ON COARSE GRAINING AND OTHER FINE PROBLEMS

by

MATE PULJIZ

A thesis submitted to
the University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Mathematics
College of Engineering and Physical Sciences
University of Birmingham
September 2016

ABSTRACT

We study coarse grainings — reductions of a dynamical system to its factor systems. In the literature, different variations of this problem are known under various denominations; including lumping, model reduction, aggregating, semi-conjugacy, etc.

In the first half we investigate the problem of simplifying a dynamical system by reducing the number of variables and give an algorithm achieving this in some special cases. Building on the known results we extend the theory of aggregations of heuristics. We then turn to a probabilistic generalisation of these models and show that in certain cases they coarse grain onto the well-known probabilistic game of Gambler's ruin for which we prove some new results.

In the second half coarse graining is used to motivate questions in topological dynamics. Given a system (X, T) we study the induced system $(2^X, 2^T)$ on the hyperspace of compact non-empty subsets of X and its periodic points. Related to this we construct an almost totally minimal system on the Cantor set. We also give a solution to a certain problem in topological dynamics related to ω-limit sets and show how a known result on the Cantor set dynamics can be seen as a consequence of a structural result about shift spaces.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the help and guidance of Chris Good in the production of this work. Without his continuous support and encouragement this thesis would probably never get written up.

Thanks also go to Jonathan E. Rowe, David Parker, Leobardo Fernández, John Haslegrave, Chris Cannings, Peter Dittrich, Martin N. Jacobi, Jan Huwald, Richard Henze, Bahsar Ibrahim, Nishanthan Kamaleson, and Chunyan Mu. The author is indebted to all of them for many helpful discussions during which the problems he attempts in this thesis were formulated, and quite often, directions towards their solutions were hinted.

A heartfelt thank you goes to my family, especially my parents Dijana and Ante for always being by my side even when I was miles away.

Thanks also go to my friends and colleagues for making my time spent in Birmingham an enjoyable one. Primarily among these I want to thank Alessandro, Ani, Betty, Chunlei, Columba, Cristina, David, Esen, Kristina, and Valentina.

Lastly, the author acknowledges support from the European Union through funding under FP7-ICT-2011-8 project HIERATIC (316705).

CONTENTS

Chapter 1

INTRODUCTION

LTHOUGH very simple, it could be argued that the idea of representing complex (or complicated) systems in terms of their simplified models is fundamental to all the sciences —for applied disciplines such as physics and engineering this goes without saying, but also in mathematics itself working with so called "toy models" often offers a great deal of insight. In this thesis we hope to show how this simple idea can be formalised in the framework of dynamical systems, how it can help to solve a number of different problems, and how it guided our own research in unexpected directions.

A *system* for us always means an ordered pair (X, T) where X is a topological space (often compact and metric), and $T\colon X \to X$ is a continuous (and often surjective) map. Given such a system, one would like to know what its trajectories

$$x, T(x), T^2(x), T^3(x), \dots$$

look like for various initial conditions $x \in X$. But iterating the map T many times may be infeasible, either because it is computationally demanding or because it is just technically inaccessible. In that case, one might hope to find a projection $\Xi\colon X \to Y$ onto a simpler system (Y, S) which is compatible with the dynamics (X, T). That is to say that the trajectories of (Y, S)

$$y, S(y), S^2(y), S^3(y), \dots$$

should coincide with the projections under Ξ of the trajectories of (X, T). In other words (Y, S) should be a model for (X, T).

It requires only a little thought to see that this will be fulfilled provided that the following diagram commutes

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \downarrow{\scriptstyle \Xi} & & \downarrow{\scriptstyle \Xi} \\ Y & \xrightarrow{S} & Y \end{array}$$

or written symbolically

$$\Xi \circ T = S \circ \Xi.$$

When this happens we shall say that (X, T) *coarse grains*, or *factors* onto (Y, S); that Ξ is a *coarse graining*, or a *semi-conjugacy*, or a *factor map* of (X, T); and that (Y, S) is a *factor* of (X, T).

It is only natural to require the projection Ξ to be continuous and surjective. Surjective as we do not wish to introduce additional dynamical features that did not previously exist in the system (X, T), and continuous as we would like that tiny perturbations in the initial condition of the original system correspond to tiny perturbations in the coarse grained system.

To give the reader a better idea of what a coarse graining is, let us consider a simplified description of our Solar System as it is being taught to the first year undergraduates in any classical mechanics course. The initial positions of the Sun, the planets and their moons are given as a set of N position vectors in $\mathbb{R}^3$. Newtonian mechanics teach us that once we know their initial velocities, which is another set of N vectors in $\mathbb{R}^3$, we can predict the behaviour of this system i.e. give the positions and velocities of any celestial body at any point in the future (and indeed in the past). Here we assume that the masses of all bodies are known and do not change over the course of time.

To fit this within our framework, we define a map $T\colon (\mathbb{R}^3)^{2N} \to (\mathbb{R}^3)^{2N}$ which to a given set of positions and velocities assigns the new set of data representing the system after one year has elapsed. This is a standard way to discretise a continuously running system.

