location and n represents height (or "floor number"):

$$\Delta = \{(z, n) : z \in \Delta_0;\ n \in \mathbb{Z}_{\geq 0};\ n < R(z)\}.$$

The setup also uniquely determines the *tower map* $F : \Delta \circlearrowleft$. This sends everything upwards, except at the top of the tower, where it maps back down into the base Δ_0:

$$F(z, n) = \begin{cases} (z, n+1) & n+1 < R(z) \\ (\zeta_i(z), 0) & n+1 = R(z) \quad \text{(where } i \text{ is such that } z \in \Delta_{0,i}). \end{cases}$$

To ensure Δ actually contains its base set Δ_0, we abuse notation and imagine Δ_0 to formally be a set of pairs, rather than a set of points, when this is needed: $\Delta_0 = \{z : z \in \Delta_0\} = \{(z, 0) : z \in \Delta_0\} \subseteq \Delta$.

We then also define Δ_l to be the l^{th} *floor* of the tower Δ, and define $\Delta_{l,i}$ to be the part of floor number l lying above the i^{th} base partition set:

$$\Delta_l = \Delta \cap \{n = l\}; \quad \Delta_{l,i} = \Delta_l \cap \{z \in X_{0,i}\}.$$

Each $\Delta_{l,i}$ is one level of a *column* of the tower, with base $\Delta_{0,i}$. In column i, the highest level is $\Delta_{R_i-1,i}$. The collection $\mathcal{Z} := \{\Delta_{l,i} : i \geq 1, 0 \leq l < R_i\}$ partitions Δ.

The *induced system* is the dynamical system $F^R : \Delta_0 \circlearrowleft$ given by $F^R(z) = F^{R(z)}(z, 0) = \zeta_i(z)$ (where i is such that $z \in \Delta_{0,i}$).[3]

The *separation time* of two points in the base Δ_0 says how long it takes for their orbits to move apart:

$$\forall x, y \in \Delta_0 : s(x, y) = \min\{n \geq 0 : (F^R)^n(x), (F^R)^n(y) \text{ lie in separate } \Delta_{0,i}\}.$$

If we need to extend separation time to the whole tower Δ, we let $s(x, y)$ equal 0 unless x, y lie in the same $\Delta_{l,i}$, in which case $s(x, y) := s(x', y')$ where x', y' are the corresponding points in $\Delta_{0,i}$.

[2] Usually the base set Δ_0 will be a set of significance in the system we are trying to represent as a tower, and m will be something like Lebesgue measure.

[3] We will tend to use F^R rather than the individual ζ_i bijections, since F^R is defined on the entire base set Δ_0. However, think of F^R on Δ_0 as having as many *branches* ζ_i as there are $\Delta_{0,i}$ in the base set.

The *hitting time* of a point $(z, l) \in \Delta$ is the smallest number of iterations of F needed to enter the base Δ_0:

$$\hat{R}(z,l) = \min\{n \geq 0 : F^n(z,l) \in \Delta_0\} = \begin{cases} 0 & l = 0 \\ R(z) - l & 0 < l < R(z). \end{cases}$$

We then define the *size of the tails* τ_n to be the measure of all points more than n iterates away from the base:

$$\tau_n = m(\{x \in \Delta : \hat{R}(x) > n\}).$$

Finally, the measure space for the full tower $(\Delta, \mathcal{B}, m)$ is obtained by translating the measurable sets in $\mathcal{B}|_{\Delta_0}$ up the tower via F, and preserving measure as follows: if $A \subseteq \Delta_{l,i}$, then $m(A) = m(F^{-l}A)$.[4]

We call the system defined above a *Young Tower* if the following conditions are satisfied:

Y1. (Measurability). All the sets mentioned are $\mathcal{B}$-measurable, as are the ζ_i, as well as F and its inverses.

Y2. (Strong Generation). $\mathcal{Z}$ is a strong generator for $\mathcal{B}$.

Y3. (Aperiodicity). $\gcd\{R_i : i = 1, 2, \dots\} = 1$.

Y4. (Bounded Distortion). For each i, $\zeta_i : \Delta_{0,i} \to \Delta_0$ and its inverse are nonsingular with respect to m. Also, $\exists C > 0$ and $\beta \in (0,1)$ such that:

$$\forall i \geq 1 : \forall x, y \in \Delta_{0,i} : \left| \frac{JF^R(x)}{JF^R(y)} - 1 \right| \leq C\beta^{s(F^R x, F^R y)}$$

where J denotes the Jacobian[5] (which exists and is positive by nonsingularity).

Y5. (Return Time Integral). $\int R \, dm < \infty$.

I should note that the notation is taken to be identical to [You99], but the definition has been re-ordered in the hopes of making it easier to understand and to check. The bijections ζ_i have also been added; their action on Δ is simply referred to as F^R in Young's paper.

This definition is a bit of a mouthful, so let's take some time to pick it apart and discuss how it can actually be useful to us.

Most of the definition is there to define notation and terms that are useful when talking about how the tower map acts on the tower Δ. When actually proving that something is a Young tower, we only need to show that it satisfies the conditions of the first paragraph, plus conditions **Y1–Y5**.

In practice, the idea will be to construct a Young tower from a system that we already have, and that we wish to study—say, $(X, \mathcal{E}, \mu, f)$. Then, we can let the base Δ_0 equal the "base set" B of our system, and we can carefully choose R and ζ_i in the definition in such a way that we obtain a Young tower Δ. Note that this tower will not be equal to X, despite it being a representation of X. Instead, Δ will contain all the points in B, and then copies of B further up the tower in the form (b, n) (for $b \in B$ and $n < R_B(b)$). The other discrepancy in representation is that the measure of sets in $X \setminus B$ may not equal the measure of their representation in Δ, even if the reference measures agree on B. This is because F carries the m-mass of each $\Delta_{0,i}$ evenly up the tower, while

[4] Young calls this "carrying" the measure. The full measure space need not be finite, but it is always σ-finite.

[5] We will be looking at interval maps later, so the Jacobian will just be the derivative.

the original system f may map subsets of B to sets of smaller measure. Finally, as we will see later, this construction means that one single point in $X \setminus B$ may be represented by multiple points in Δ.

Despite the discrepancies in representation, useful conclusions can still be drawn from the Young tower representation of a map. In the following definition, I propose a way to formalise the construction of a Young tower from a system $(X, \mathcal{E}, \mu, f)$ and a base set B. In the interest of having analogous notation to (Δ, F), the base set B is referred to here as X_0.

Definition 2.3 (Young Tower Representation)**.** Let $(X, \mathcal{E}, m)$ be a finite measure space and $f : X \circlearrowleft$ be measurable. Let $X_0 \in \mathcal{E}$ be a base set of positive measure whose iterates cover almost all of X (i.e. $m(\cup_{n\geq 0} f^n X_0) = m(X)$) and such that the return time function R_{X_0} is finite for m-almost every point in X_0. Suppose X_0 admits a measurable partition $\{X_{0,i}\}_{i=1,2,\ldots}$ where R_{X_0} is constant on each partition set.[6] Finally, assume that f and its inverse are nonsingular with respect to m. We say that (X, f) has a *Young tower representation* as (Δ, F) with base X_0 if $f^{R_{X_0}}|_{X_{0,i}} : X_{0,i} \to X_0$ is a bijection for all i, and if it is possible to construct a valid Young tower with the following parameters from Definition 2.2:

- $(\Delta_0, \mathcal{B}|_{\Delta_0}, m) = (X_0, \mathcal{E}|_{X_0}, m)$;
- $\Delta_{0,i} = X_{0,i}$;
- $R = R_{X_0}$;
- $\zeta_i = f^{R_{X_0}}|_{X_{0,i}}$.

If a tower can be constructed but it does not satisfy some or all of the conditions **Y2–Y5**, we still call the above a Young tower representation, but we agree to explicitly state which conditions are not met. **Y1** is more essential as measurability is always desirable.

Remark 2.4. This definition will allow us to convert an arbitrary system into a Young tower in a standardised way. Note that the various conditions on $(X, \mathcal{E}, m)$ are all there to ensure we can slot the system nicely into Young's framework. Additionally, we require that the iterates of X_0 sweep m-almost all of the space X: this is so that for m-almost every $x \in X$, there exists $z \in X_0$ and $n \geq 0$ such that $x = f^n(z)$. This means x will have a representation in the tower as (z, n). Conversely, each point $(z, n) \in \Delta$ will correspond to $f^n(z)$, and so each set $\Delta_{l,i}$ will correspond to $f^l X_{0,i}$. The final nonsingularity condition is to ensure that the assumption that F carries m up the tower Δ is not entirely unreasonable when looking at the dynamics on X.

2.1.3 Some immediate results

As mentioned earlier, if (Δ, F) is a Young tower, this implies that it has a number of nice properties. These include statements on invariant measures, the system's rate of mixing, and some statistical limit laws. I invite the reader to look at Young's paper [You99], where these results are contained in Theorems 1–4 with the same notation as we have used here, then proved in the latter part of the paper. These will be of use to us in the next chapter.

[6] In practice, if we know the return time function, we can use it to partition X_0 by setting $X_{0,i} = \{x \in X_0 : R_{X_0} = i\}$.

Rather than repeating Young's work here, I suggest we use the tower model to prove a weaker, more accessible version of one of Young's theorems which will benefit us when studying the Manneville-Pomeau map in Chapter 3. The idea is to work through the proof to see the tower in action. At the end, we will be able to move away from Young tower representations entirely, keeping just the intuition developed, to show that the proposition holds for general systems too.

Proposition 2.5 (From [You99], Theorem 1)**.** *Let (Δ, F) be a Young tower as in Definition 2.2. If there is a probability measure ν_0 on $(\Delta_0, \mathcal{B}|_{\Delta_0})$ such that $\nu_0 \ll m$ and ν_0 is F^R-invariant, then $(\Delta, \mathcal{B})$ admits an F-invariant measure*[7] *$\nu \ll m$.*

Proof. The idea for this proof is to say the following: "to get the measure space $(\Delta, \mathcal{B}, m)$, we took the measure m on Δ_0 and carried it up the tower. Can we just do the same with ν_0?"

The answer, conveniently, is yes. So for a set $A \in \mathcal{B}$, we somehow "project" A down onto the base and measure it there. A closed form for the resulting measure is as follows:

$$\nu(A) = \sum_{j=1}^{\infty} \sum_{k=0}^{R_j - 1} \nu_0(F^{-k} A \cap \Delta_{0,j}). \tag{2.1}$$

This is the measure given by Young in the proof of Theorem 1, but no justification of invariance or absolute continuity is given, so I will fill in the blanks to make the proof easier to follow.

Basic measure theory shows that ν is indeed a measure, since it is a linear sum of measures all constructed from ν_0 using restrictions and push-forwards (see Proposition 1.4).

We need to prove that this measure is F-invariant, and absolutely continuous with respect to m. First, it may help to calculate a simpler expression for $\nu(A)$ for particular values of A, namely if A is fully contained in one element of the partition $\mathcal{Z}$. To this end, suppose $A \subseteq \Delta_{l,i}$. Iterates $F^{-k}A$ will only intersect non-trivially with a $\Delta_{0,j}$ within $k \leq R_j - 1$ steps if $j = i$, and only $k = l$ will take us back to level 0. Noting that $F^{-l}A \subseteq \Delta_{0,i}$, we get:

$$\nu(A) = \nu_0(F^{-l}A).$$

This is a slightly clearer demonstration of the fact that ν projects A to the base and measures it there. (The double sum in the definition of ν is just there to deal with the case where A may overlap with multiple sets in $\mathcal{Z}$; we need different negative powers of A to map different parts of A to the base).

We're now able to prove that ν is F-invariant on sets $A \subseteq \Delta_{l,i}$. First suppose $l > 0$; then $F^{-1}A \subseteq \Delta_{l-1,i}$, and

$$\nu(F^{-1}A) = \nu_0(F^{-(l-1)}F^{-1}A) = \nu_0(F^{-l}A) = \nu(A).$$

Secondly suppose $l = 0$. Taking the inverse of a set at the bottom of the tower gives a union of sets at the top, as in the diagram below. These sets will all be some levels directly above $(F^R)^{-1}A \subseteq \Delta_0$.

[7] The measure ν is not necessarily a probability measure.

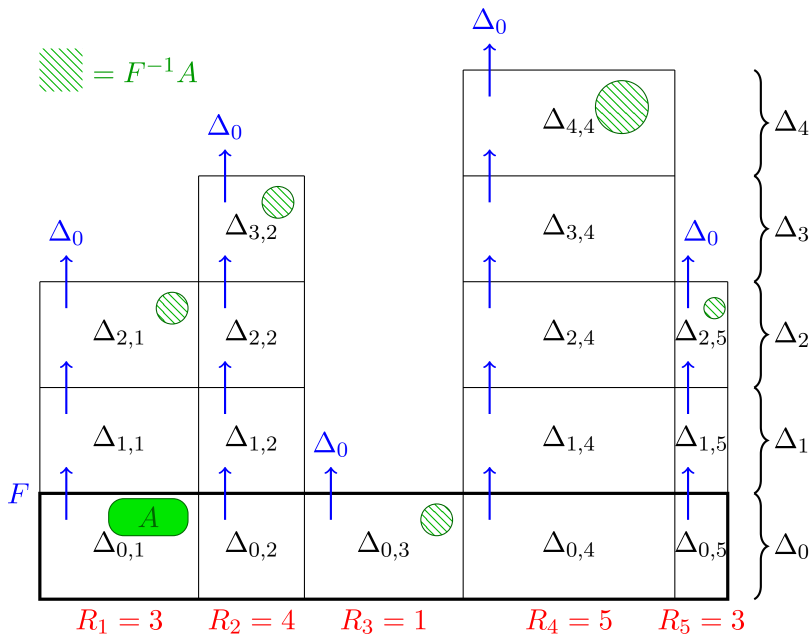


Figure 2.2: A Young tower with the pre-image of A marked.

When we take their ν-measure, all these sets will be projected back down to their analogous points in Δ_0, and so cancelling the gain in levels with the loss of levels due to ν we get

$$\begin{aligned}
\nu(F^{-1}A) &= \sum_{j=0}^{\infty} \nu_0((F^R)^{-1}A \cap \Delta_{0,j}) \\
&= \nu_0((F^R)^{-1}A \cap \Delta_0) && \text{since we are summing over all } j \\
&= \nu_0((F^R)^{-1}A) && \text{Since } (F^R)^{-1}A \subseteq \Delta_0 \\
&= \nu_0(A) && \text{by invariance} \\
&= \nu(A)
\end{aligned}$$

Having covered the case where A sits nicely inside one partition set, we can now generalise to any measurable set A as follows. Let $A = A_1 \cup A_2 \cup \dots$ be a partition of A into disjoint sets that all sit inside one element of $\mathcal{Z}$. (Certainly we don't need more than a countable partition here, since $\mathcal{Z}$ is countable). Each A_i can be chosen to be measurable[8], and furthermore $F^{-1}A_i$ is disjoint from $F^{-1}A_j$ for all $i \neq j$. So:

$$\nu(F^{-1}A) = \nu(F^{-1}A_1) + \nu(F^{-1}A_2) + \cdots = \nu(A_1) + \nu(A_2) + \cdots = \nu(A).$$

To prove that $\nu \ll m$, suppose that $A \in \mathcal{B}$ has positive ν-measure. Then at least one set in $\{A \cap Z : Z \in \mathcal{Z}\}$ must have positive ν-measure. Call this set A' and suppose it lies at level l. We

[8]The easiest choice is to set each partition element to $A \cap \Delta_{l,i}$ for some l, i.

have

$$\begin{aligned}
\nu(A') &= \nu_0(F^{-l}A') && \text{since } A' \text{ lies in one partition element} \\
&\leq m(F^{-l}A') && \text{since } \nu_0 \ll m \\
&= m(A') && \text{since } m \text{ is carried up the tower by assumption} \\
&\leq m(A) && \text{since } A' \subseteq A.
\end{aligned}$$

Since we had $\nu(A') > 0$, this concludes the proof. □

Remark 2.6. This proof did not use any of the conditions **Y2–Y5**! This suggests that we are looking at something more general than Young's structures, and indeed we will see a generalisation of this just after having discovered the limitations through an example. In the original statement of Young's Theorem 1, the assumptions are necessary in order to reach a stronger conclusion: an F-invariant measure always exists (we don't even need to find ν_0 first—its existence is guaranteed by the theorem, and it is absolutely continuous with respect to m).[9] Furthermore, this invariant measure is ergodic and finite (with this proof, finiteness is not guaranteed, though we correct this in Proposition 2.9).

We have established that if we can represent a dynamical system as a Young tower (Δ, F), and if we can find a probability measure on its base Δ_0 which is invariant under the induced map F^R, then we can somehow "push" that base measure so that it extends to the full tower Δ. However, this is not very useful unless we can turn this measure into an invariant measure for our original space X, with our original function f acting on it: currently, we have an invariant measure for a system similar to, but not actually isomorphic to, the original system. Fortunately, we are quite close to deducing from ν an invariant measure on X. Let's look at an example to see why the two systems (X, f) and (Δ, F) are not analogous.

2.1.4 Non-uniqueness of representation

In this subsection, I provide an example to see how to construct a Young tower. We will also encounter an issue in the way Young towers represent the space they are constructed from, which we will then attempt to reconcile in following subsections.

Consider the set of states $S = \{1, 2, 3\}$ and the topological transition matrix

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

This gives rise to a shift space Σ_A^+ (see Definition 1.17), containing all infinite sequences of digits $1, 2, 3$ where a 1 can be followed by any digit, but all 2s are followed by a 1 and all 3s are followed by a 2. We will try to represent this system as a Young tower, but will not worry about the technical conditions.

Note that because of the rules on what digits may follow a 2 or a 3, we actually have e.g. $[1, 2] = [1, 2, 1]$ and $[1, 3] = [1, 3, 2] = [1, 3, 2, 1]$, where the sequences surrounded by brackets denote cylinder sets. This will be useful later.

[9] The proof of existence of ν_0 can be seen as one form of the Ergodic Folklore Theorem (Theorem 3.3).

When we apply the one-sided shift σ to Σ_A^+, we get, for example,

$$\sigma((1,1,2,1,3,2,1,1,3,2,1,\dots)) = (1,2,1,3,2,1,1,3,2,1,\dots).$$

We might notice, looking at A, that if a point lies in the cylinder [1] (all words starting with 1), then we are guaranteed to return to [1] very quickly. Indeed, if $\boldsymbol{w} \in [1]$, then one of the following will happen:

- $\sigma(\boldsymbol{w}) \in [1]$
- $\sigma(\boldsymbol{w}) \in [2] \implies \sigma^2(\boldsymbol{w}) \in [1]$
- $\sigma(\boldsymbol{w}) \in [3] \implies \sigma^2(\boldsymbol{w}) \in [2] \implies \sigma^3(\boldsymbol{w}) \in [1]$.

Hence the return time function to [1], denoted $R_{[1]}$, takes values in $\{1,2,3\}$. In fact, we can characterise this function fully:

$$R_{[1]}(\boldsymbol{w}) = \begin{cases} 1 & \boldsymbol{w} \in [1,1] \\ 2 & \boldsymbol{w} \in [1,2] \\ 3 & \boldsymbol{w} \in [1,3]. \end{cases}$$

This is exactly the partition we need to start building our tower. Using the notation from Definition 2.3, define $X_0 = [1]$, and partition this base set into three parts on which the return time function is constant: $X_{0,i} = [1,i]$ with $R_i = i$ for $i = 1,2,3$. In our tower, elements $\boldsymbol{w} \in X_{0,i}$ are represented by $(\boldsymbol{w}, 0) \in \Delta_{0,i}$.

To construct the higher tower levels, we add $i-1$ floors above each $\Delta_{0,i}$. So, $\Delta_{0,1}$ has nothing above it, since the return time on $X_{0,1}$ is 1. The base set $\Delta_{0,2}$ has one level above it, and $\Delta_{0,3}$ has two above it. This fully describes the tower structure as in the picture below:

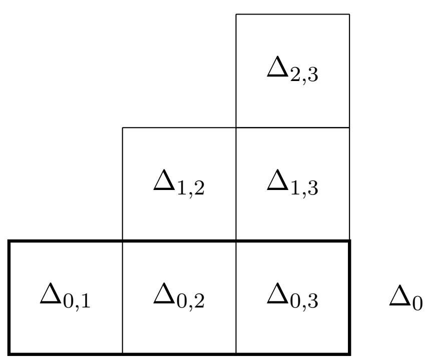


Figure 2.3: Symbolic structure of the Young tower

This is where the difference between Σ_A^+ and Δ, mentioned in the previous section, becomes noticeable. In Remark 2.4, we said that every set $\Delta_{l,i}$ was meant to correspond to $\sigma^l X_{0,i}$ in the original shift space. In our example, this namely means that:

- $\Delta_{1,2}$ corresponds to $\sigma X_{0,2} = \sigma[1,2] = \sigma[1,2,1] = [2,1]$;
- $\Delta_{2,3}$ corresponds to $\sigma^2 X_{0,3} = \sigma^2[1,3] = \sigma^2[1,3,2,1] = [2,1]$.

So, the same set $[2,1]$ in the shift space is represented by two *distinct* sets in the tower: $\Delta_{1,2}$ and $\Delta_{2,3}$. This shows that while the tower representation is a good way to visualise the rough structure of the dynamics on Σ_A^+, the representation is not in bijection with the original set. It may duplicate elements of the partition so that the same point in Σ_A^+ could be in multiple different parts of the tower, depending on where in the base set X_0 the orbit started. In the worst case, this representation could be countable-to-one (a point in the state space could be represented by a different point in every column of the tower, and the tower could have countably many columns). This will require some thought when we look to convert an invariant measure on Δ into an invariant measure on Σ_A^+.

The key when converting back will be to "squash" down all the different points in Δ that actually correspond to the same point, down to that one point. Returning to full generality, this leads naturally to a notion of projection as alluded to in Remark 2.4. This is a known construction in the literature, but I propose some clarity on the projection that I have had trouble finding elsewhere (likely because these sorts of properties seem to be considered trivial).

2.1.5 The Young tower projection

Definition 2.7 (Young Tower Projection)**.** Let $(X, \mathcal{E}, m)$ be a finite measure space and $f : X \circlearrowleft$ be measurable. Suppose (X, f) admits a Young tower representation (see Definition 2.3) as (Δ, F) with base X_0, partitioned into $\{X_{0,i}\}_{i=1,2,\ldots}$. (We do not require that **Y2–Y5** are necessarily met). Then we define the *Young tower projection* π_Δ as follows.

$$\begin{aligned} \pi_\Delta : \qquad \Delta &\longrightarrow X \\ (z, n) &\longmapsto f^n(z). \end{aligned}$$

Proposition 2.8. *Let $(X, \mathcal{E}, m)$ be a finite measure space and $f : X \circlearrowleft$ be measurable. Suppose (X, f) admits a Young tower representation with base X_0 on the measure space $(\Delta, \mathcal{B}, m)$, and suppose that (Δ, F) has an invariant measure $\nu \ll m$. Then, the measure $\nu \circ \pi_\Delta^{-1}$ on X is f-invariant and a.c. w.r.t. m. This is the case regardless of whether **Y2–Y5** are met.*

Proof. Caution is necessary throughout this proof, as we have slightly abused notation by considering m to be a measure both on X and on $\mathcal{B}$. In reality, m is the ambient measure on X, and we extend it to a measure on $\mathcal{B}$ by assuming that m is the same on X_0 and on Δ_0, and then carrying m up the tower Δ.

Our first aim is to show that $\pi_\Delta \circ F = f \circ \pi_\Delta$ (i.e. the projection *conjugates* the two systems). Let $(z, n) \in \Delta$.

- If $n + 1 < R(z)$, then $\pi_\Delta(F(z, n)) = \pi_\Delta(z, n + 1) = f^{n+1}(z) = f(f^n(z)) = f(\pi_\Delta(z, n))$.
- If $n + 1 = R(z)$, then $\pi_\Delta(F(z, n)) = \pi_\Delta(f^{R(z)}(z), 0) = f^{R(z)}(z) = f^{n+1}(z) = f(f^n(z)) = f(\pi_\Delta(z, n))$.

This proves $\pi_\Delta \circ F = f \circ \pi_\Delta$, and we can invert this to get

$$F^{-1} \circ \pi_\Delta^{-1} = \pi_\Delta^{-1} \circ f^{-1}.$$

Now let $E \in \mathcal{E}$. To show $\nu \circ \pi_\Delta^{-1}$ is f-invariant:

$$\begin{aligned} \nu \circ \pi_\Delta^{-1}(f^{-1}E) &= \nu(\pi_\Delta^{-1} \circ f^{-1}E) \\ &= \nu(F^{-1} \circ \pi_\Delta^{-1}E) \\ &= \nu(\pi_\Delta^{-1}E) \\ &= \nu \circ \pi_\Delta^{-1}(E). \end{aligned}$$

To show $\nu \circ \pi_\Delta^{-1}$ is absolutely continuous with respect to m, let $E \in \mathcal{E}$ have positive $\nu \circ \pi_\Delta^{-1}$-measure. Then $\nu(\pi_\Delta^{-1}E) > 0$, hence $m(\pi_\Delta^{-1}E) > 0$ (since we assumed $\nu \ll m$) where here, m is a measure on Δ. Since $\mathcal{Z}$ is a countable partition of Δ, this means we can find a $\Delta_{l,i}$ such that $E' := (\pi_\Delta^{-1}E) \cap \Delta_{l,i}$ has positive m-measure. Now, by the conjugation properties of π_Δ, we have

$$f^l \pi_\Delta F^{-l} E' \subseteq f^l \underbrace{\pi_\Delta F^{-l} \pi_\Delta^{-1}}_{=f^{-l}} E = E. \tag{2.2}$$

Since m on Δ is generated by carrying m on Δ_0 upwards and $E' \subseteq \Delta_{l,i}$, we have $m(F^{-l}E') = m(E') > 0$. Furthermore, this new set $F^{-l}E'$ lies in the base Δ_0 where m agrees between both measure spaces, so the projection has no effect: $m(\pi_\Delta F^{-l}E') = m(F^{-l}E') > 0$. By nonsingularity of the inverse of f (assumed in Definition 2.3), this further implies $m(f^l \pi_\Delta F^{-l}E') > 0$. By (2.2) we conclude that $m(E) > 0$.

So $\nu \circ \pi_\Delta^{-1}$ does indeed satisfy the conditions. □

Using the above theorem, we can recover from a measure on a Young tower an analogous projected measure on the original system.

Returning to the case of expressing a pre-existing system (X, f) as a Young tower, our workflow for finding an absolutely continuous measure has therefore been:

1. Express the original system (X, f) as a Young tower (e.g. using Definition 2.3).
2. Find an invariant measure on the induced map F^R acting on the base Δ_0 of the Young tower (e.g. using Young's theorems in [You99], or directly on the original system).
3. Push this measure up the tower to find a measure ν for Δ (using Proposition 2.5).
4. Squash this measure down to X via the projection π_Δ (using Proposition 2.8).

If we don't have an interest in expressing the system as a Young tower at all, then there is an analogous method of "pushing" a measure for the induced map back to the original system that stays in X the whole time. If we are inducing on a base set $X_0 \in \mathcal{E}$ that satisfies the conditions of Definition 2.3 (not necessarily **Y2–Y5**) and get a function f^R which has an invariant measure $\kappa_0 \ll m$, then we can push this measure back onto (X, f) using:

$$\kappa(A) = \sum_{j=1}^{\infty} \sum_{k=0}^{R_j - 1} \kappa_0(f^{-k}(A) \cap X_{0,j}), \tag{2.3}$$

where we have used the same partition of X_0 into $\{X_{0,j}\}_j$, sets on which R_{X_0} is constant and equal to R_j. This is equivalent to the measure ν we found for the Young tower, although the proof of invariance is slightly less visual. In fact, one can even show that the measures obtained via both methods are the same: $\nu \circ \pi_\Delta^{-1} = \kappa$, provided that ν was generated by pushing κ_0 too.[10]

However, this approach has the advantage of skipping steps 1 and 4 from the list above; and since all the inducing is happening in the original space $(X, \mathcal{E}, m)$, we have less to keep track of when ensuring absolute continuity of measures and conversion of σ-algebras. This is the technique we will use later when working on the Manneville-Pomeau map.

2.1.6 Getting a probability measure

The measure κ, which we just found to be f-invariant, is also absolutely continuous with respect to the reference measure m on X. However, one key property of this measure remains to be proved: can it be made into a probability measure?

This all depends on the value of $\kappa(X)$: if the size of the space is finite, we can normalise our measure. If it's infinite, then we haven't found a probability measure, though it is still invariant.

To compute $\kappa(X)$, it's helpful to partition X in a way which simplifies the expression of κ. The best way to do this is in fact to go back to the equivalent Young tower expression, where every block of the tower is a set $\Delta_{l,i}$, and the set of all $\Delta_{l,i}$ partitions Δ. So, we consider

$$\kappa(X) = \nu(\pi_\Delta^{-1} X) = \nu(\Delta).$$

Conveniently, this means all we have to do is compute the measure of the tower Δ. Recall from the proof of Proposition 2.5 that the measure of $A \subseteq \Delta_{l,i}$ is $\nu(A) = \nu_0(F^{-l}A)$. Namely, this implies $\nu(\Delta_{l,i}) = \nu_0(\Delta_{0,i})$.

Partitioning Δ as mentioned, we get:

$$\begin{aligned}
\kappa(X) &= \nu(\Delta) \\
&= \sum_i \sum_{l=0}^{R_i-1} \nu(\Delta_{l,i}) \\
&= \sum_i \sum_{l=0}^{R_i-1} \nu_0(\Delta_{0,i}) \\
&= \sum_i R_i \nu_0(\Delta_{0,i}) \\
&= \sum_i \int_{\Delta_{0,i}} R \, d\nu_0 && \text{since } R \text{ is constant and equal to } R_i \text{ on } \Delta_{0,i} \\
&= \int_{\Delta_0} R \, d\nu_0 && \text{since } \{\Delta_{0,i}\}_i \text{ partitions } \Delta_0
\end{aligned}$$

If we have derived κ without using a Young tower, this works out to:

$$\kappa(X) = \int_{X_0} R_{X_0} \, d\kappa_0.$$

[10]Note that κ_0, as a measure on the base, can be seen both as a measure on X_0 and on Δ_0, hence our ability to push it through two different models.

We also note that by partitioning Δ by $\{\Delta_l\}_l$, and using the fact that ν_0 is a probability measure, we can similarly show that κ is always σ-finite.

We have proved:

Proposition 2.9. *Let $(X, \mathcal{E}, m)$ be a finite measure space and $f : X \circlearrowleft$ be measurable. Suppose (X, f) admits a Young tower representation (see Definition 2.3) as (Δ, F) with base X_0, but not necessarily satisfying **Y2**–**Y5**. Suppose further that there exists an f^R-invariant measure $\kappa_0 \ll m$.*

Let κ be the pushed invariant measure on X as defined in (2.3). Then κ is σ-finite, and furthermore, it is finite if and only if

$$\int_{X_0} R \, d\kappa_0 < \infty.$$

Conventionally, when the measure is indeed finite, we normalise it by dividing through by the integral (i.e. $\kappa(X)$). We may refer to this as $\bar{\kappa}$, and this is a probability measure.

Note that this condition on finiteness of the invariant measure is equivalent to **Y5**! So we have discovered the importance of Young's "return time integral" condition.

We are done with Young towers for now, but we should keep in mind that the construction we have studied in this section is a useful symbolic representation for a system (X, f). The latter subsections have also shown that although the representation via a tower (Δ, F) may not be in bijection with X, it is close enough that we can make inferences about X by applying the tower projection. Deciding how well properties translate between spaces will also be a question we will need to ask in the next section, where we move on to a different symbolic representation.

2.2 Markov maps and Markov shifts

We saw in section 1.3 how to construct a symbolic dynamical system called a CMS, where points in the state space are infinite walks along a topological transition graph given by a matrix A of zeros and ones. We will now see how CMS can be used to symbolically model dynamical systems on other spaces. The general idea will be to break down the state space of some dynamical system (X, f) into an at most countable partition, and then only care about the partition element that each $x \in X$ lies in, rather than its precise location. In a sense, we are discretising the system.

Some restrictions are necessary on the partition to ensure that the resulting CMS appropriately "represents" f. Different sources will do this in different ways. For a taste of the general concept, I refer the reader to [Aar97, Chapter 4], which provides a definition appropriate for abstract measure spaces. To simplify the setting, in this section we will only consider definitions of Markov concepts appropriate for interval maps, and to do this we take inspiration from [BG97, Chapter 9]. Interval maps will be the main point of focus for the rest of this section (and indeed also in the next chapter).