This map T is all we need to compute the state of our Solar System in 1-year time steps. If we wonder what the position of the planets will be in exactly 1000 years from now, we simply run the map T iteratively a thousand times on today's data and we shall obtain the answer.

But unfortunately the map T is not easily obtainable, let alone computable. As we have seen above, in order to produce T one would have to solve the equations of motion for our system. This is known as solving the N-body problem and is still very much an active area of research.

Consider now a projection map $\Xi\colon (\mathbb{R}^3)^{2N} \to \mathbb{R}^3 \times \mathbb{R}^3$ which to any configuration of celestial bodies assigns the position vector of its centre of mass $x \in \mathbb{R}^3$ and its velocity $v \in \mathbb{R}^3$. This is just the weighted average of all the position vectors and velocities in the system respectively. It is then a classical fact that the system $(\mathbb{R}^3 \times \mathbb{R}^3, S)$ given (in the appropriate units) by

$$S(x, v) = (x + v, v)$$

is a factor of $((\mathbb{R}^3)^{2N}, T)$ via coarse graining Ξ.

The crucial point here is that the dynamics of S requires no knowledge of the behaviour of the system given by T. The description of the dynamics on the coarser level is independent of the description of the dynamics on the finer level. And if one is interested only in the long term behaviour of the centre of mass of our Solar System then it suffices to study the system S which is a simple translation by a fixed velocity vector v.

Let us remark that for the purposes of this discussion we assumed that our System is in isolation and not a part of any other celestial structure. We also remark that various energy and momenta conservation laws can be seen as coarse grainings to a stationary system in a similar fashion.

In the first part of this thesis we shall mostly be concerned with the existence of coarse grainings and algorithms to find them. Stated in this generality the question is elusive and some restrictions have to be imposed in order to be able to say anything meaningful. Therefore in Chapter 2 we restrict our attention to systems where the underlying space is $\mathbb{R}^n$ (or some subset of $\mathbb{R}^n$) and where the dynamics is given by an analytic map. We also require coarse grainings to be linear projections of a very special form — we require that they are induced by the partitions of the set $\{1, 2, \ldots, n\}$. These special kinds of coarse grainings we call *aggregations*, or *lumpings*.

In that chapter we extend the results of Rowe, Vose and Wright [RVW06] and give a criterion deciding when an aggregation is a valid coarse graining. This, in theory, gives an algorithm to find all

such aggregations of a given map. It turns out that this is not always very effective as we also show that the problem of existence of a non-trivial aggregation is NP-complete when the family of maps consists of Markov chain transition matrices i.e. linear maps. The results in this chapter form joint work with Chris Good, David Parker, and Jonathan E. Rowe.

In Chapter 3 we study a special class of Markov chains with the state space consisting of different colourings of a fixed connected graph. The vertices of the graph are meant to represent individuals and the transition probabilities are modelling interactions between them where each individual adopts the colour (in that chapter we use the term strategy rather then colour) of its neighbour with certain probability. These kinds of models have previously been studied in relation to disease spreading and voting models. We show how the coarse graining approach can be used to study their statistical properties. When the underlying graph is the complete graph, these dynamics coarse grain onto a generalised instance of *Gambler's ruin* — a classical problem in probability theory for which we prove some new results related to the mean duration of the game. The results in this chapter form joint work with John Haslegrave.

We begin Chapter 4 by defining the *transfer operator* on the space of measures over X. Given a system (X, T) imagine that the space X is filled with an inhomogeneous gas whose varying density is given by a map $g\colon X \to [0, \infty)$. It is then natural to ask how the density of this gas evolves under the map T. Under the assumption that no other dynamics (such as diffusion) takes place — the transfer operator is exactly what provides the dynamics in this case. We then use a particular coarse graining to motivate the study of induced systems on the hyperspace of compact non-empty subsets of X. In the central part of this chapter we investigate the interactions of the original system and its induced systems. In particular we seek to explain how dynamical properties — such as having periodic points of certain period — transfer from the original system to its induced systems. As a by-product we construct a dynamical system on the Cantor set which is *almost totally minimal*, as defined there, and obtain an interesting result on dynamical embeddings (Theorem 4.21). The results in this chapter form joint work with Leobardo Fernández and Chris Good.

Finally Chapter 5 contains two additional results regarding the Cantor set dynamics. In the first part we reprove a result by Sherman from [She12] in which he shows that a system consisting entirely of

periodic points must either have a *finitely based spectrum*, a notion defined below, or otherwise have a non-trivial connected component. We show how, by using an appropriate coarse graining, his theorem can be seen as a corollary of a structural result about *shift spaces*.

The second part of this chapter revolves around ω-limit sets and draws from the theory advanced in recent years by Barwell, Davies, Good, Knight, Meddaugh, Oprocha, and Raines. There we resolve a conjecture related to ω-limit sets posed by Barwell in [Bar11, BDG12]. He asked if ω-limit sets are characterised by the ICT property in the systems with *shadowing*. We construct a Cantor system showing that the answer in general is no.

As we mentioned before, the idea of coarse graining is quite broad and it would be extremely difficult, if not impossible, to give a comprehensive treatment of the topic. This is why we chose to focus on different problems in which we could make use of this method. As a consequence the structure of this thesis is somewhat unconventional; the four main chapters are all written and can be read as self-contained units.