2.2.1 Definition

Definition 2.10 (Markov Transformation)**.** Let X be an interval and $\mathcal{B}(X)$ be the Borel σ-algebra on X. Let f be a transformation of the measure space $(X, \mathcal{B}(X), m)$, where m is equivalent to Lebesgue measure.[11] Let $\mathcal{P} = \{P_i\}_{i \in S} \subseteq \mathcal{B}(X)$ be a non-trivial and at most countable partition of X into intervals of positive m-measure. Then f is a *Markov transformation* with respect to $\mathcal{P}$, and $\mathcal{P}$ is a *Markov partition* for f, if:

1. (Bicontinuity). For each $P \in \mathcal{P}$, $f|_P$ is a continuous bijection with continuous inverse (i.e. $f|_P$ is a homeomorphism);

2. (Markov Condition). $\forall i, j \in S$, either $P_j \cap f(P_i) = \emptyset$ or $P_j \subseteq f(P_i)$;

3. (Shift Condition). $\forall s \in S, \exists i, j \in S$ such that:

$$\begin{aligned} P_i &\subseteq f(P_s) \\ P_s &\subseteq f(P_j). \end{aligned}$$

Furthermore, the Markov partition $\mathcal{P}$ is called a *strong generator* if:

4. (Strong Generation). $\bigvee_{i=0}^{\infty} f^{-i}\mathcal{A}(\mathcal{P}) \overset{\circ}{=} \mathcal{B}(X)$, where $\mathcal{A}(\mathcal{P})$ is the algebra of all finite unions of partition sets in $\mathcal{P}$.

Sometimes we will refer to this Markov transformation as $(X, \mathcal{B}(X), m, f, \mathcal{P})$.

Remark 2.11 (Interpretation of definition).

1. Bicontinuity is a niceness condition on f which allows us to talk about the inverse function $(f|_P)^{-1}$ of f on each branch $P \in \mathcal{P}$. This condition also ensures that the image of any Markov partition element is not only a union of intervals: it is in fact connected, and so is itself an interval.

2. The Markov condition is the main condition to be expected in this definition. It requires that every set in the partition maps cleanly onto other sets, and this will allow us to talk about going "from" some P_i "to" some P_j (which is possible if $P_j \subseteq f(P_i)$).

3. The shift condition is a technicality required here so that we can match Markov maps with the CMS defined in the previous chapter. In Definition 1.16 we assumed that the topological transition matrix contains at least one 1 in every row and every column, and the condition here is analogous to that: each partition set must be reachable from at least one (not necessarily distinct) set, and each partition set must map to at least one (not necessarily distinct) set. The map f would have to be silly for this condition not to be satisfied. I have not seen this condition used elsewhere, but will require it here for consistency across chapters.

4. Finally, the optional strong generation condition will come to light when we attempt to create a conjugacy between (X, f) and a CMS.

The following observation will help us characterise partition sets later.

Lemma 2.12. *The n^{th} partition of X into "cylinders" given by:*

$$\mathcal{P}_n := \bigvee_{i=0}^{n-1} f^{-i}\mathcal{P}$$

consists only of intervals (considering singletons as intervals).

[11] This means $\mu \ll$ Lebesgue and Lebesgue $\ll \mu$.

Proof. Proceed by induction on n. The claim is trivial for $n = 1$. For the inductive step note that $\mathcal{P}_{n+1} = f^{-1}\mathcal{P}_n \vee \mathcal{P}_n$. Any set in this partition is $f^{-1}Q \cap Q'$ for some $Q, Q' \in \mathcal{P}_n$. The function f is invertible on each branch $P \in \mathcal{P}$ by assumption, and these inverses are continuous. So, $f^{-1}Q = \cup_{P \in \mathcal{P}} (f|_P)^{-1}Q$, and each of these pre-images of Q is an interval (by the inductive hypothesis and by continuity of $(f|_P)^{-1}$) and a subset of P. (Depending on whether f maps each given $P \in \mathcal{P}$ onto Q, some of these pre-images may be empty, but in that case they do not matter.) Also, Q' is a subset of some $P' \in \mathcal{P}$. So,

$$f^{-1}Q \cap Q' = \underbrace{(f^{-1}Q \cap P')}_{\text{an interval}} \cap \underbrace{Q'}_{\text{an interval}}$$

so $f^{-1}Q \cap Q'$ is an interval (or empty), as required. □

2.2.2 The CMS conjugacy Φ

Next, as promised, we draw a link between Markov transformations and Markov shifts. Let $(X, \mathcal{B}(X), m, f, \mathcal{P})$ be a Markov transformation as in Definition 2.10. We construct a CMS based on this system using Definition 1.17.

In our setup, S is already an at most countable set, so take it to be our set of states and define the topological transition matrix $A = (a_{ij})$ to have entries:

$$a_{ij} = \begin{cases} 1 & P_j \subseteq f(P_i) \\ 0 & \text{otherwise.} \end{cases}$$

The shift condition ensures that A is a topological transition matrix, so we have created the CMS (Σ_A^+, σ). The idea is that points in x can be represented by points (sequences) in Σ_A^+, and we can do this via the following function. First, define $\tau : X \to S$ such that $\tau(x) = i$, where i is the unique element of S such that $x \in P_i$. Then, we consider:

$$\begin{aligned} \Phi : X &\longrightarrow \Sigma_A^+ \\ x &\longmapsto (\tau(x), \tau(f(x)), \tau(f^2(x)), \dots). \end{aligned}$$

One can verify that Φ does indeed map into Σ_A^+. To that end, let $y \in X$. Then $y \in P_{\tau(y)}$ and $f(y) \in P_{\tau(f(y))}$ by definition of τ. But then $f(y) \in P_{\tau(f(y))} \cap f(P_{\tau(y)})$, so the intersection is non-empty. This implies $a_{\tau(y)\tau(f(y))} = 1$. Set $y = f^i(x)$ for arbitrary i to conclude that the sequence $\Phi(x)$ is admissible in the CMS.

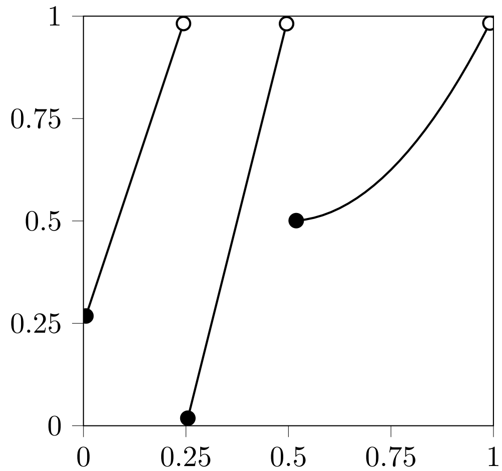
One further desirable property of Φ is that it conjugates the pair of dynamical systems it maps between, i.e. $\Phi \circ f = \sigma \circ \Phi$. This is easily checked: for $x \in X$, we have

$$\begin{aligned} \Phi(f(x)) &= (\tau(f(x)), \tau(f^2(x)), \tau(f^3(x)), \dots) \\ &= \sigma((\tau(x), \tau(f(x)), \tau(f^2(x)), \dots)) \\ &= \sigma(\Phi(x)). \end{aligned}$$

These initial properties alone give us confidence that representing a Markov transformation as a CMS is reasonable. Furthermore, if nothing else, we can keep in mind that drawing the transition graph of the CMS of f may provide us with some helpful intuition about where orbits go; the first few sections of the next chapter make heavy use of this intuition.

The remainder of this section has arisen from my initial concern when writing this project that Φ may conjugate the two systems, but could still be terrible at converting between the two in every other sense: for example, it could happen that Φ is not a bijection. Hence, I propose some further assumptions on the Markov transformation that ensure that Φ provides a fairly faithful conversion. I believe the claims here are not unlike the content of some textbooks, but the assumptions on those claims may be slightly different. The idea is to give a taste of what properties are needed for there to be a strong link between a map and its representation.

Example 2.13. Consider the interval map f on $X = [0, 1)$ with Lebesgue measure, given by the following graph:



It is piecewise continuous and invertible on the intervals $P_1 = [0, 1/4)$, $P_2 = [1/4, 1/2)$ and $P_3 = [1/2, 1)$. The images also map cleanly:

$$f(P_1) = P_2 \cup P_3; \qquad f(P_2) = P_1 \cup P_2 \cup P_3; \qquad f(P_3) = P_3.$$

The shift condition is easy to check and we get the following topological transition matrix:

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that in this case, $3 \in S$ is an absorbing state since the only valid transition from 3 is back to 3. In the original map, therefore, we expect orbits to eventually get trapped in the set P_3. This also implies that Φ is not almost everywhere injective in this case, since any $x \in P_3$ will map under Φ to the sequence $(3, 3, 3, \dots)$.

With this example, we may wonder whether the presence of an absorbing state is preventing injectivity of Φ. However, this is not the only problem, as the next example shows.

Example 2.14. Consider the rotation map by 1/2 on S^1, i.e. $f : [0, 1) \to [0, 1)$ given by $f(x) = x + 1/2 \bmod 1$. An appropriate Markov partition is $\{P_1 = [0, 1/2), P_2 = [1/2, 1)\}$ (this becomes clear if you draw the graph of f), and $f(P_1) = P_2$, $f(P_2) = P_1$. So the topological transition matrix is

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

which has no absorbing states. But, any $x \in X$ has a 2-periodic orbit, since $f^2(x) = x{+}1 \bmod 1 = x$. So every $x \in P_1$ has $\Phi(x) = (1, 2, 1, 2, \dots)$, and conversely every $x \in P_2$ has $\Phi(x) = (2, 1, 2, 1, \dots)$.

It turns out that ensuring Φ is a bijection is where strong generation comes in handy.

Proposition 2.15. *Let $(X, \mathcal{B}(X), m, f, \mathcal{P})$ be a Markov transformation as in Definition 2.10, giving rise to a CMS (Σ_A^+, σ). Suppose Σ_A^+ is equipped with a measure that assigns zero mass to singletons.*[12] *Then if $\mathcal{P}$ is a strong generator, Φ is almost everywhere bijective.*

Proof. Surjectivity actually holds regardless of the strong generator condition, as we will see here. Let $\boldsymbol{w} = (w_0, w_1, w_2, \dots) \in \Sigma_A^+$. We wish to find $x \in X$ such that $\Phi(x) = \boldsymbol{w}$, i.e. $x \in P_{w_0}, f(x) \in P_{w_1}, f^2(x) \in P_{w_2} \dots$. Equivalently, we need

$$x \in \bigcap_{n \geq 1} C_{\boldsymbol{w}}^n,$$

where $C_{\boldsymbol{w}}^n := \bigcap_{i=0}^{n-1} f^{-i} P_{w_i} \in \mathcal{P}_n$ is a cylinder set. Thinking about the arguments used in the proof of Lemma 2.12, since each transition in the word $\boldsymbol{w}$ is a valid transition of f on the partition $\mathcal{P}$, the cylinders $C_{\boldsymbol{w}}^n$ must be non-empty as finite intersections of preimages. The tricky part is proving that the infinite intersection remains non-empty. To do this, note that there are two cases. Firstly, it may be that one of the $C_{\boldsymbol{w}}^n$ is a singleton. However, in this case, every subsequent $C_{\boldsymbol{w}}^{n+i}$ is also a singleton, and so the intersection is non-empty. In all other cases, each cylinder in the sequence is an interval with more than one point, so the interior $\mathring{C}_{\boldsymbol{w}}^n$ of $C_{\boldsymbol{w}}^n$ is an open interval (a_n, b_n) for some $a_n < b_n$. Furthermore, we easily get $C_{\boldsymbol{w}}^1 \supseteq C_{\boldsymbol{w}}^2 \supseteq \dots$ by definition, so $\mathring{C}_{\boldsymbol{w}}^1 \supseteq \mathring{C}_{\boldsymbol{w}}^2 \supseteq \dots$, and so $(a_n)_n$ is increasing to a limit a while $(b_n)_n$ is decreasing to a limit b, and $a \leq b$. By real analysis, $\bigcap_n (a_n, b_n)$ is certainly non-empty if neither endpoint sequence is eventually constant[13] (in which case e.g. a lies in the intersection). But for each n, the endpoints a_n, b_n are necessarily endpoints of elements of the finite interval partition $\mathcal{P}_n$. So, there are only countably many eventually constant sequences $(a_n)_n$, and the same goes for $(b_n)_n$. Hence if $\boldsymbol{w}$ has no preimage under Φ, it is part of a zero measure set, and in all other cases we can find:

$$x = a \in \bigcap_{n \geq 1} \mathring{C}_{\boldsymbol{w}}^n \subseteq \bigcap_{n \geq 1} C_{\boldsymbol{w}}^n,$$

as required.

For injectivity, suppose $x, y \in X$ are such that $x < y$ and $\Phi(x) = \Phi(y) = \boldsymbol{w} = (w_0, w_1, \dots)$ for some $\boldsymbol{w} \in \Sigma_A^+$. Then as above, $x, y \in C_{\boldsymbol{w}}^n$ for all n. Since the $C_{\boldsymbol{w}}^n$ are all intervals by Lemma 2.12, this implies that each partition $\mathcal{P}_n$ contains no intervals that are proper subsets of (x, y). This must then also hold for the limit σ-algebra $\bigvee_{n=0}^{\infty} \mathcal{P}_n = \bigvee_{i=0}^{\infty} f^{-n} \mathcal{A}(\mathcal{P}) \overset{\circ}{=} \mathcal{B}(X)$ (the last equality being deduced from the strong generator condition). However, $\mathcal{B}(X)$ contains open proper subsets of (x, y) since (x, y) is open, a contradiction. □

[12] If we want the measure on Σ_A^+ to be something other than the zero measure, then Σ_A^+ should be uncountable, which is the case provided for example that A is topologically transitive and at least one state maps to at least two others.

This is good news, but the strong generator condition can be difficult to check. Fortunately, it is implied by a natural dynamical property.

Proposition 2.16. *Let $(X, \mathcal{B}(X), \mu, f, \mathcal{P})$ be a Markov transformation as in Definition 2.10, and suppose μ is an invariant probability measure equivalent to Lebesgue measure such that $(X, \mathcal{B}(X), \mu, f)$ is weakly mixing. Then $\mathcal{P}$ is a strong generator for $\mathcal{B}$.*

Proof. Suppose this is not the case. We know that if the diameter of a sequence of Borel partitions tends to 0, then the σ-algebra they generate is the Borel algebra $\mathcal{B}(X)$. So, it must be that $\operatorname{diam}(\mathcal{P}_n) \not\to 0$. Since each $\mathcal{P}_n$ is a refinement of the previous partition, the diameter sequence must necessarily decrease, so we must have $\operatorname{diam}(\mathcal{P}_n) \searrow \delta > 0$. Also because the partitions are refinements, this means we can find a sequence of sets $K_n \in \mathcal{P}_n$ such that $\forall n \geq 1 : K_{n+1} \subseteq K_n$ and $\operatorname{diam}(K_n) \geq \delta$. Furthermore, by Lemma 2.12, the K_n are intervals. Let $K = \cap_n K_n$. Then K is an interval of diameter at least δ. Also, K is a subset of K_1, an element of $\mathcal{P}$ which we will refer to as P.

Let $\mathcal{F} := \bigvee_{i=0}^{\infty} \mathcal{P}_n$. Then certainly $\mathcal{F} \subseteq \mathcal{B}(X)$, but $\mathcal{F}$ cannot contain proper subsets of K, since none of the $\mathcal{P}_n$ do. Since μ is equivalent to Lebesgue measure and K is an interval of positive measure, we have $k := \mu(K) > 0$. Furthermore, $k \leq \mu(P)$, but since $\mathcal{P}$ is non-trivial, it contains other sets of positive Lebesgue measure: combined with the fact that X is an interval equipped with Lebesgue measure, we conclude that $\mu(P) < \mu(X) = 1$. So, $0 < k < 1$.

Now, let $Q \in \mathcal{P}$ be any partition element. By the same arguments, $0 < q := \mu(Q) < 1$. Let $\varepsilon > 0$. Since the system is weakly mixing and K, Q are measurable, $\exists n \geq 1$ such that

$$|\mu(f^{-n}Q \cap K) - qk| < \varepsilon.$$

But $f^{-n}Q \cap K \in \mathcal{F}$ and this intersection is a subset of K. By our assumptions, this would mean that either $f^{-n}Q \cap K = K$ or $f^{-n}Q \cap K = \emptyset$, and this implies $\mu(f^{-n}Q \cap K) \in \{0, k\}$. But $0 < qk < k$, so picking ε small enough gives a contradiction. □

These two propositions should give us some confidence that the conjugacy $\Phi : X \to \Sigma_A^+$ is a reasonable tool to use; some ergodic problems are more easily solved in a symbolic setting, and we can convert problems on the interval into symbolic problems if we are careful with the conversion.

Note that we have not yet assigned a measure to Σ_A^+, and it is not essential to do so for the purposes of this project. However, if we have a measure μ on X, one natural choice for a measure on Σ_A^+ could be $\mu \circ \Phi^{-1}$. Furthermore, in the cases where Φ is a bijection, we can actually do the reverse: if we have a measure ν for Σ_A^+, then $\nu \circ \Phi$ is a measure on X. This could be a starting point for further investigation but we will not do this here.

When f is also assumed to be linear on each $P \in \mathcal{P}$, the behaviour of the system is close enough to a CMS that f can be studied using matrices with entries related to the distortion of each interval. This is done in [BG97, Chapter 9].

Remark 2.17. Looking back at the Young tower construction from section 2.1, and the representations of towers that we drew in Figure 2.2 and Figure 2.3, we might notice that Young towers and CMS are both methods of symbolically "discretising" the system. Technical conditions in Definition 2.2 aside, Young towers are a way of breaking up the state space into a countable number

[13] This is a sufficient but not necessary condition; e.g. $a < b$ would also work, but this actually cannot happen here since $\mathcal{P}$ would not be a strong generator.

of sections. If we view the base set Δ_0 as one state, for example, we can construct a CMS where the remaining states are $\Delta_{l,i}$ for $l \geq 1$, and there are transitions leading from each $\Delta_{l,i}$ to $\Delta_{l+1,i}$, or from $\Delta_{l,i}$ back to Δ_0 when the top of the tower is reached. The base Δ_0 is then the only state which can have multiple admissible transitions, and leaving Δ_0 makes an orbit $O(x)$ embark on a predictable loop which inevitably leads back to Δ_0 after "$R(x)$" steps. The non-uniqueness of representation from subsection 2.1.4 remains an issue that is less prominent with CMS, but this is nevertheless a good excuse to think about the similarities between the models—especially if the Markov partition we choose is somehow also linked to inducing, which it will be in the next chapter.

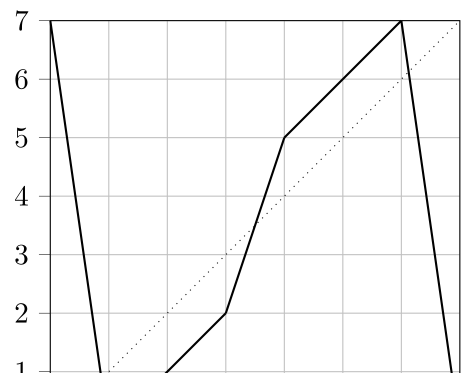
2.2.3 Looking for intermittency in a CMS

Up until this point, this section has contained ideas applicable to all kinds of systems, intermittent or not. However, the CMS representation of a map allows us to look at the rough movement of orbits, and so we may be able to search for certain properties of the transition graph of a Markov map in order to characterise it as intermittent. In this section I propose a heuristic for this, along with an example.

Recall that we defined intermittency as being the alternation of laminar phases with chaotic bursts of motion (Definition 1.29). With walks on a graph, a laminar phase could be interpreted as a sequence of states each with only one admissible transition, leading from one to the next. On the other hand, for a chaotic burst, we can look for states that have multiple admissible transitions.

Intermittency is a qualitative concept so this is not a perfect characterisation, but it will in some sense prove true in the next chapter when we study the Manneville-Pomeau map. For now, let's design a map where we can distinguish between these behaviours.

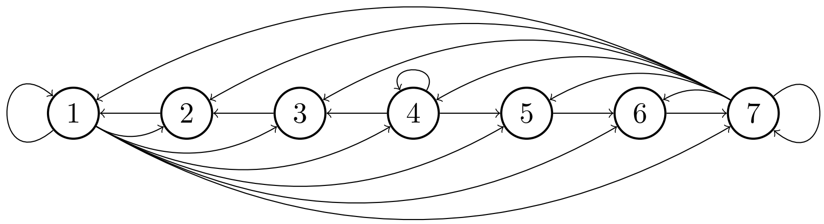
Example 2.18. Consider the function $f : [0, 7] \to [0, 7]$ given by the following graph:



We can show without too much difficulty that an appropriate Markov partition for f on $[0, 7]$ is $\mathcal{P} = \{P_i\}_{i \in S}$, where $S = \{1, 2, \dots, 7\}$ and:

$$P_1 = [0, 1); \quad P_2 = [1, 2); \quad \dots \quad P_6 = [5, 6); \quad P_7 = [6, 7].$$

Calculating the image of each set gives us the following graph.



This seems to show that an orbit passing through the cylinder [4] can stay there for a certain number of iterates (since $(4,4)$ is an admissible word), but once it leaves, its path must be either $4 \to 5 \to 6 \to 7$ or $4 \to 3 \to 2 \to 1$. This part of the orbit is determined and hence we might view it as being "laminar". On the other hand, only when an orbit reaches the end of this part of the flow does it arrive in either state 7 or state 1; in both cases it can be mapped from there to anywhere in the space. This is our chaotic step or "bursting region".

It seems as though the bursting regions correspond to areas where $|f'|$ is large in this case, and this should be expected: the map is expanding here (i.e. it has high derivative, so points close together are mapped $|f'| = 7$ times further apart than they were).

Another important observation on this map is that $\{[0,1),[1,6),[6,7]\}$ is also an appropriate Markov partition, and the associated CMS is the full shift on three symbols. However, this somehow gives us less information, because the laminar behaviour inside $[1,6)$ is factored out. Also, f is not piecewise linear with respect to this partition. In any case, we should keep in mind that there are often several valid choices of partition, and we should select carefully based on what we are trying to show.

We should be cautious about jumping to conclusions about the intermittency of a system based on a CMS representation. The discussion above remains heuristic, and in fact we can find Markov partitions for decidedly non-intermittent systems that make them look like periods of laminar flow are possible. There is an example of this in the next chapter with the doubling map; the key there is that while laminar flow is possible, the laminar region of the graph is mapped to only rarely. Hence, we should consider the relative sizes of the partition elements with respect to a measure of interest, such as Lebesgue, to see whether the laminar regions we discover may only correspond to a very small part of the space.

Alternatively, we might conclude that the takeaway from this chapter is that there is a little bit of intermittency in everyone.

Chapter 3

The Manneville-Pomeau Map

In this chapter, our aim is to apply some of the notions discussed in previous chapters to a concrete example. One of the simplest examples of a dynamical system with intermittent properties is the Manneville-Pomeau map, of which there are multiple parametrisations.

Over the course of this chapter, we will discover many classical results from the literature about the Manneville-Pomeau map that relate to its intermittent properties, and will present proofs of these results where possible. First, we should agree on an appropriate definition for this map.

3.1 Definition

The term *Manneville-Pomeau map* is used to refer to a class of interval maps that are topologically similar to the famous doubling map $x \mapsto 2x \bmod 1$, but with a tangency at the fixed point at the origin, which distinguishes them from faster mixing systems.

This means that at the fixed point $x = 0$, the derivative of the map is 1, and things move slower under iterated applications of the map than they do elsewhere in the space. We call this point a *neutral fixed point* or an *indifferent fixed point*.

We'll see that this is what creates the intermittent behaviour we're looking for.

Throughout this chapter, we'll use the definition of the Manneville-Pomeau map from [LSV99]. This leads some sources to refer to this parametrisation as the *Liverani-Saussol-Vaienti map* or *LSV map*. Let $I := [0, 1]$, and define $T_\alpha : I \to I$ as follows:

$$T_\alpha(x) = \begin{cases} x(1 + 2^\alpha x^\alpha) & x \in [0, 1/2) \\ 2x - 1 & x \in [1/2, 1]. \end{cases}$$

It will be helpful to have names for these two branches, so let $u : [0, 1/2) \to I$ be given by $u(x) = x(1 + 2^\alpha x^\alpha)$ and $v : [1/2, 1] \to I$ be given by $v(x) = 2x - 1$.

Notice that the tangency in the first half of the interval can be controlled using the parameter α. When $\alpha = 0$, this is just the doubling map; as α increases, the tangency gets stronger. For $\alpha > 0$, the derivative T'_α is equal to 1 at $x = 0$, making the curve tangent to the $y = x$ line.

Let's plot the Manneville-Pomeau map for a few values of α, including the $\alpha = 0$ case (doubling map) for comparison.

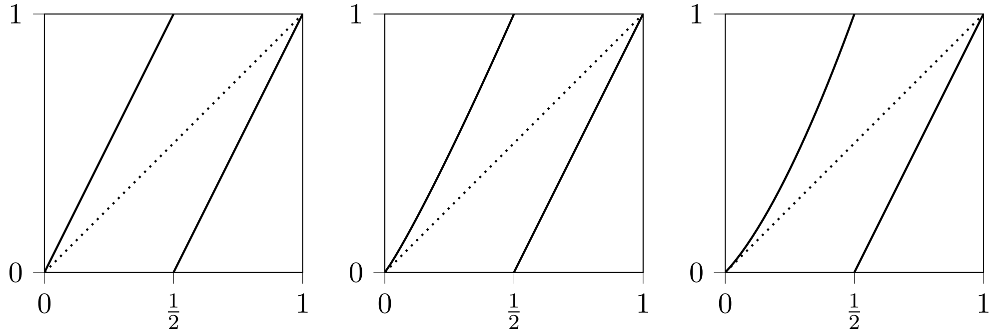


Figure 3.1: Plots of the Manneville-Pomeau map on $[0,1]$ for $\alpha = 0, \frac{1}{4}, \frac{3}{4}$.

To study this map, we will place ourselves in the usual σ-algebra for the unit interval: $([0,1], \mathcal{B}, \lambda)$ where λ denotes Lebesgue measure and $\mathcal{B}$ denotes the Borel algebra. From this point onwards, unless stated otherwise, all relations between sets (such as equalities and inclusions) are considered true as long as they hold up to a λ-zero set.

3.2 A Markov representation for T_α

To understand this interval map better, we begin by finding a suitable Markov partition—hopefully one which will demonstrate the system's slow dynamics about the origin. We can do this inductively, beginning with the right-hand half of the interval $I_0 := [1/2, 1]$. This set maps bijectively onto I under T_α, and we now seek to suitably partition the other half of the interval $I \setminus I_0$.

To demonstrate the flow away from 0, we can imagine the orbit of any point $x \in [0, 1/2)$ as being an increasing sequence of points $x \le T_\alpha(x) \le T_\alpha^2(x) \le \dots$ that eventually reaches I_0, where it is then mapped to some other (unpredictable) point in the space. Then, the final point before the orbit reaches I_0 necessarily lies in $T_\alpha^{-1} I_0 \cap [0, 1/2)$. The point before that lies in the pre-image of *that* set, i.e. $T_\alpha^{-1} I_1 \cap [0, 1/2)$, and so on. The endpoints of each interval in this sequence cannot be expressed algebraically for general α, but it can be shown without too much difficulty that they are the pre-images of the point $1/2$. Therefore, we define the following sequence $(x_n)_n \subseteq I$:

$$x_0 = 1, \quad x_1 = \frac{1}{2}, \quad x_{n+1} = u^{-1}(x_n) \; \forall n \ge 1.$$

Our pre-image intervals are then:

$$I_0 = [1/2, 1], \quad I_n = [x_{n+1}, x_n) \; \forall n \ge 1.$$

Note that $x_n \to 0$ as $n \to \infty$, so the union of these intervals is in fact I (without $x = 0$, but this doesn't matter). The first few elements of the partition are plotted below:

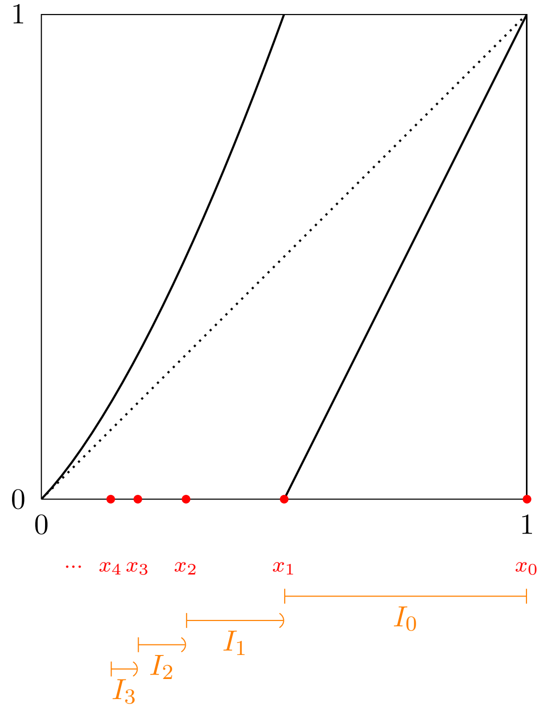


Figure 3.2: The first few intervals of the Markov partition for $\alpha = \frac{3}{4}$.

By construction, the intervals in our partition have the following images:

$$T_\alpha I_0 = I = \bigcup_{k=0}^{\infty} I_k; \quad T_\alpha I_n = I_{n-1} \; \forall n \geq 1.$$

We can then easily check the conditions of Definition 2.10 and conclude that $(I, \mathcal{B}, \lambda, T_\alpha, \{I_0, I_1, \dots\})$ is a Markov transformation. This allows us to represent the system as a countable Markov shift, using the technology from section 2.2. We can also convince ourselves that the Markov partition $\{I_i\}_{i\geq 0}$ that we have chosen for I is a strong generator. So by Proposition 2.15, Φ is a bijection up to a countable set of exceptions for surjectivity. The topological transition matrix A for this CMS is the adjacency matrix of the following transition graph:

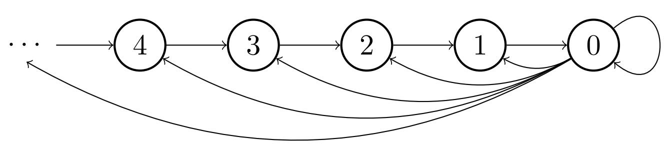


Figure 3.3: The transition graph of the CMS corresponding to T_α.

A CMS with the above transition graph is called a *renewal shift*, since the system "resets" every time the 0 state is reached. Choosing this representation allows us to look at the dynamics from the point of view of the second half-interval I_0. From there, any orbit leaving I_0 looks like it jumps away by a certain amount (the closer the orbit is to the fixed point at $x = 0$, the further away it is from I_0) before steadily returning to I_0, interval by interval. Although this is one of many possible Markov representations, looking at T_α as a renewal shift will be a useful perspective to have in subsequent sections.

Note that expressing T_α as a renewal shift says nothing about its intermittent properties. In fact, the construction of $(I_n)_n$ above is possible for any value of $\alpha \geq 0$. Namely, choosing $\alpha = 0$ allows us to express the doubling map $x \mapsto 2x \bmod 1$ as a renewal shift, despite this map being a textbook example of a chaotic, non-intermittent system.

However, the $\alpha = 0$ and $\alpha > 0$ partitions do differ in their asymptotics. When $\alpha = 0$, the sequence $(x_n)_n$ is easily computed exactly to be

$$x_n = 2^{-n},$$

so an exponentially decreasing sequence. On the other hand, it is well-known (see for example Lemma 2.1 in [Iso95]) that for $\alpha > 0$ we have

$$x_n = (\alpha 2^\alpha (n+1))^{-1/\alpha}(1 + \mathcal{O}(n^{-1})), \tag{3.1}$$

which is only polynomial. This is not the only time that a contrast between polynomial and exponential decay will appear in this chapter.

Remark 3.1. This is not the only possible CMS we can get from T_α; for example, a simpler choice would come from a partition of I into $\{[0, 1/2), [1/2, 1]\}$. This is also a Markov partition for the system, but the lower resolution means that less inferences will be possible. There is a qualitative reason for this similar to what we noted in example 2.18: we will not be able to see the laminar flow that we can see here. On a more technical level, results that we can prove for a lower-resolution CMS will in general not translate well to the original system because of the *distortion* this map presents. Making this rigorous requires thermodynamic formalism; the idea is that if we pick the standard potential $\phi = -\log|T'_\alpha|$, then this partition will not give *summable variations* of ϕ (this means ϕ varies too much on the depth-n cylinders for the series given by supremum of these variations on each level to be finite). More detail about summable variations can be found in [Sar99].

3.3 Finding an invariant measure

We now turn to the classical ergodic problem of finding an invariant measure for T_α. Remember that out of the many possible invariant measures, there are many uninteresting ones, such as $\frac{1}{2}(\delta_0 + \delta_1)$ (0 and 1 being the fixed points of the map we are working with). So to narrow things down, we wish to find an invariant measure that is absolutely continuous with respect to Lebesgue measure λ. Ideally we would also like it to be a probability measure.