Lastly, we acknowledge that this research was funded by the HIERATIC project which aimed to develop a framework for understanding complex systems through the hierarchical decomposition — and the notion of coarse graining was used as a main tool for detecting higher level dynamics. As a consequence, large portions of this thesis appeared in technical reports of the HIERATIC project [GPPR14, GPPR15, FGP16b, CGH+14, CGH+15, FGH+16, GJP+14, GJP+15, DHH+15].

In addition to this the material from Chapter 3 has been accepted for publication as a joint paper [HP17] with John Haslegrave. The material from Chapter 4 is currently under review in a joint paper [FGP16a] with Leobardo Fernández and Chris Good. Lastly, the material from Chapter 2 forms a joint paper [GPPR16] with Chris Good, David Parker, and Jonathan E. Rowe which is also currently under review. Other material from the HIERATIC project has not been included in this thesis but may appear in future papers.

Chapter 2

A CHARACTERISATION OF STATE SPACE AGGREGATIONS

We study coarse grainings of discrete time dynamical systems on $\mathbb{R}^n$. In particular we are interested in state space (or dimensional) aggregations. An aggregation is a simple change of variables transformation mapping $\mathbb{R}^n$ to $\mathbb{R}^m$ which is associated to a particular partition of the set $\{1,2,\ldots,n\}$. This means that there are only finitely many tentative aggregations available and in theory one could go through the whole list checking each if it produces a well-defined dynamics on $\mathbb{R}^m$, i.e. if it is a coarse graining of our system $(\mathbb{R}^n, T)$. The actual difficulty is thus to find an efficient way to target only (or some of) those partitions which produce aggregations compatible with $(\mathbb{R}^n, T)$.

In this chapter we show that if T is in a certain class of maps (weighted binary tournaments) one can efficiently characterise and find all the compatible aggregations. The characterisation used for these maps can be generalised to work with analytic maps on $\mathbb{R}^n$, but it is no longer efficient enough to give a quick algorithm to identify compatible aggregations. Indeed, we are able to show that in any class of maps containing Markov chains (i.e. linear maps) the decision problem whether there exist a non-trivial aggregation is NP*-complete.*

We also show how our results apply to artificial chemistries, Random Heuristic Search models, and related finite population models.

The results in this chapter form joint work with Chris Good, David Parker, and Jonathan E. Rowe — the paper [GPPR16] *containing this material is currently under review.*

Modelling real life dynamical systems is a challenging problem. Whether we are exploring social dynamics on a large scale, or trying to understand biological interactions in a cell at the molecular level, the number of variables involved to describe these processes is often too large for any meaningful

computer simulation or analysis to take place in a reasonable amount of time. One then has to resort to various techniques of simplifying the underlying model, thus reducing its dimension and enabling computation.

One typical approach is that of coarse graining. Given a (continuous) map $T\colon \mathbb{R}^n \to \mathbb{R}^n$, the map $\Xi\colon \mathbb{R}^n \to \mathbb{R}^m$ is a *coarse graining* of T provided the following diagram commutes.

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{T} & \mathbb{R}^n \\ \downarrow{\scriptstyle \Xi} & & \downarrow{\scriptstyle \Xi} \\ \mathbb{R}^m & \xrightarrow{S} & \mathbb{R}^m \end{array}$$

The point here is that the dynamics are preserved by Ξ in the sense that $\Xi(T^k(x)) = S^k(\Xi(x))$ for any k (in topological dynamics, such a Ξ is called a semi-conjugacy). Typically, of course, in practical applications one wants m to be very much smaller than n.

In this chapter we look at a systematic approach for finding those coarse grainings which are a result of aggregating or lumping dimensions. Our approach stems from analysing state space aggregations in the Random Heuristic Search framework developed by Vose [Vos99], but extends to provide a complete characterization of such coarse grainings for discrete time Markov processes and analytic maps on $\mathbb{R}^n$. The theory provides an algorithm for determining such coarse grainings. However, we also show that the problem of deciding whether a Markov chain has a non-trivial state space aggregation is NP-complete.

Consider the following problem from evolutionary biology. Suppose that for a set of n states, which we might call genotypes, G (for simplicity we frequently assume that $G = \{1, 2, \ldots, n\} = [n]$) we are given a set of rules that determines which genotype is dominant in any given pair, i.e. the genotype inherited by the offspring. Equivalently, the rules specify an orientation for each edge of the complete graph over the set of n vertices G. We write $k \to i$ to mean that mating between k and i produces an offspring of type i. One would then like to predict how a population of individuals with these genotypes evolves over time.

The first problem one faces is how to model this evolution. One common approach is to assume that the population is large and well-mixed. This leads to a model whose dynamics is given by a quadratic polynomial $T\colon \mathbb{R}^n \to \mathbb{R}^n$, $T = (T_1, \ldots, T_n)$, where

$$T_i(p) = T_i(p_1, \ldots, p_n) = p_i \left(p_i + 2 \sum\nolimits_{k \to i} p_k \right), \text{ for any } 1 \leq i \leq n, \tag{2.1}$$

and where the sum runs over all the genotypes k which are dominated by i. The function T is the *heuristic* in the aforementioned Random Heuristic Search framework, but we shall often say that T is a *continuous*, or *infinite population model*. The interpretation is as follows: suppose that we start with a population in which the proportions of different genotypes are given by $p = (p_1, \ldots, p_n)$, then $T_i(p)$ gives the probability that the offspring of two individuals chosen at random has genotype i. In particular, note that the non-negative portion ($p \geq 0$) of the hyperplane $p_1 + \cdots + p_n = 1$ is invariant under T. We remark that this construction generalises *binary tournaments* given in [RVW05b, RVW06].