There are several ways to prove the existence of an a.c. invariant measure. A non-constructive proof in [LSV99] is able to establish some basic properties of this measure (or rather, its density) which we will discuss later—but this uses transfer operators, and we will instead exhibit a more elementary proof which makes good use of the inducing scheme method seen in section 2.1. The general idea is to try to factor out the "problem point" at $x = 0$ where the derivative is 1, which is preventing the system from expanding nicely everywhere, by picking a base set away from 0. We will in fact not need Young towers for this proof, working instead with the direct pushing formula in (2.3). We will, however, need to induce on a base set. The rest of this section demonstrates a standard inducing-pushing technique that I have worked through with guidance from my supervisor Dr Mike Todd, with a historical interlude as we invoke a powerful theorem from ergodic theory.

Fix $\alpha \in (0,1)$. Our first task is to find a nice subset of $[0,1]$ to induce on. To construct the CMS in the previous section we looked at things "from the point of view" of I_0, and the same strategy will work here; pick $A := I_0 = [1/2, 1]$.

Let R denote the return time function $R_A : A \to \mathbb{N}$ under T_α, and consider now the induced system $T_A := T_\alpha^R|_A$ as in previous examples, i.e.

$$\begin{aligned} T_A : A &\longrightarrow A \\ x &\longmapsto T_\alpha^{R(x)}(x). \end{aligned}$$

This produces an induced system $(A = [1/2, 1], T_A)$, and we will use Lebesgue measure restricted to this interval (but not normalised) as the base measure here, so that $\lambda(A) = 1/2$. To find an a.c. invariant measure for T_α, it now suffices to find one for T_A and then push it back to T_α. But what makes T_A any easier to deal with than T_α?

The key is that T_A turns out to be a *full-branched expanding map*, and we will define this below. Note that this terminology is not standard and different papers will provide similar, but not identical, definitions for similar terms; this is adapted from [Tod20] but slightly generalised in favour of a historical note.

Definition 3.2 (Full-Branched Expanding Map)**.** We say that an interval map $f : J \to J$ (for some interval $J \subseteq \mathbb{R}$) is a *full-branched expanding map* or *FBEM* if:

1. (Full-Branched Condition). There exists an at most countable partition $\mathcal{P} = \{J_i\}_i$ of J such that each J_i is an interval, f is continuously differentiable on the closure of each J_i, and $f(J_i) = J$;

2. (Bounded Distortion Condition). Letting $\mathcal{P}_n := \bigvee_{j=0}^{n-1} f^{-j}\mathcal{P}$ denote the set of n-cylinders, we have
$$\sup_{n\geq 1} \sup_{Z\in\mathcal{P}_n} \sup_{x,y\in Z} \left| \frac{(f^n)'(x)}{(f^n)'(y)} \right| < \infty;$$

3. (Expanding Condition). There exists $\gamma > 1$ such that $|f'(x)| \geq \gamma$ for all $x \in J$ (taking appropriate derivatives at boundary points).

The first and third conditions reflect the properties alluded to in the name "FBEM", while the second—the bounded distortion condition—imposes an additional technical restriction. Quoting from [Adl73], this condition can be viewed as measuring the map's "departure from linearity": if f is piecewise linear then the supremum will equal 1, while a non-constant derivative will drive the supremum upwards.

Since bounded distortion is difficult to check directly because of the need to bound the derivatives of arbitrary iterates of f, more straightforward tests that imply the same results are used instead. Usually, one of the following works:

D1. f is uniformly $1 + \beta$-Hölder on each branch J_i, for some choice of $\beta > 0$;

D2. (Adler-type condition). f is (not necessarily uniformly) C^2 on each J_i, and the following holds:
$$\sup_{x\in I} \frac{|f''(x)|}{|f'(x)|^2} < \infty;$$

Proof that **D1** implies the bounded distortion condition can be found in [Tod20], while justification for **D2** is given in [Adl79].

We might notice that T_α itself is very nearly a FBEM, failing only on the final condition: we cannot uniformly bound the derivative away from 1. We might, however, have more luck with T_A. The reason we want to prove this is that we will be able to apply the following powerful theorem on expanding maps:

Theorem 3.3 (Ergodic Folklore Theorem)**.** *Every FBEM f has an acip μ. In fact, the sequence of measures $(\mu_n)_n$ given by:*

$$\mu_n = \frac{1}{n}\sum_{j=0}^{n-1} f_*^j \lambda$$

has a subsequence $(\mu_{n_k})_k$ which converges weak to μ.*[1]

The earliest account of this theorem is difficult to track down. Furthermore, slight variations on the FBEM assumptions that do not change the conclusion are common (one important one being that we can weaken the full-branched assumption to a Markov condition, but we will not need that here). The literature often cites Adler [Adl73], but in his famous afterword to Bowen [Adl79], Adler provides a chain of mathematicians who in turn had discovered the theorem before him. Within the field of ergodic theory, this has brought about the name of "Folklore Theorem", although since many theorems in mathematics go by that name, it is perhaps safer to refer to it here as the *Ergodic* Folklore Theorem.

For our purposes, a good proof of the Ergodic Folklore Theorem can be found in [Tod20], and we will not repeat it here; however, the idea of the proof is to show firstly that the subsequence above does indeed converge to an invariant measure (using weak* compactness of the space of invariant measures), and secondly that this limit measure is absolutely continuous (using the bounded distortion condition).

Returning now to the case of the inducing scheme T_A, we are hoping to show that it is a FBEM so that this theorem may apply. The hardest part here will be the bounded distortion condition; to tackle it, proving that the "Adler condition" **D2** applies will be a good choice, since this condition is fairly robust under iterates, and T_A is defined through iterates of T_α. We will work through the proof of this in more detail than is usually given when checking this sort of condition in the literature. Because we are expecting polynomial rather than exponential decay, the bounds we have to establish in order to verify the condition are quite tight, and we will need to follow orbits very closely to achieve this.

Lemma 3.4. *T_A is a FBEM.*

Proof. The following proof is mostly an exercise in basic analysis and algebraic manipulation, but demonstrates the sort of argument that may in general need to be provided in order to show that a given map is indeed a FBEM. However, note that the exact bounds used in the latter part of the proof are specific to this map T_A.

Begin by partitioning A by return time. For $n \geq 0$ let

$$A_n = \{x \in A \mid R(x) = n + 1\}.$$

[1] Saying that a sequence of measures converges weak* to another measure means that for any continuous function φ, the sequence of integrals of φ with respect to each measure converges to the integral of φ with respect to the limit measure.

We can show that

$$A_0 = [3/4, 1], \quad A_n = T_\alpha^{-1} I_n \cap [1/2, 1] \ \forall n \geq 0,$$

which shows that this is a partition into intervals. Similarly to the Markov partition in the previous section, the endpoints here are another sequence of pre-images of the point 1, namely

$$a_n = v^{-1}(x_n)$$

for each $n \geq 0$. Then, our partition intervals are

$$A_n = [a_{n+1}, a_n)$$

for $n \geq 0$ (with a closed interval in the case $n = 0$ only).

Thinking about the first few steps of the orbit of $x \in A_n$ through the transition graph in Figure 3.3, we also find that on A_n, we have $T_A(x) = u^n(v(x))$. This is continuous, and at the endpoints of A_n we have:

$$T_A(a_{n+1}) = u^n(x_{n+1}) = 1/2$$
$$\lim_{x \to a_n^-} T_A(x) = \lim_{x \to x_n^-} u^n(x) = 1$$

So this map is indeed full-branched on $[1/2, 1]$. A numerical plot is given below; each A_n is the domain of the $(n+1)^{\text{th}}$ branch from the right.

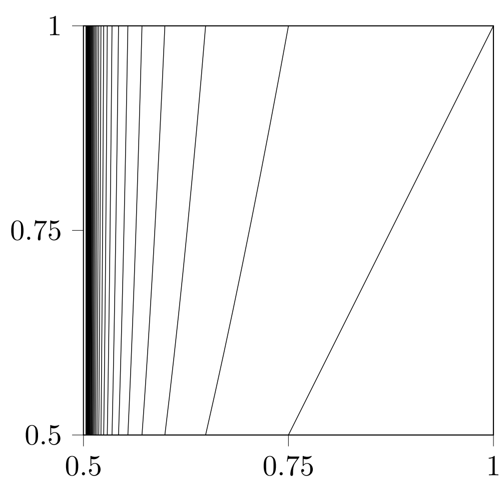


Figure 3.4: The induced map T_A for $\alpha = 3/4$.

The final condition from the definition is also easy to show. Note that $T'_\alpha(x) \geq 1$ everywhere, and T_A always begins with an application of v, so iterating the chain rule gives that T'_A is a product of v' with some number of T'_α. But $v' = 2$, so picking $\gamma = 2$ works. This is also evident from the plot above.

Now we need to prove bounded distortion and we will do this using the Adler condition. For any function f, let $\text{Adl}(f) := |f''/(f')^2|$, so that we are trying to bound the function $\text{Adl}(T_A)$ on A.

The idea is to find an expression for $\text{Adl}(T_\alpha^n)$ for each n, since taking powers of the original map is how we get T_A. The approach that gives $\text{Adl}(f^n)$ for any n is given, for example, in the proof of

[Coa20, Theorem A], although it's done there for a different map. The trick is to repeatedly apply the chain rule and the triangle inequality to get that in our case, for any $n \geq 0$,

$$\mathrm{Adl}(T_\alpha^n) \leq |\,\mathrm{Adl}(T_\alpha) \circ T_\alpha^{n-1}| + \sum_{k=1}^{n-1} \left| \frac{\mathrm{Adl}(T_\alpha) \circ T_\alpha^{k-1}}{(T_\alpha^{n-k})' \circ T_\alpha^k} \right|.$$

This will appear in the bound we are about to establish. In preparation, we should find bounds for $\mathrm{Adl}(T_\alpha)$ and $(T_\alpha^l)'$ on each Markov interval I_n. For every $n \geq 1$, we have $T|_{I_n} = u$, and we can compute derivatives:

$$\begin{aligned} u(x) &= x(1 + 2^\alpha x^\alpha) \\ u'(x) &= 1 + (1+\alpha)2^\alpha x^\alpha \\ u''(x) &= \alpha(1+\alpha)2^\alpha x^{\alpha-1} \end{aligned}$$

Remember that throughout this section we assumed $\alpha \in (0,1)$, so u, u' are increasing while u'' is decreasing. All functions are positive. So on I_n, we can bound these derivatives using the values of u', u'' at the endpoints x_{n+1} and x_n of the interval. This is straightforward for $\mathrm{Adl}(T_\alpha)$ and gives us:

$$\sup_{I_n} \mathrm{Adl}(T_\alpha) \leq \frac{\sup_{I_n} |u''|}{\inf_{I_n} |u'|^2} \leq \frac{u''(x_{n+1})}{u'(x_n)^2} \leq \alpha(\alpha+1)2^\alpha x_{n+1}^{\alpha-1}. \tag{3.2}$$

To bound $(T_\alpha^l)'$ from below, we iterate the chain rule. Let $x \in I_n$. We are only interested in the case $n = l$, since this is what will arise when considering the induced map T_A; so assume $n = l$.

$$\begin{aligned} (T_\alpha^l)'(x) &= (u^l)'(x) \\ &= u'(\underbrace{x}_{\in I_n}) \cdot u'(\underbrace{u(x)}_{\in I_{n-1}}) \cdot u'(\underbrace{u^2(x)}_{\in I_{n-2}}) \cdot \dots \cdot u'(\underbrace{u^{l-1}(x)}_{\in I_{n-l+1}}) \\ &\geq u'(x_{l+1}) \cdot u'(x_l) \cdot u'(x_{l-1}) \cdot \dots \cdot u'(x_2) \qquad \text{(since } n = l) \\ &= (1 + (1+\alpha)2^\alpha x_2^\alpha)(1 + (1+\alpha)2^\alpha x_3^\alpha) \dots (1 + (1+\alpha)2^\alpha x_{l+1}^\alpha). \end{aligned}$$

Denote this product by t_{l+1}, so that we get

$$\inf_{I_l} (T_\alpha^l)' \geq t_{l+1}. \tag{3.3}$$

We are now ready to compute $\mathrm{Adl}(T_A)$. Suppose $x \in A$ has return time $n+1$, so that $x \in A_n$. If $n = 0$, then $T_A(x) = v(x) = 2x - 1$ and so $\mathrm{Adl}(T_A)(x) = 0$. Turning to the case $n \geq 1$, we recall from above that $T_A(x) = T_\alpha^{n+1}(x) = u^n(v(x)) = T_\alpha^n(v(x))$. So,

$$\begin{aligned} \mathrm{Adl}(T_A)(x) &= \left| \frac{(T_\alpha^n \circ v)''(x)}{(T_\alpha^n \circ v)'(x)^2} \right| \\ &= \left| \frac{\cancel{v'(x)^2} \cdot (T_\alpha^n)''(v(x))}{\cancel{v'(x)^2} \cdot (T_\alpha^n)'(v(x))^2} \right| \qquad \text{(by chain and product rules and since } v'' = 0) \\ &= \mathrm{Adl}(T_\alpha^n)(v(x)) \\ &\leq |\,\mathrm{Adl}(T_\alpha)(T_\alpha^{n-1}(v(x)))| + \sum_{k=1}^{n-1} \left| \frac{\mathrm{Adl}(T_\alpha)(T_\alpha^{k-1}(v(x)))}{(T_\alpha^{n-k})'(T_\alpha^k(v(x)))} \right|. \end{aligned}$$

Since the induced map and its derivatives are all positive, we can drop the absolute values.

At this point, notice that we are applying functions to points of the form $T_\alpha^l(v(x))$ for some l. But $x \in A_n \implies v(x) \in I_n \implies T_\alpha^l(v(x)) \in I_{n-l}$, so applying our estimates in (3.2) and (3.3), this yields:

$$\begin{aligned}
\mathrm{Adl}(T_A)(x) &\le \sup_{I_1} \mathrm{Adl}(T_\alpha) + \sum_{k=1}^{n-1} \frac{\sup_{I_{n-k+1}} \mathrm{Adl}(T_\alpha)}{\inf_{I_{n-k}} (T_\alpha^{n-k})'} \\
&\le \alpha(\alpha+1)2^\alpha x_2^{\alpha-1} + \sum_{k=1}^{n-1} \frac{\alpha(\alpha+1)2^\alpha x_{n-k+2}^{\alpha-1}}{t_{n-k+1}} \\
&\le \alpha(\alpha+1)2^\alpha x_2^{\alpha-1} + \sum_{k=2}^{n} \frac{\alpha(\alpha+1)2^\alpha x_{k+1}^{\alpha-1}}{t_k} \qquad \text{(relabelling } n-k+1 \to k\text{)}
\end{aligned}$$

Many terms in this bound are independent of n. Filtering these out, to show that $\mathrm{Adl}(T_\alpha)$ is bounded uniformly on every interval A_n, it suffices to show that the series

$$\sum_{k=2}^{\infty} \frac{x_{k+1}^{\alpha-1}}{t_k}$$

converges.

We will analyse the asymptotics of the k^{th} term of the series by taking logarithms. We will also use the estimates for x_k given in Equation 3.1. In this next calculation, we write $a_k \approx b_k$ if the sequence $|a_k - b_k|$ is bounded (since we are working with logs, this corresponds to the original sequences differing by a multiplicative constant).

$$\begin{aligned}
\ln(x_{k+1}^{\alpha-1}/t_k) &= \ln(x_{k+1}^{\alpha-1}) - \ln(t_k) \\
&= (\alpha-1)\ln(x_{k+1}) - \sum_{i=2}^{k} \ln(1+(1+\alpha)2^\alpha x_i^\alpha) \\
&\approx (\alpha-1)\ln(x_{k+1}) - \sum_{i=2}^{k} (1+\alpha)2^\alpha x_i^\alpha \\
&\qquad\qquad \text{(using the approximation for } \ln(1+x) \text{ as } x \text{ for small } x^2\text{)} \\
&\approx (\alpha-1)\ln((\alpha-1)2^a(k+2)^{-1/\alpha}) - \sum_{i=2}^{k}(1+\alpha)2^\alpha((\alpha 2^\alpha(i+1))^{-1/\alpha})^\alpha \quad \text{(using (3.1))} \\
&\approx \frac{1-\alpha}{\alpha}\ln(k+2) - \frac{1+\alpha}{\alpha} \underbrace{\sum_{i=2}^{k}\frac{1}{i+1}}_{\text{sum of reciprocals} \sim \ln(k+2)} \\
&\sim -2\ln(k+2).
\end{aligned}$$

Taking exponentials of both sides, we conclude that our series is bounded above by a constant multiple of

$$\sum_{k=2}^{\infty} \frac{1}{k^2},$$

which is convergent. Hence, there is some $C < \infty$ such that

$$\mathrm{Adl}(T_A) < C,$$

and this verifies the bounded distortion condition. □

Combining Lemma 3.4 and Theorem 3.3, we conclude that T_A has an acip which we will denote μ_α, and furthermore μ_α is the weak* limit of a subsequence $(\mu_{\alpha,n_k})_k$ of:

$$\mu_{\alpha,n} := \frac{1}{n} \sum_{j=0}^{n-1} (T_A^j)_* \lambda.$$

Furthermore, recalling Proposition 2.9, we find that the original Manneville-Pomeau map T_α has an invariant measure which we will call ν_α, and this measure is given by the "pushing" method in (2.3) (but note that indexing is different and we have pre-computed return times):

$$\forall E \in \mathcal{B} : \nu_\alpha(E) = \sum_{j=0}^{\infty} \sum_{k=0}^{j} \mu_\alpha(T_\alpha^{-k}(E) \cap A_j). \tag{3.4}$$

We also get from this proposition that the pushed measure is absolutely continuous with respect to λ, i.e. $\nu_\alpha \ll \lambda$.

We claimed at the beginning of this section that finding an invariant *probability* measure is preferable. Again by Proposition 2.9, we know that certainly ν_α is σ-finite (at least when $0 < \alpha < 1$ which we have assumed throughout); the condition for finiteness is

$$\int_A R \, d\mu_\alpha < \infty.$$

Proposition 3.5. *The Manneville-Pomeau map T_α has an acip for any $\alpha \in (0, 1)$.*

Proof. The above derivations gave an invariant, absolutely continuous invariant measure ν_α for T_α which we now wish to normalise to get a probability measure, which will then be an acip.

This can be done provided ν_α is finite, which we said above is equivalent to the condition

$$\int_A R \, d\mu_\alpha < \infty,$$

where μ_α is the invariant measure of the induced map.

However, we do not currently have a way of integrating against μ_α. We have established the existence of this measure in Lemma 3.4, and the only other thing we know is that it is the limit of a subsequence $(\mu_{\alpha,n_k})_k$ of:

$$\mu_{\alpha,n} := \frac{1}{n} \sum_{j=0}^{n-1} (T_A^j)_* \lambda.$$

[2]Some thought is necessary to justify that summing k approximations of $\ln(1+x)$ will be a uniform constant away from the sum of the x as $k \to \infty$. The error in the approximation has order x^2, and the terms we are plugging in are $x = x_i$ which behave like $i^{-1/\alpha}$, so the sum of the errors is a convergent series.

This expression as a limit will actually be enough to prove that the integral of the return time is finite, and we will do this by showing that integrating with respect to μ_α is not so different from integrating with respect to λ. More precisely, we will show that there exists a constant D such that

$$\forall E \in \mathcal{B}|_{[1/2,1]} : \mu_\alpha(E) \le D \cdot \lambda(E),$$

which will give us the following bound:

$$\int_A R \, d\mu_\alpha \le D \int_A R \, d\lambda.$$

Using the fact that T_A is a FBEM by Lemma 3.4, we consider the base partition $\mathcal{P} := \{A_0, A_1, \dots\}$, and then as in the definition, we consider the sets of cylinders $\mathcal{P}_j := \bigvee_{i=0}^{j-1} T_A^{-i}\mathcal{P}$ for each j. Similarly to what we did in the proof of Lemma 2.12, we can show that the partition $\mathcal{P}_j$ breaks down $[1/2, 1]$ into intervals on which T_A^j is full-branched, and by the lemma, we know that:

$$\sup_{k\ge 1} \sup_{Z\in\mathcal{P}_k} \sup_{x,y\in Z} \left| \frac{(T_A^k)'(x)}{(T_A^k)'(y)} \right| < \infty.$$

We claim that the constant D we require can be taken to equal this supremum. (We will also need that $D \ge 1$, which is actually guaranteed to be the case for the supremum above; e.g. we can take $x = y$).

The proof of this claim is essentially contained in the proof of the Ergodic Folklore Theorem in [Tod20], but we will replicate the relevant parts here. Note first that it suffices to show that the bound with D holds for each push-forward measure $(T_A^j)_*\lambda$, since it will then hold in the limit. So, fix $j \ge 0$.

Let $E \in \mathcal{B}|_{[1/2,1]}$ be an interval; intervals generate the Borel algebra, so proving the bound for every interval E will prove it for all measurable sets. This way we know that $\lambda(E)$ is just the distance between the endpoints of the interval E, which we may write as $\lambda(E) = |E|$.

If $j = 0$ the bound holds trivially. Now suppose $j \ge 1$. Then let $Z \in \mathcal{P}_j$ be arbitrary, and consider the interval $T_A^{-j}E \cap Z$. Since T_A^j is increasing and bijective on Z, and $T_A^j(Z) = [1/2, 1]$, the interval E must have a preimage inside Z under T_A^j, and this preimage is precisely $T_A^{-j}E \cap Z$. So the endpoints of the interval $T_A^{-j}E \cap Z$ must map to the endpoints of E under T_A^j, and hence we can apply the Mean Value Theorem to find a $y \in T_A^{-j}E \cap Z$ such that

$$|(T_A^j)'(y)| = \frac{|E|}{|T_A^{-j}E \cap Z|},$$

and similarly applying the MVT to $T_A^j(Z) = [1/2, 1]$ produces $x \in Z$ such that

$$|(T_A^j)'(x)| = \frac{|[1/2, 1]|}{|Z|} = \frac{1}{2|Z|}.$$

Now,

$$\frac{|T_A^{-j}E \cap Z|}{|Z|} = \frac{|(T_A^j)'(x)|}{|(T_A^j)'(y)|} \cdot 2|E| \le 2D|E|,$$

Where D is the distortion constant. Hence $|T_A^{-j}E \cap Z| \le 2D \cdot \lambda(E)\lambda(Z)$.

Now if we sum over every $Z \in \mathcal{P}_j$ we get

$$\begin{aligned}
(T_A^j)_*(E) &= \sum_{Z\in\mathcal{P}_j} |T_A^{-j}E \cap Z| \\
&\leq 2D \cdot \lambda(E) \sum_{Z\in\mathcal{P}_j} \lambda(Z) \\
&= 2D \cdot \lambda(E)\lambda([1/2, 1]) \\
&= D \cdot \lambda(E).
\end{aligned}$$

Now that the bound holds, proving that the measure ν_α is finite boils down to showing $\int_A R\, d\lambda < \infty$. To evaluate this integral, note that R is an integer-valued function and is equal to n precisely on A_{n-1} for each $n \geq 1$. So,

$$\begin{aligned}
\int_A R\, d\lambda &= \sum_{n=1}^{\infty} \int_{a_n}^{a_{n-1}} n\, d\lambda \\
&= \sum_{n=0}^{\infty} (a_n - 1/2) && \text{(“summing horizontally rather than vertically”)} \\
&= \frac{1}{2} \sum_{n=0}^{\infty} x_n \\
&= \frac{1}{2} \sum_{n=0}^{\infty} (\alpha 2^\alpha (n+1))^{-1/\alpha}(1 + \mathcal{O}(n^{-1})) && \text{by (3.1)} \\
&\leq \frac{\alpha^{-1/\alpha}}{4} \left(\sum_{n=0}^{\infty} \frac{1}{(n+1)^{1/\alpha}} + \sum_{n=0}^{\infty} \frac{A}{(n+1)^{1+1/\alpha}} \right) && \text{for some } A \gg 1 \\
&< \infty,
\end{aligned}$$

since $\alpha \in (0, 1) \implies 1/\alpha > 1$, so both series converge. □

The work in this section has therefore allowed us to conclude that for $\alpha \in (0, 1)$, the map T_α has an invariant, absolutely continuous, probability measure. We get this measure by taking the invariant measure ν_α and normalising by dividing by the size of the space $\nu_\alpha([0, 1])$, which we know to be finite by the proof above. Call this normalised measure $\bar{\nu}_\alpha$, so that we have:

$$\bar{\nu}_\alpha = \frac{\nu_\alpha}{\nu_\alpha([0, 1])} \qquad \text{and} \qquad \forall E \in \mathcal{B} : \bar{\nu}_\alpha(T_\alpha^{-1}E) = \bar{\nu}_\alpha(E).$$

Remark 3.6. The proof above relied on the existence of a constant D such that the T_A-invariant measure μ_α that we obtained in Equation 3.4 satisfies

$$\mu_\alpha \leq D\lambda.$$

While we only needed this upper bound, a lower bound can also be obtained with exactly the same arguments. There is only one inequality in the proof, and it can be reversed by a symmetry argument: since our distortion constant is

$$D = \sup_{k\geq 1} \sup_{Z\in\mathcal{P}_k} \sup_{x,y\in Z} \left| \frac{(T_A^k)'(x)}{(T_A^k)'(y)} \right|,$$

then by exchanging x and y, we get that

$$\frac{1}{D} = \inf_{k\geq 1} \inf_{Z\in\mathcal{P}_k} \inf_{x,y\in Z} \left| \frac{(T_A^k)'(x)}{(T_A^k)'(y)} \right|,$$

and so in fact:

$$\frac{1}{D}\lambda \leq \mu_\alpha \leq D\lambda.$$

This is a sufficient (but not necessary) condition for μ_α to be *equivalent* to Lebesgue measure: the two measures have the same null sets.

3.4 Invariant densities

In the previous section, we searched for an invariant measure for T_α which satisfies the additional restriction of being absolutely continuous with respect to our reference measure λ (Lebesgue). When working with finite absolutely continuous measures, the Radon-Nikodym Theorem (Theorem 1.24) applies, and so each a.c. measure has a *derivative* (or "*density*") with respect to Lebesgue measure.

The goal of this section is to produce graphs for some of these densities. Their properties are already well-known in the literature, as we will discuss at the end of the section. However, actual graphs for the purposes of visualisation are difficult to find,[3] so this section has arisen from my desire to link up measures and densities and actually plot the results. Namely, we will derive a pointwise closed form for the invariant density of T_α.

Keeping notation from the beginning of this chapter, the two densities we will look at now are the ones corresponding to the two a.c. invariant measures from the previous section: μ_α for the FBEM T_A, and $\bar{\nu}_\alpha$ for T_α. Both, we have seen, are finite. We have proven the existence of both of these invariant measures, but that does not necessarily give us an explicit way of finding their associated densities.

Conveniently, both of our invariant measures are given as algebraic combinations of measures that have a known density. Let's recall the results from the previous section:

$$\frac{1}{n_k} \sum_{j=0}^{n_k-1} f_*^j \lambda \xrightarrow{w^*} \mu_\alpha \tag{3.5}$$

$$\nu_\alpha(E) = \sum_{j=0}^{\infty} \sum_{k=0}^{j} \mu_\alpha(T_\alpha^{-k}(E) \cap A_j) \tag{3.6}$$

$$\bar{\nu}_\alpha = \frac{\nu_\alpha}{\nu_\alpha([0,1])}. \tag{3.7}$$

We start with the density of the T_A-invariant measure μ_α. We noted in Remark 3.6 that this measure was uniformly bounded above and below by D and $1/D$ respectively, for some distortion constant D. So, the same will hold for the density.[4]

[3] This is not to say that no-one has been able to create such graphs, but rather that since they are specific to the parametrisation of the map, it is usually more efficient to prove general properties of the density which give a rough idea of its graph—in a manner which is also applicable to other parametrisations.

[4] To a certain extent, a density uniformly bounded above and below is not interesting to study, because often upper and lower bounds are all we need. Nevertheless we can view finding this density as a means to an end, which is to study the more interesting $\bar{\nu}_\alpha$.

The sequence converging to the measure is a linear sum of push-forwards (see Definition 1.25) of Lebesgue measure. Since the density of Lebesgue measure is the constant function $\mathbf{1}$, and since taking the push-forward of a measure corresponds to applying the transfer operator to its density (recall Proposition 1.27), we are actually able to express the density of each element in the subsequence:

$$\rho_{\alpha,n_k} := \frac{1}{n_k} \sum_{j=0}^{n_k-1} \mathcal{L}^j \mathbf{1}, \tag{3.8}$$

Where $\mathcal{L}$ denotes the transfer operator associated with T_A and Lebesgue measure. We will further denote by ρ_α the density $d\mu_\alpha/d\lambda$. Since $\mu_{\alpha,n_k} \xrightarrow{w^*} \mu_\alpha$, it would be nice to claim that the convergence indeed also holds (in some sense—perhaps pointwise) for the densities:

$$\mu_{\alpha,n_k} \xrightarrow{w^*} \mu_\alpha \implies \rho_{\alpha,n_k} \longrightarrow \rho_\alpha.$$

This is not immediately obvious, but it is certainly the case *if* we assume that the densities converge pointwise to *something*:

Lemma 3.7. *If ρ_{α,n_k} converge pointwise to some bounded measurable function, then that function is ρ_α.*

Proof. Suppose $\rho_{\alpha,n_k} \longrightarrow \rho$ for some measurable ρ. Define a measure μ through its density: $d\mu = \rho\, d\lambda$. Then for any continuous function $f \in C([1/2, 1])$, we have:

$$\int f\, d\mu_{\alpha,n_k} = \int f \rho_{\alpha,n_k}\, d\lambda \xrightarrow[k\to\infty]{} \int f\rho\, d\lambda = \int f\, d\mu$$

by the Dominated Convergence Theorem, since $f\rho_{\alpha,n_k} \leq f \cdot D$, with f bounded since continuous on a closed interval (the bound with D comes from the proof of Proposition 3.5).

However, since $\mu_{\alpha,n_k} \xrightarrow{w^*} \mu_\alpha$ and f continuous, we also have:

$$\int f\, d\mu_{\alpha,n_k} \xrightarrow[k\to\infty]{} \int f\, d\mu_\alpha.$$

Hence, μ and μ_α agree on integrals of continuous functions, and they are both finite, so they must be equal (this follows, for example, from Theorem 1.2 in [Bil99]). So the densities are equal, i.e. $\rho = \rho_\alpha$. □

To prove that the densities ρ_{α,n_k} do in fact converge, we can use a well-known theorem on sufficiently nice Markov maps, that implies that better still, the sequence of densities $(\mathcal{L}^n\mathbf{1})_n$ converges uniformly to some continuous function ρ. Since T_A is a FBEM, we can take the partition $\mathcal{P}$ to be a Markov partition and treat T_A as a CMS, since it will in this case satisfy all the niceness properties necessary for the two systems to be in bijection. Then, we can apply for example [Sar99, Theorem 5] with the potential $\phi = -\log|f'|$ to conclude the convergence we need (in fact Sarig shows this convergence is even exponentially fast). If $\mathcal{L}^n\mathbf{1} \to \rho$ uniformly, then certainly the n_k-averages of the sequence (equal to ρ_{α,n_k}) also converge to ρ pointwise, and then we apply Lemma 3.7 to conclude that

$$\rho_\alpha = \lim_n \mathcal{L}^n \mathbf{1}.$$

So, the idea is to iteratively apply the transfer operator to the constant function **1**. For large n, when little evolution is noticed from step to step, we get the following density profile:

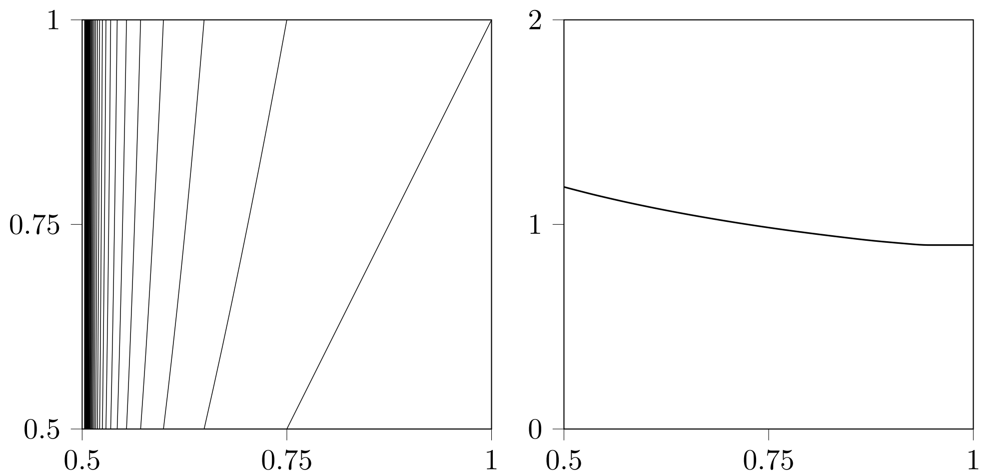


Figure 3.5: The inducing scheme T_A and its associated density ρ_α, both for $\alpha = 0.75$.

Despite the build-up of branches near the origin $x = 1/2$ for T_A, the associated invariant density remains bounded above and below, as predicted.