Another approach is to consider finite population models. Assume that our population is of fixed size r and that this total number does not change over time. Let $\mathbb{N}_0$ denote the set of non-negative integers and let X_r^n be the set of all vectors in $\mathbb{N}_0^n$ whose terms sum to r, so that $v = (v_1, \ldots, v_n) \in X_r^n$ represents a population of r individuals v_i of whom have genotype i, for each $i \leq n$. Clearly X_r^n has cardinality $C(n, r) := \binom{r+n-1}{r}$. We then define a discrete-time Markov chain with the state space X_r^n where the transition probabilities are given by

$$\mathbb{P}\left[v \to w\right] = \frac{r!}{w_1! \cdots w_n!} \left(T_1(v/r)\right)^{w_1} \cdots \left(T_n(v/r)\right)^{w_n} = \frac{r!}{w!} \left(T\left(v/r\right)\right)^{w}, \tag{2.2}$$

where T is as in (2.1). The equation above implies that, starting from a population $v \in X_r^n$, the transition probabilities are given by a multinomial random variable with the expectation $T\left(v/r\right)$ and the parameter r. In other words, if we start with the population v, in order to obtain the generation in the next time step we first compute $T\left(v/r\right)$, this represents the proportions of different labels in an urn from which we draw r of them independently with replacement to form the new generation w.

Alternatively, one can think of this Markov chain purely in algorithmic terms. Starting with a population represented by $v \in X_r^n$, one randomly chooses two individuals from it with replacement.

They then produce an offspring which inherits the dominant genotype of the parents (if the parents have the same genotype or if the same parent is chosen twice, then the offspring inherits that same genotype). This 'mating' process is repeated r times and results in a new generation of r offspring represented by w. It was shown in [Vos99] that this leads to the same transition probabilities as given in (2.2).

Once a particular model has been chosen, one might try to reduce the state space in order to facilitate computations. This clustering of portions of the state space should be done in such a way that the dynamics on these 'higher level' states is still well-defined. In the present chapter we shall be interested only in reductions that work across the range of derived models we mentioned above. In Section 2.1 we argue that the only simplifications to be considered under this requirement are state space aggregations that correspond to the changes of variables of the form $P_i = p_{i_1} + \cdots + p_{i_k}$ where each of the variables p_j occurs exactly once in exactly one of P_is. This effectively gives a partition of the set of variables. It was shown by Vose in [Vos99] that for aggregations of this type it suffices to find those that are compatible with the continuous model as they naturally extend to aggregations for the derived models described above. We further show that they are also compatible with the dynamics of derived models from a larger class (Theorem 2.1) reinforcing the view that it is only aggregations of the continuous model that matter.

It turns out that for the map T as in (2.1) it is possible to characterise all the valid aggregations efficiently using the notion of *contiguous partitions*. We say that an equivalence relation (or the corresponding partition) $\equiv$ on the set G is *contiguous* with respect to an orientation $\to$ on G if for all $i, j, k \in G$ we have

$$i \equiv j \equiv k \qquad \text{whenever} \qquad i \equiv k \quad \text{and} \quad k \to j \to i.$$

In words, any two genotypes i and k that are in the same block of the partition must for each individual j from any other block either both dominate or both be dominated by it. In Section 2.2 we prove

Theorem 2.3. *Let T be a heuristic as in (2.1). An equivalence relation on G is compatible (i.e. gives a coarse graining) with T if and only if it is contiguous with respect to $\to$.*

This can be seen as an extension of the well-known condition for lumping states of a Markov chain (see e.g. [KS76]) which says that any two states that are lumped together must have exactly the same outgoing transition probabilities towards any of the blocks in the partition. Theorem 2.3 itself can further be seen as a Corollary of Theorem 2.11 where, more generally, we give a criterion for finding compatible aggregations in terms of the Taylor coefficients of an analytic map.

Section 2.3 is about the computational complexity of our problem and there we show that even for Markov chains the problem of deciding whether a non-trivial aggregation of the system exists is NP-hard (Theorem 2.6). It is then perhaps somewhat surprising to learn that in that same section we give a polynomial algorithm deciding the very same question in the class of weighted binary tournaments. Theorem 2.3 above and its analogue, Theorem 2.4, play a decisive role in constructing this algorithm.

In Section 2.4 we show how the systems we are studying can be interpreted in the context of artificial chemistries. And lastly Section 2.5 deals with aggregations of analytic maps. This can be seen as a general model encompassing all those discussed in the sections preceding it. The contiguity criterion manifests itself there as a condition on certain Taylor series coefficients, see (2.10).