Remark 3.8. As an interesting heuristic exercise, we can think about why the density ρ_α is higher towards 1/2. Its value is determined by the iterates of the transfer operator, and in its pointwise form, this operator is a sum over pre-images of T_A, where summands are divided by the derivative of T_A at each pre-image. This derivative is smaller towards the left-hand side of each branch (this is visible on the second branch from the right in the plot above, or alternatively, can be shown directly from the definition of T_A).

Armed with an estimate for ρ_α, we can now express the density of ν_α as a combination of known functions. This is a basic computation that requires some jumping back and forth between measures and densities. Let ψ_α be the Radon-Nikodym density of ν_α, and let $E \in \mathcal{B}|_{[1/2,1]}$. We now take $\mathcal{L}$ to denote the transfer operator of T_α, rather than T_A, with respect to λ. Then:

$$
\begin{aligned}
\int_E \psi_\alpha \, d\lambda &= \mu_\alpha(E) \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{j} \mu_\alpha(T_\alpha^{-k} E \cap A_j) \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{j} \int_{T_\alpha^{-k} E} \chi_{A_j} \, d\mu_\alpha \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{j} \int \chi_{T_\alpha^{-k} E} \cdot \chi_{A_j} \cdot \rho_\alpha \, d\lambda \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{j} \int \chi_E \circ T_\alpha^k \cdot \chi_{A_j} \cdot \rho_\alpha \, d\lambda && \text{since } \chi_B \circ f = \chi_{f^{-1}B} \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{j} \int \chi_E \cdot \mathcal{L}^k(\chi_{A_j} \cdot \rho_\alpha) \, d\lambda && \text{by properties of the transfer operator} \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{j} \int_E \mathcal{L}^k(\chi_{A_j} \cdot \rho_\alpha) \, d\lambda \\
&= \int_E \left(\sum_{j=0}^{\infty} \sum_{k=0}^{j} \mathcal{L}^k(\chi_{A_j} \cdot \rho_\alpha) \right) d\lambda.
\end{aligned}
$$

Since this holds for all Borel sets E, we have found an explicit expression for the density ψ_α. This simplifies further once we consider its pointwise definition, using Proposition 1.28.

$$
\begin{aligned}
\psi_\alpha(x) &= \sum_{j=0}^{\infty} \sum_{k=0}^{j} \sum_{T_\alpha^k z = x} \frac{\chi_{A_j}(z) \rho_\alpha(z)}{|(T_\alpha^k)'(z)|} \\
&= \sum_{j=l}^{\infty} \frac{\rho_\alpha(v^{-1}(u^{-(j-l)}(x)))}{(T_\alpha^{1+j-l})'(v^{-1}(u^{-(j-l)}(x)))} && \text{where } x \in I_l \\
&= \sum_{i=0}^{\infty} \frac{\rho_\alpha(v^{-1}(u^{-i}(x)))}{(T_\alpha^{i+1})'(v^{-1}(u^{-i}(x)))} && \text{relabelling } i = j - l.
\end{aligned}
$$

The last line of this calculation is a simplification that comes from considering orbits carefully. Suppose $x \in I_l$. Fix j and fix $k \leq j$; then the characteristic function will cause the innermost sum to evaluate to all zeros, *unless* we can find a $z \in A_j$ such that $T_\alpha^k(z) = x$. Since $k \leq j$ and $T_\alpha A_j = I_j$ by definition, by the Markov structure of the map we have $T_\alpha^k A_j = I_{j-k+1}$. So there exists a z giving non-zero inner sum if and only if $l = j - k + 1$, and this z is unique if it exists, since it is equal to $z = v^{-1}(u^{-(j-l)}(x)) \in T_\alpha^{-(j-l+1)} x$.

This formula is a pointwise closed form, and it allows the density ψ_α to be computed approximately: the sum is convergent a.e. since we know the integral of ψ_α is $\nu_\alpha([0,1])$, which we proved in Proposition 3.5 is finite.

Now, we simply normalise to get:

$$\bar{\psi}_\alpha := \frac{d\bar{\nu}_\alpha}{d\lambda} = \frac{\psi_\alpha}{\nu([0,1])}.$$

We can estimate this computationally, which gives the following density, plotted next to the map itself.

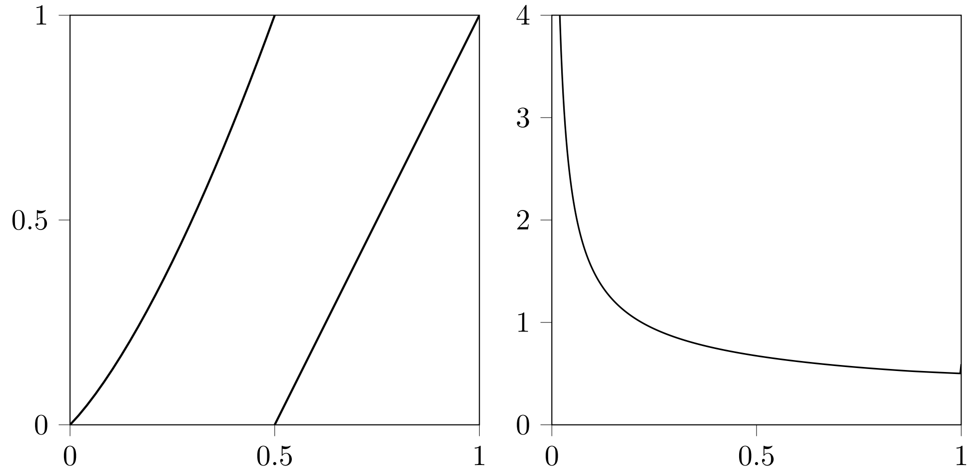


Figure 3.6: The Manneville-Pomeau map T_α and its associated invariant probability density $\bar{\psi}_\alpha$, both for $\alpha = 0.75$.

Remark 3.9. The graph suggests that $\bar{\psi}_\alpha(x) \geq C > 0$ for some constant C, which would imply that not only $\bar{\nu}_\alpha \ll \lambda$ but also $\lambda \ll \bar{\nu}_\alpha$, i.e. the two measures are equivalent (have the same null sets). Rigorously, we might look back at the pointwise expression for ψ_α and notice that all the summands are positive, so just singling out the first one:

$$\psi_\alpha(x) \geq \frac{\rho_\alpha(v^{-1}(x))}{T'_\alpha(v^{-1}(x))} \geq \frac{1}{2D},$$

where D was the distortion bound for T_A. Dividing through by $\nu([0,1])$, it holds true that the density can be bounded from below by a positive constant.

This density is much more interesting, firstly because it actually relates directly to the map we are studying (rather than the inducing scheme, which is just a necessary step towards understanding T_α better). It also seems from this estimate that the density is unbounded towards $x = 0$, and getting a better understanding of this phenomenon requires more work.

While this could perhaps be done carefully using the pointwise definition we have above, a more elegant approach is carried out by Liverani, Saussol and Vaienti in [LSV99]. I refer the reader to section 2 of their paper for the full proof, which is cleaner than what we have done here with the inducing and pushing technique.

The key difference in Liverani, Saussol and Vaienti's approach is that they do not use the inducing scheme at all, instead searching for a fixed point under the transfer operator (referred to in their paper as P) of T_α.

The idea in [LSV99] is to find a compact set of continuous functions on $[0,1]$ which is preserved under the transfer operator, and then construct a sequence of functions in this set which is invariant in the limit (and which must have a convergent subsequence by compactness). The properties shared by all functions in the compact set are then properties of the invariant density. In their paper, the properties they manage to extract are that $\bar{\psi}_\alpha$ must be locally Lipschitz, and that

$$\bar{\psi}_\alpha(x) = \mathcal{O}(x^{-\alpha})$$

as $x \to 0$.

On another note, estimates for the invariant densities of more general maps with indifferent fixed points have been known for some time; see for example [Tha80].

3.5 Interpretation

Now that we have an invariant measure for T_α (and a plot of its density with respect to Lebesgue measure), we can reap the benefits. At the most basic level, for example, Poincaré's Recurrence Theorem (Theorem 1.11) applies and we conclude that any set of positive Lebesgue measure is visited infinitely often by the orbit of λ-almost every $x \in [0,1]$. (We have made the leap from $\bar{\nu}_\alpha$ to λ using equivalence, which we noted in Remark 3.9).

But we can do much better. Additional interpretation abilities become available if we can prove that $\bar{\nu}_\alpha$ is not only invariant, but also ergodic. We can do this in a number of ways, but the route of least effort is to use the Young tower technology we have already accumulated. So, it is time to properly link up the Manneville-Pomeau map T_α with our work in section 2.1.

Lemma 3.10. *Let $I = [0,1]$. The map T_α on $(I, \mathcal{B}, \lambda)$ admits a Young tower representation (in the sense of Definition 2.3) with base set $A = [1/2, 1]$ and with the return time partition $\{A_n\}_{n=0,1,\ldots}$.*

Proof. Clearly the iterates of A cover the whole space since $T_\alpha(A) = I$, and each of the countably many partition elements A_i has finite return time $i+1$. Since T_α is bounded and piecewise continuous, and each positive-measure subinterval of I has a positive-measure preimage under T_α, the map is also nonsingular. Furthermore, the induced map $T_A := T_\alpha^{R_A}$ is full-branched and bijective on each branch (we've already seen that T_A is a FBEM).

We now just need to prove that Young's conditions **Y1**–**Y5** apply. Showing measurability and strong generation are not interesting exercises, but in the case of a relatively nice interval map like T_α, they certainly hold. Aperiodicity (**Y3**) follows from the fact that every return time is possible on A. As for finiteness of the return time integral (**Y5**), this was the final calculation in the proof of Proposition 3.5.

The last thing to check is the distortion condition **Y4** on T_A. We have already spent quite some time in this chapter proving another distortion condition on T_A—namely, Adler's condition **D2**. For the sake of time and sanity we will consider Young's condition to be similar enough that an analogous calculation will bring us close. ☐

Now that we can place ourselves in the setting of Young's paper [You99], we can use its powerful theorems. Call (Δ, F) the Young tower obtained from T_α, and take the notation of section 2.1. Namely, we will once again abuse notation and consider λ to be a measure on the tower obtained by carrying Lebesgue measure up from $A = [1/2, 1]$.

Proposition 3.11. *The T_α-invariant probability measure $\bar{\nu}_\alpha$ is ergodic.*

Proof. A result in [You99, Theorem 1] guarantees the existence of an ergodic probability measure ν for F. In Young's proof of Theorem 1, ν is obtained using exactly the pushing formula (2.1) followed by normalisation, and the measure getting pushed has density equal to the limit of ρ_{α,n_k} from Equation 3.8. We noted at the end of subsection 2.1.5 that this means that the pushing method we used here, and the one used by Young, are in fact equivalent, and so we should have $\bar{\nu}_\alpha = \nu \circ \pi_\Delta^{-1}$.

For ergodicity, then, suppose $T_\alpha^{-1}E = E$ for some $E \in \mathcal{B}$. Then $\pi_\Delta^{-1}T_\alpha^{-1}E = \pi_\Delta^{-1}E$. By the conjugation properties of the projection π_Δ (see Proposition 2.8), this implies $F^{-1}(\pi_\Delta^{-1}E) = \pi_\Delta^{-1}E$. Since ν is ergodic, we have either $\nu(\pi_\Delta^{-1}E) = 0$ or $\nu(\pi_\Delta^{-1}E) = 1$, i.e. $\bar{\nu}_\alpha(E)$ is either 0 or 1. □

This is an ideal setting: we have an ergodic probability measure equivalent to Lebesgue measure. Birkhoff's Ergodic Theorem will allow us to convert between time averages and space averages; one basic application of this is as follows.

Corollary 3.12. *Let E be a measurable subset of the interval. For λ-almost every $x \in I$, the limiting proportion of time that the orbit of x under T_α spends in E is $\int_E \bar{\psi}_\alpha(x)\,dx$.*

Proof. The limiting proportion we want is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(T_\alpha^k(x)),$$

which by Birkhoff's Ergodic Theorem (Theorem 1.14) is equal to $\int \chi_E \, d\bar{\nu}_\alpha$. Convert to Lebesgue measure using the Radon-Nikodym derivative $\bar{\psi}_\alpha$ to conclude. □

Broadly, this means that the density $\bar{\psi}_\alpha$ (with graph given in Figure 3.6) shows how "attractive" different parts of the space I are to orbits: the higher the density, the more attractive the area.[5] We conclude from the singularity near $x = 0$ that something about the dynamics of T_α is creating a build-up of orbits near the origin. For comparison, the invariant density for the doubling map $\alpha = 0$ is easily checked to be the uniform density $\mathbf{1}$, so the indifferent fixed point has indeed had an effect on the dynamics.

3.6 Decay of correlations

The final ergodic question about T_α that we will seek to answer is about decay of correlations. This question emerges as a finer, and stronger, version of the mixing properties of dynamical systems (recall the definition of mixing from Definition 1.15). Mixing and its generalisations are phrased in terms of an ergodic probability measure, so our work in previous sections finding $\bar{\nu}_\alpha$ has been necessary to get to this point. We will only briefly discuss a few results here, to show the techniques involved and the historical progression on the bounds. All of the results in this section are taken from recent papers and talks and credited as such, apart from Proposition 3.16 where we have to do some of our own work to reconcile (I, T_α) with its Young tower representation. A good historical overview of the progress made on the question of decay of correlations for the Manneville-Pomeau map is available in [Bal00, Section 3.5], but it does not contain the two most recent advances made since its publication by Sarig then Gouëzel (Proposition 3.18 and Proposition 3.19 respectively).

[5] Quite the opposite of the modern "urban flight" phenomenon.

Let $(X, \mathcal{E}, \mu, T)$ be an ergodic ppt. Mixing is a statement about iterates of sets becoming independent of one another. But rather than thinking about mixing as a statement on sets:

$$\mu(A \cap T^{-n}B) \longrightarrow \mu(A)\mu(B),$$

we can think about this same problem as a statement on characteristic functions:

$$\int \chi_B \cdot \chi_A \circ T^n \, d\mu \longrightarrow \int \chi_A \, d\mu \int \chi_B \, d\mu.$$

The natural question to ask now is whether we can replace χ_A and χ_B with arbitrary functions in some function space, and the answer is yes if the function space is L^2 (and this is in fact an alternative formulation of strong mixing). If we consider these functions to be random variables on a probability space, this becomes a statement about the independence of stochastic random variables defined by the dynamics T.

Definition 3.13. Let $(X, \mathcal{E}, \mu, T)$ be an ergodic ppt and let $f, g \in L^2$. Then the n^{th} *correlation* of f and g under T is

$$\operatorname{Cor}(f, g \circ T^n) = \int f \cdot g \circ T^n \, d\mu - \int f \, d\mu \int g \, d\mu.$$

Note that the integrals are taken with respect to the invariant measure, rather than the reference measure (e.g. λ or m)! If the invariant measure is ever unclear, we can specify on the correlation function using subscripts, e.g. Cor_μ.

If we have a strongly mixing system, then it is clear that $\operatorname{Cor}(f, g \circ T^n)$ converges to 0. The finer question is then to determine the asymptotic speed at which this happens (in terms of n). We refer to this as *decay of correlations* or the *rate of mixing*.

For f and g in L^2, the rather unsatisfactory answer to this question is that nothing much can be said at all. Gouëzel in [Gou21] provides the following nice counterexample for the doubling map using Fourier series:

Proposition 3.14. *Let T be the doubling map on $[0,1]$ and let $(a_n)_n$ be a sequence of real numbers such that $\sum_n |a_n|^2 < \infty$. Then there exist $f, g \in L^2(\mathbb{R})$ such that $\forall n \in \mathbb{N} : \operatorname{Cor}(f, g \circ T^n) = a_n$ (i.e. we can get functions whose correlations decay however we like, so long as they do decay).*

Proof. Take $g(x) = \cos 2\pi x$ and $f(x) = \sum_k 2a_k \cos(2\pi 2^k x)$. Then $\int f \int g = 0$, and

$$\int f \cdot g \circ T^n \, d\lambda = 2a_n \int_0^1 \cos(2\pi 2^n x)^2 \, dx = a_n.$$

□

The doubling map being a chaotic FBEM, where we expect things to happen exponentially, this serves as a warning that L^2 is not restrictive enough. Hence, questions about decay of correlations usually treat a subset of L^2, often with a smoothness assumption such as Lipschitz continuity or differentiability.

Let's now return to our familiar example $(I, \mathcal{B}, \bar{\nu}_\alpha, T_\alpha)$. Our first upper bound for decay of correlations comes from the paper that we have used as the definition of our parametrisation of T_α:

Proposition 3.15 ([LSV99]). *Let $\alpha \in (0,1)$. Then for $f \in C([0,1])$ and $g \in L^\infty$, we have*

$$|\mathrm{Cor}(f, g \circ T_\alpha^n)| = \mathcal{O}(n^{1-1/\alpha}(\log n)^{1/\alpha}).$$

The approach to arrive at this bound is to introduce a random perturbation to the system, and this seems to be what generates the additional log term.