Coarse graining has been previously studied by Vose and his collaborators and some useful criteria have been devised that guarantee existence of coarse grainings in certain cases [RVW06, RVW05a, RVW05b, Bur07]. Other authors, most prominently Rabitz and collaborators [LR89, LR90, LRT94], and more recently Jacobi [Jac05], and Tomlin et al. [TLRT97] explored the coarse grainings of continuous systems given by differential equations. It is not hard to see that the theory of linear coarse grainings coincides in both the discrete and continuous systems.

We would like to stress that there are other approaches to simplifying the dynamics of iterated dynamical systems. These model reduction techniques include, most notably, singular value decomposition and variants thereof (see, for example, [JG09, GKK+06]). However, in general these do not preserve the inner dynamics of the system, which is our primary concern.

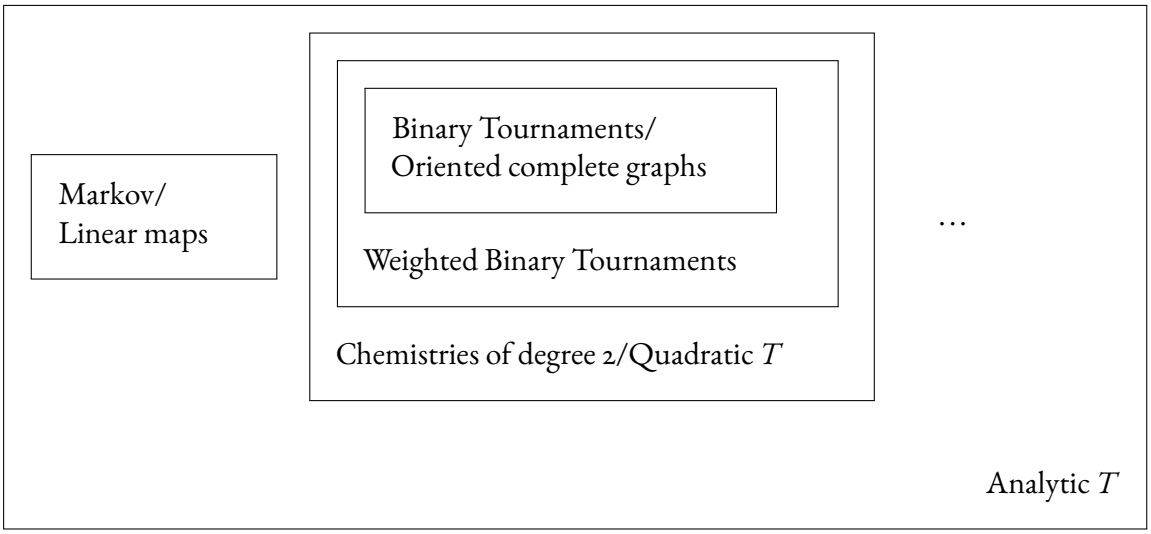

Figure 2.1: Maps for which we characterise aggregations

2.1 DIFFERENT MODELS

In this section we describe a general method for generating finite population Markov chain models akin to those we discussed in the introduction. Our starting point is an infinite model given by a map $T\colon \mathbb{R}^n \to \mathbb{R}^n$. We further assume that T maps the unit simplex $\Delta^n = \{p \in \mathbb{R}^n \mid \sum_{i=1}^n p_i = 1 \text{ and } p_i \geq 0\}$ into itself.

Let

$$\left\{Z_\alpha^{k,i}, Y_\beta^{k,i} \mid k, i \in \mathbb{N} \text{ and } \alpha, \beta \in \Delta^n\right\}$$

be a family of independent random variables where

$$Z_\alpha^{k,i} \sim \begin{pmatrix} 1 & 2 & \dots & n \\ \alpha_1 & \alpha_2 & \dots & \alpha_n \end{pmatrix} \quad \text{and} \quad Y_\beta^{k,i} \sim \begin{pmatrix} 1 & 2 & \dots & n \\ \beta_1 & \beta_2 & \dots & \beta_n \end{pmatrix}.$$

Let $r \in \mathbb{N}$, the population size, be fixed and recall that for our purposes $G = \{1, 2, \dots, n\} = [n]$. The family above acts as a stock of independent random variables we use to define a Markov chain $(S_k\colon k \in \mathbb{N})$ over the state space X_r^n. The evolution of this chain is prescribed by a transition function

$$F\colon X_r^n \times [n]^\infty \times [n]^\infty \to X_r^n,$$

and the rule

$$S_{k+1} = F(S_k; Z^{k,1}_{T(S_k/r)}, Z^{k,2}_{T(S_k/r)}, \dots; Y^{k,1}_{S_k/r}, Y^{k,2}_{S_k/r}, \dots).$$

The idea is that by varying F one can emulate the effect of changing the simulation algorithm by means of which the finite population model evolves. This will better be understood in an example.

If we let F be

$$F_1(v; i_1, i_2, \dots; j_1, j_2, \dots) = \sum_{l=1}^{r} e_{i_l},$$

where e_i is the i^{th} vector of the canonical basis in $\mathbb{R}^n$, then the chain $(S_k \colon k \in \mathbb{N})$ satisfies the following transition rule

$$S_{k+1} = \sum_{l=1}^{r} e_{Z^{k,l}_{T(S_k/r)}}.$$

In particular, this is precisely the same chain as the one described in the introduction. Each new generation S_{k+1} is formed by drawing r genes with replacement from an urn in which the distribution of different genes is given by $T(S_k/r)$.

Another useful choice is taking

$$F_2(v; i_1, \dots; j_1, \dots) = v + e_{i_1} - e_{j_1}. \tag{2.3}$$

In this case our chain $(S_k \colon k \in \mathbb{N})$ will satisfy

$$S_{k+1} = S_k + e_{Z^{k,1}_{T(S_k/r)}} - e_{Y^{k,1}_{S_k/r}}$$

and the next generation S_{k+1} is produced by throwing out one individual uniformly chosen from the current population (the third term accounts for this) in order to free up one space for another individual carrying a gene randomly chosen from $[n]$ with weights $T(S_k/r)$ (the second term). When T is quadratic as in (2.1) this evolution can, perhaps more naturally, be explained by saying that in each step a couple[1] is chosen that produces an offspring with the gene inherited from the dominant parent while at the same time, independently, one individual dies.

[1] As mentioned before, one feature of this model is that it allows for the same parent being chosen twice to form a couple.

Clearly F_1 and F_2 represent two extreme approaches to modelling this system. The first is generational, as all the individuals get replaced at each step, while the other represents one change at a time evolution. There are various other possibilities in between which can be modelled by choosing a different F.

We now wish to show that any aggregation of genes that coarse grains the dynamics of T also works for the models induced by F_1 and F_2. Indeed in Theorem 2.1 below we give a sufficient condition on F for this to happen.

Recall that an aggregation is a partition of the set $G = [n]$. By choosing some ordering on the blocks of the partition we can identify this with a function $\pi\colon [n] \to [m]$ where $m \leq n$. We set Ξ_π to be an $m \times n$ matrix associated to this aggregation where

$$\Xi_\pi(i,j) = \begin{cases} 1, & \text{if } \pi(j) = i, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the matrix Ξ_π is the change of variable transformation corresponding to the aggregation π. Asking that π is compatible with a map $T\colon \mathbb{R}^n \to \mathbb{R}^n$ amounts to asking that there exists a map $\tilde{T}\colon \mathbb{R}^m \to \mathbb{R}^m$ such that Ξ_π semi-conjugates T and $\tilde{T}$, i.e.

$$\Xi_\pi \circ T = \tilde{T} \circ \Xi_\pi.$$

Note that π aggregates $G = \{1, 2, \dots, n\}$ but the induced models we now wish to consider have X_r^n for the state space. There is, however, a natural partition on X_r^n that π induces and it is given by Ξ_π. In particular Ξ_π maps X_r^n to X_r^m.

Theorem 2.1. *Let $F\colon X_r^n \times [n]^\infty \times [n]^\infty \to X_r^n$ be a modelling scheme which is compatible with any aggregation, i.e. such that for any $m \leq n$ and any $\pi\colon [n] \to [m]$ the map $\tilde{F}\colon X_r^m \times [m]^\infty \times [m]^\infty \to X_r^m$ is well-defined by the formula*

$$\tilde{F}(\Xi_\pi(v); \pi(i_1), \dots; \pi(j_1), \dots) = \Xi_\pi(F(v; i_1, \dots; j_1, \dots)).$$

Then given any system $T\colon \mathbb{R}^n \to \mathbb{R}^n$ *all the aggregations compatible with* T *induce coarse grainings of the Markov model* $(S_k\colon k \in \mathbb{N})$ *obtained via* F.

Proof. Let $\pi\colon [n] \to [m]$ be an aggregation compatible with T. Employing the usual criteria for matching outgoing probabilities, it suffices to see that for any three states $v, u, w \in X_r^n$ the equation

$$\mathbb{P}\left[\Xi_\pi(S_{k+1}) = \Xi_\pi(w) \mid S_k = v\right] = \mathbb{P}\left[\Xi_\pi(S_{k+1}) = \Xi_\pi(w) \mid S_k = u\right]$$

holds as soon as $\Xi_\pi(v) = \Xi_\pi(u)$. The LHS of this expression is

$$\begin{aligned}
&\mathbb{P}\left[\Xi_\pi(S_{k+1}) = \Xi_\pi(w) \mid S_k = v\right] = \\
&\qquad = \mathbb{P}\left[\Xi_\pi\left(F\left(S_k; Z^k_{T(S_k/r)}; Y^k_{S_k/r}\right)\right) = \Xi_\pi(w) \mid S_k = v\right] = \\
&\qquad = \mathbb{P}\left[\tilde{F}\left(\Xi_\pi(S_k); \pi\left(Z^k_{T(S_k/r)}\right); \pi\left(Y^k_{S_k/r}\right)\right) = \Xi_\pi(w) \mid S_k = v\right] = \\
&\qquad = \mathbb{P}\left[\tilde{F}\left(\Xi_\pi(v); \pi\left(Z^k_{T(v/r)}\right); \pi\left(Y^k_{v/r}\right)\right) = \Xi_\pi(w)\right] = \\
&\qquad = \mathbb{P}\left[\tilde{F}\left(\Xi_\pi(v); Z^k_{\Xi_\pi(T(v/r))}; Y^k_{\Xi_\pi(v)/r}\right) = \Xi_\pi(w)\right].
\end{aligned}$$