Using Young's method, a tighter bound is obtained, and this follows from the same paper we have been frequently quoting. This is for functions $f : \Delta \to \mathbb{R}$ rather than $f : [0,1] \to \mathbb{R}$, but to a certain extent this does not matter. This is because recalling the Young tower projection from Definition 2.7, we can convert functions f on $[0,1]$ into functions on Δ by setting e.g. $f_\Delta(x) = f(\pi_\Delta(x))$. Of course, we would then need to check that f_Δ was is member of Young's function spaces on the tower, and finally we would have to convert the conclusion back into a statement about the integrals of the functions on $[0,1]$.

Proposition 3.16. *Let $\alpha \in (0,1)$. Let (Δ, F) be the Young tower representation of $(I, \mathcal{B}, \bar{\nu}_\alpha, T_\alpha)$. Then for $f \in \mathcal{C}_\beta(\Delta)$[6] and $g \in L^\infty(\Delta, \lambda)$, we have*

$$|\mathrm{Cor}(f, g \circ F^n)| = \mathcal{O}(n^{1-1/\alpha}).$$

Proof. We already know that the Manneville-Pomeau map admits a Young tower representation in the sense of Definition 2.3, since we showed this in Lemma 3.10. The idea is therefore to apply [You99, Theorem 3], for which we need that $\tau_n = \mathcal{O}(n^{-\omega})$ for some $\omega > 0$. This ω will give us the rate of mixing.

Young's paper provides a formula for the size of the tails that we can work with:

$$\tau_n = \sum_{l>n} \lambda(\Delta_l) = \lambda(\cup_{l>n} \Delta_l).^{7} \tag{3.9}$$

This may be a good time to draw the Young tower obtained from T_α by inducing on $A = [1/2, 1]$. Remember that we have partitioned A into $A_n = [a_{n+1}, a_n)$ with return time $n+1$ (see the proof of Lemma 3.4, namely Figure 3.4, where each A_n is the domain of one of the branches of the induced map). We have that $a_n \to 1/2$, and we extend out the base so that each A_n has the same size for readability (in reality these sets get smaller as $n \to \infty$).

[6] The set $\mathcal{C}_\beta(\Delta)$ is a space of functions defined in Young's paper with a tower-specific Hölder condition attached.

[7] Note that although decay of correlations is phrased with respect to an ergodic probability measure, Young's condition applies to the size of the tails with respect to the reference measure (λ in this case).

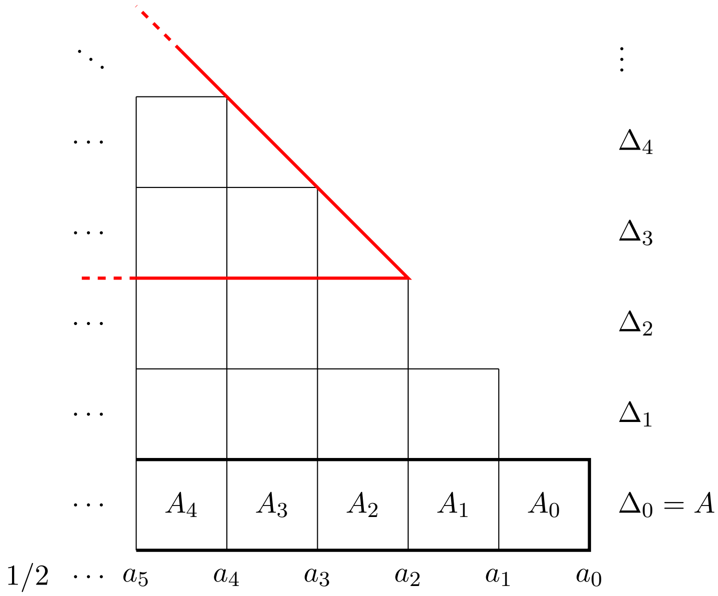


Figure 3.7: The Young tower for the Manneville-Pomeau map, with a tail marked.[8]

From (3.9) we obtain that the tail τ_n is just the measure of the triangle made up of all levels greater than n (for $n = 2$ this is the red triangle in the figure). Summing the measures row by row, each level has measure $a_l - 1/2$. The following calculation is quite similar to the proof of Proposition 3.5.

$$\begin{aligned}
\tau_n &= \sum_{l>n} (a_l - 1/2) \\
&\sim \sum_{l>n} x_l \\
&\sim \sum_{l>n} l^{-1/\alpha} \\
&= \sum_{l=n+1}^{\infty} \int_l^{l+1} l^{-1/\alpha}\, dt && \text{(forcing an integral to appear)} \\
&\sim \sum_{l=n+1}^{\infty} \int_l^{l+1} t^{-1/\alpha}\, dt \\
&= \int_{n+1}^{\infty} t^{-1/\alpha}\, dt \\
&\sim n^{1-1/\alpha} && \text{(by basic calculus (since } \alpha \neq 1\text{))} \\
&= \mathcal{O}(n^{-\omega}),
\end{aligned}$$

Where $\omega = 1/\alpha - 1$.[9] Since we are assuming $\alpha \in (0, 1)$, ω is positive, and so Young's Theorem 3 applies, and this implies exactly the asymptotics we were looking for. □

[8]The tail τ_n denotes the measure of the set of all points that have hitting time greater than n, and Young's formula says that this measure is equal to that of a triangle at the top of the tower (for $n = 2$, the red triangle in the figure). However, the sets are not equal, since points with high hitting time are located near the base, not the top, of the tower. The reason the two sets have the same measure is because we got the red one by shifting the points with high hitting time up to the top of the tower (and this doesn't change the measure since λ is lifted up the columns).

[9]The "ω" is referred to as "α" in Young's paper, but that letter was already used here for our parameter.

Remark 3.17. Theorem 4 in [You99] is also highly sensitive to the size of tails τ_n. Under the right conditions, it implies that the Central Limit Theorem (CLT) applies, i.e. for a (measure-theoretic) random variable $\varphi : \Delta \to \mathbb{R}$ with integral 0, the sum $\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \varphi \circ F^i$ converges in law to a normal distribution centred at 0. However, the condition for this is $\tau_n = \mathcal{O}(n^{-\omega})$ for some $\omega > 1$ (rather than $\omega > 0$). In the case of the Manneville-Pomeau map, this means we can only show that the CLT holds for $\alpha \in (0, 1/2)$. For larger α and bad enough observables, (namely, observables that are non-zero at 0), the limiting distribution becomes a weaker statistical law called an *α-stable law* (see, for example, [MZ15]). The reason for this (at least if $\varphi(0) \neq 0$) is that an orbit entering a small neighbourhood of 0 stays there for many more iterations of T_α, and so in the sum of $\varphi \circ T_\alpha^i(x)$, very similar non-zero values $\varphi(z)$ for $|z| \ll 1$ appear multiple times in a row, creating a large jump in the value of the sum. This can be studied using *piling processes* (see [FFT20]).

Returning now to decay of correlations, we have found function spaces in which the decay for T_α is polynomial of order $1 - 1/\alpha$. This is already in stark contrast with the doubling map, our canonical example of chaos, where Lebesgue measure is ergodic and decay (at least for C^∞ functions) is *superexponential*, i.e. faster than ε^n for any $\varepsilon > 0$ (a proof was given in [Gou21], once again using Fourier series). However, all we have currently is an upper bound for decay of correlations for T_α: it could be that this is a bad upper bound, and actually decay is exponential for the Manneville-Pomeau map too.

In 2002, Sarig showed that the bound from Young's paper is indeed optimal, i.e. we can find reasonable functions whose correlations decay exactly at the speed given in Proposition 3.16. The proof of this result uses *renewal theory*, a functional analysis approach where rather than using transfer operators (Definition 1.26), we use restrictions of transfer operators to sets of significance in inducing schemes. This fits in nicely with Markov maps. Sarig's result then finds the asymptotics of the correlation function in terms of the sequence $\bar{\nu}_\alpha(\{x \in A : R(x) > n\})$. (Note that this is related to the size of the tails τ_n in Young's model.)

Proposition 3.18 ([Sar02])**.** *Let $\alpha \in (0, 1/2)$. Then for f Lipschitz and g bounded measurable, we have*

$$|\mathrm{Cor}(f, g \circ T_\alpha^n)| = \Theta(n^{1-1/\alpha}).$$

This was improved shortly afterwards by Gouëzel, who extended the valid range for α to the full interval we have been looking at:

Proposition 3.19 ([Gou04])**.** *Let $\alpha \in (0, 1)$. Then for f Lipschitz and g bounded measurable, we have*

$$|\mathrm{Cor}(f, g \circ T_\alpha^n)| = \Theta(n^{1-1/\alpha}).$$

So we can indeed find fairly large classes of functions whose correlations, under the Manneville-Pomeau map, decay only exponentially: this is comparatively slow. Thinking back to Pomeau and Manneville's quotes in section 1.6, it had been observed that the "bursting region" of the map was where correlations are broken. Informally, we might conclude that correlations decay more slowly with an indifferent fixed point because along laminar portions of orbits (i.e. those passing near $x = 0$), the dynamics are not expanding enough to allow for fast mixing. Hence, all the mixing has to happen in the areas of the map with higher derivative. As the tangency is stronger for higher choices of α, this is also consistent with the asymptotic speed $n^{1-1/\alpha}$ being slower for larger α.

Conclusion

We stated at the beginning of this project that we were looking to observe and study intermittency in dynamical systems. While the Manneville-Pomeau map is one example amongst many, it seems to have served us well. Let's recap what we have established about this specific map T_α.

Firstly, we defined intermittency to be the alternation of laminar flow and chaotic bursts. The representation of T_α as a renewal shift seems to embody this well, with the return toward the renewal state 0 being inevitable and laminar, while the state mapped to after visiting 0 remains completely unpredictable. However, we noted that this alone was not enough to claim that intermittency was present, and we considered the relative sizes (in Lebesgue measure) of the Markov partition elements to conclude that they decay polynomially rather than exponentially. This means that picking a point in $[1/2, 1]$ uniformly at random gives a higher expected return time for T_α than it does for the doubling map; entering into a laminar phase is somehow *more likely*.

The tangency at the origin is to blame for this, as orbits arriving near $x = 0$ are then "captured" there for a while, in the sense that applying T_α to a small value of x barely increases it at all. We need many iterations of T_α to escape the slow, laminar area: here again we echo Pomeau and Manneville's words from section 1.6. This capture of orbits is also reflected in the system's invariant density, with a singularity at the origin.

On a more technical level, we can also make similar "exponential versus polynomial" statements about the system's decay of correlations, which are once again comparatively slow since they are also polynomial. This has a number of applications, from stochastic processes generated from dynamical systems to more basic questions about how quickly orbits move apart. Studying decay of correlations remains fashionable (notice that the results in the last section are comparatively recent), especially using modern methods such as transfer operators.

One of the limitations of studying the Manneville-Pomeau map in depth in this project, however, is that we might somewhat lack perspective coming out the other end. Many of the calculations we carried out in the third chapter were very specific to this map, and even to this parametrisation (by Liverani, Saussol and Vaienti). Nevertheless, the *methods* we used were not. In the process of proving results for T_α, we applied knowledge of both Young towers and countable Markov shifts, and I should stress again that these techniques are incredibly valuable in many scenarios: not just for T_α, and not even just for intermittent systems. They appear often in the literature and I hope that reading through this project may have provided an adequate warm-up for some more technical reading.

Despite some of the work being parametrisation-specific, much of what we did holds more generally for non-uniformly expanding maps with a neutral fixed point at the origin. Indeed, in some papers such as [Iso95], the situation is treated in more generality, and when given a specific map, it can be plugged into the derived formulae.

In fact, there are certainly many other maps (and even other interval maps) that are also intermittent, but for different reasons. One good example is *unimodal maps*, which it is shown in [Zwe04] have similar properties to the intermittent system we have studied here.

For an interested reader wanting to learn more about intermittency or ergodic theory in general, there are plenty of places to go. A common extension of the search for an a.c. invariant measure is thermodynamic formalism: we mentioned this briefly at the end of section 1.4, and pointed to [Sar99] as a good introduction for this area since it contains all the important definitions, and applications to CMS. We also only briefly mentioned limit theorems (in Remark 3.17)—these warrant a much more in-depth study, especially since finding the distribution of $\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\varphi\circ F^i$ can be generalised to describing $\frac{1}{\sqrt{nt}}\sum_{i=0}^{\lfloor nt\rfloor}\varphi\circ F^i$ as a function of t in the limit $n\to\infty$. This gives rise to *functional limit theorems*. Going in another direction, one can introduce *holes* to a dynamical system's state space, for example in such a way that they absorb orbits. We can then study the set of all points not yet absorbed at iteration number n and find conditional invariant measures for these sets; for the Manneville-Pomeau map, this is done in [DT17].

Bibliography

[Aar97] J. Aaronson. *An introduction to infinite ergodic theory.* American Mathematical Society, Providence, R.I, 1997.

[Adl73] R. L. Adler. F-expansions revisited. In Anatole Beck, editor, *Recent Advances in Topological Dynamics*, pages 1–5, Berlin, Heidelberg, 1973. Springer Berlin Heidelberg.

[Adl79] R. L. Adler. Afterword to R. Bowen. *Communications in Mathematical Physics*, 69(1):15–17, 1979.

[Bal00] V. Baladi. *Positive transfer operators and decay of correlations.* World Scientific Pub. Co, Singapore, 2000.

[Bar95] R. Bartle. *The elements of integration and Lebesgue measure.* Wiley, New York, 1995.

[BG97] A. Boyarski and P. Gora. *Laws of Chaos : Invariant Measures and Dynamical Systems in One Dimension.* Birkhäuser, Boston, Mass, 1997.

[Bil99] P. Billingsley. *Convergence of probability measures.* Wiley, New York, 1999.

[Coa20] D. Coates. *Semi-stable laws for intermittent maps.* PhD thesis, University of Exeter, 2020.

[DT17] M. F. Demers and M. Todd. Slow and fast escape for open intermittent maps. *Communications in Mathematical Physics*, 351(2):775–835, 2017.

[FFT20] A. C. M. Freitas, J. M. Freitas, and M. Todd. Enriched functional limit theorems for dynamical systems. Preprint, 2020.

[Gou04] S. Gouëzel. Sharp polynomial estimates for the decay of correlations. *Israel Journal of Mathematics*, 139(1):29–65, 2004.

[Gou21] S. Gouëzel. Ruelle resonances for geodesic flow on noncompact manifolds. Lecture, Spring School on Transfer Operators, Bernoulli Center, EPFL, March 2021.

[Iso95] S. Isola. Dynamical zeta functions and correlation functions for non-uniformly hyperbolic transformations. Preprint, 1995.

[LSV99] C. Liverani, B. Saussol, and S. Vaienti. A probabilistic approach to intermittency. *Ergodic Theory and Dynamical Systems*, 19(3):671–685, 1999.

[MZ15] I. Melbourne and R. Zweimüller. Weak convergence to stable Lévy processes for nonuniformly hyperbolic dynamical systems. *Annales de l'I.H.P. Probabilités et statistiques*, 51(2):545–556, 2015.

[Nik30] O. Nikodým. Sur une généralisation des intégrales de M. J. Radon. *Fundamenta Mathematicae*, 15:131–179, 1930.

[Pet83] K. E. Petersen. *Ergodic Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1983.

[PM80] Y. Pomeau and P. Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Communications in Mathematical Physics*, 74(2):189–197, 1980.

[Sar99] O. M. Sarig. Thermodynamic formalism for countable Markov shifts. *Ergodic Theory and Dynamical Systems*, 19(6):1565–1593, 1999.

[Sar02] O. M. Sarig. Subexponential decay of correlations. *Inventiones mathematicae*, 150(3):629–653, 2002.

[SC97] L. Saloff-Coste. *Lectures on finite Markov chains*, pages 301–413. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.

[Tha80] M. Thaler. Estimates of the invariant densities of endomorphisms with indifferent fixed points. *Israel Journal of Mathematics*, 37(4):303–314, 1980.

[Tod20] M. Todd. MT5877 Ergodic Theory and Dynamical Systems. Lecture Notes, 2020.

[Wei89] B. Weiss. On the work of V. A. Rokhlin in ergodic theory. *Ergodic Theory and Dynamical Systems*, 9(4):619–627, 1989.

[You99] L-S. Young. Recurrence times and rates of mixing. *Israel Journal of Mathematics*, 110(1):153–188, 1999.

[Zwe04] R. Zweimüller. S-unimodal Misiurewicz maps with flat critical points. *Fundamenta Mathematicae*, 181:1–25, 01 2004.