It remains to notice that the last expression above will not change if we substitute u instead of v. This is because $\Xi_\pi(v) = \Xi_\pi(u)$, and hence also by assumption $\Xi_\pi(T(v/r)) = \Xi_\pi(T(u/r))$. Now backtracking the same steps we get the RHS. □

It is easy to check that both F_1 and F_2, regardless of the population size r, satisfy the hypothesis of the theorem above. Thus, we have just shown that the aggregations that work for the heuristic T, also coarse grain two associated infinite families of Markov chains. This justifies our primary concern of coarse graining heuristics.

At the end of this section we provide a simple example showing that even linear coarse grainings that are not induced by an aggregation of variables need not lead to a coarse graining of finite population models.

Example 2.2. Consider the identity map $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(p_1, p_2, p_3) = (p_1, p_2, p_3)$. This trivial dynamics is clearly coarse grained by any map. In particular, the projection map $\Xi(p_1, p_2, p_3) =$

$p_2 - p_3$ is a coarse graining albeit not induced by an equivalence relation. The transition matrix[2] of the associated Markov chain constructed via F_2 with $r = 2$ over the state space

$$X_2^3 = \{(2,0,0), (0,2,0), (0,0,2), (0,1,1), (1,0,1), (1,1,0)\}$$

is

$$\begin{bmatrix} 1 & \cdot & \cdot & \cdot & 1/4 & 1/4 \\ \cdot & 1 & \cdot & 1/4 & \cdot & 1/4 \\ \cdot & \cdot & 1 & 1/4 & 1/4 & \cdot \\ \cdot & \cdot & \cdot & 1/2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1/2 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1/2 \end{bmatrix}.$$

This matrix is clearly not compatible with the induced aggregation that merges the states $(2,0,0)$ and $(0,1,1)$. ◁

2.2 AGGREGATING WEIGHTED BINARY TOURNAMENTS

In the previous section we argued why we are interested only in aggregations of the infinite model T. Here, we seek to characterise these aggregations for the case when T is modelling a binary tournament of the form (2.1).

We start by orienting the complete graph K_n over $G = [n]$ and define a map $T\colon \mathbb{R}^n \to \mathbb{R}^n$, $T = (T_1, \ldots, T_n)$ as in (2.1) by setting

$$T_i(p) = T_i(p_1, \ldots, p_n) = p_i \left(p_i + 2 \sum\nolimits_{k \to i} p_k \right), \text{ for any } 1 \leq i \leq n.$$

Recall that an equivalence relation $\equiv$ on G is *contiguous with respect to the orientation* $\to$ if for all $i, j, k \in G$ we have

$$i \equiv j \equiv k \qquad \text{whenever} \qquad i \equiv k \quad \text{and} \quad k \to j \to i. \tag{2.4}$$

[2] The dots stand for zeros.

Note that i and k in the definition above are distinct as we cannot have $j \to i$ and at the same time $i \to j$. Also, in general, the relation $\to$ need not be transitive. An example of a non-transitive rule is the game 'Rock, paper, scissors' and extensions thereof. We can now state our characterisation of compatible aggregations for these systems. Its proof is omitted as it follows from more general Theorem 2.4 we prove below.

Theorem 2.3. *Let T be a heuristic as in (2.1). An equivalence relation on G is compatible (i.e. gives a coarse graining) with T if and only if it is contiguous with respect to $\to$.*

Suppose that it is not an orientation that is given on K_n but let $P\colon G^2 \to [0,1]$ be a function such that $P(i,j) = 1 - P(j,i)$ for all $i,j \in G$. We call such a P a *selection map*. The interpretation is that in a clash between i and j, gene i dominates with probability $P(i,j)$ and hence j dominates with the probability $1 - P(i,j) = P(j,i)$. Note that this forces $P(i,i) = 1/2$ for all $i \in G$.

If $P(i,j) \in \{0,1\}$ whenever $i \neq j$ then this reduces to the previous case as $P(i,j) = 1$ can be interpreted as orienting the edge connecting i and j by choosing $j \to i$. It turns out that there is a characterisation of admissible aggregations that is the same as the one given in Theorem 2.3. Of course, the definition of a contiguous partition requires some amendments.

An equivalence relation $\equiv$ on G is *contiguous with respect to a selection map P* if for all $i,j,k \in G$ we have

$$i \equiv j \equiv k \qquad \text{whenever} \qquad i \equiv k \quad \text{and} \quad P(i,j) \neq P(k,j). \tag{2.5}$$

We remark that this reduces to (2.4) when P maps off-diagonal pairs to $\{0,1\}$.

If we denote by $\pi\colon [n] \to [m]$ the partition induced by the classes of $\equiv$ then the condition above amounts to requiring that for any two blocks $\pi^{-1}(s)$ and $\pi^{-1}(t)$, with $s \neq t$, and for any $i,k \in \pi^{-1}(s)$ and $j \in \pi^{-1}(t)$ we have $P(i,j) = P(k,j)$.

If we let l be any other element in $\pi^{-1}(t)$, the same condition now applied with the roles of $\pi^{-1}(s)$ and $\pi^{-1}(t)$ swapped gives $P(j,k) = P(l,k)$. But since P is a selection this implies $P(k,j) = P(k,l)$ and hence $P(i,j) = P(k,l)$.

This shows that for a contiguous partition there is a well defined selection map $P\colon [m]^2 \to [0,1]$ such that $P(s,t) + P(t,s) = 1$ for all $s, t \in [m]$, and for every $i \in \pi^{-1}(s)$ and $j \in \pi^{-1}(t)$ we have $P(s,t) = P(i,j)$. Although these are two different maps, we shall keep the same notation.

How does the heuristic map look in this more general setting? Let us compute the probability that the winning gene in a random pairing is of type i.

$$T_i(p_1, \dots, p_n) = p_i^2 + \sum_{k \in [n]\setminus\{i\}} 2p_i p_k P(i,k) = 2p_i \sum_{k=1}^{n} P(i,k) p_k, \tag{2.6}$$

where we noted that $P(i,i) = 1/2$. Any such T induced by a selection map P is said to be a *weighted binary tournament*. Note that the formula for T again reduces to (2.1) if the range of P is $\{0,1\}$ on the off-diagonal pairs. We are now ready to prove our characterisation.

Theorem 2.4. *Let P be a selection map on G, and let the heuristic T be as in (2.6). An equivalence relation on G is compatible (i.e. gives a coarse graining) with T if and only if it is contiguous with respect to P.*

Proof. Let $\pi\colon [n] \to [m]$ be the map associated to the partition of $G = [n]$ induced by a contiguous equivalence relation, and let Ξ_π be the associated aggregation. Take $s \in [m]$ and calculate

$$\begin{aligned}((\Xi_\pi \circ T)(p))_s = \sum_{\pi(i)=s} T_i(p) = \sum_{\pi(i)=s} 2p_i \left(\sum_{k=1}^{n} P(i,k) p_k \right) = \\ = 2 \sum_{\pi(i)=s} p_i \left(\sum_{t=1}^{m} \sum_{\pi(k)=t} P(i,k) p_k \right) = 2 \sum_{\pi(i)=s} p_i \left(\sum_{t=1}^{m} P(s,t) \sum_{\pi(k)=t} p_k \right) \\ = 2(\Xi_\pi(p))_s \sum_{t=1}^{m} P(s,t)(\Xi_\pi(p))_t = \tilde{T}_s(\Xi_\pi(p)) = (\tilde{T}_s \circ \Xi_\pi)(p).\end{aligned}$$

Thus $\Xi_\pi \circ T = \tilde{T} \circ \Xi_\pi$, where $\tilde{T} = (\tilde{T}_1, \dots, \tilde{T}_m)$. This shows that Ξ is a coarse graining of the system and, moreover, we see that the coarse grained map $\tilde{T}$ is in the same form as the original one. It expresses the rule of transformation for meta-genes that are given by the blocks of the partition π.

Conversely, suppose that $\pi\colon [n] \to [m]$ is a partition of $[n]$ whose associated aggregation Ξ_π coarse grains T. We need to prove that the equivalence relation that it induces is contiguous with respect to P. To that end, take $s, t \in [m]$, $s \neq t$. Following the discussion after (2.5) above, it suffices

to prove that for any $i, k \in \pi^{-1}(s)$ and $j \in \pi^{-1}(t)$ we have $P(i,j) = P(k,j)$. For the sake of getting a contradiction assume that i, k and j are chosen such that $P(i,j) < P(k,j)$.

To simplify the notation we can, without loss of generality, assume that $i = 1$, $k = 2$, $j = 3$. Take vectors $v = (1/2, 0, 1/2, 0, \ldots, 0)$ and $w = (0, 1/2, 1/2, 0, \ldots, 0)$ and note that

$$\Xi_\pi(v) = \Xi_\pi(w)$$

and since Ξ_π is a coarse graining we must also have

$$(\Xi_\pi \circ T)(v) = (\Xi_\pi \circ T)(w)$$

and hence also

$$((\Xi_\pi \circ T)(v))_{\pi(1)} = \sum_{\pi(k)=\pi(1)} T_k(v) = T_1(v) = \frac{1}{2}(P(1,1) + P(1,3))$$

is equal to

$$((\Xi_\pi \circ T)(w))_{\pi(1)} = \sum_{\pi(k)=\pi(1)} T_k(w) = T_2(w) = \frac{1}{2}(P(2,2) + P(2,3)).$$

As $P(1,1) = P(2,2) = 1/2$, we get $P(1,3) = P(2,3)$, which contradicts the initial assumption $P(1,3) < P(2,3)$. This finishes the proof of the other implication. □

To finish this section we give an example of a weighted binary tournament whose one generalisation will be the main theme of Chapter 3.

Example 2.5. Let $p \in [0,1]$ be a fixed parameter. We set $G = [n] = \{1, 2, \dots, n\}$ and let $P\colon [n]^2 \to [0,1]$ be a selection map on G given by:

$$P(i,j) = \begin{cases} p & \text{if} \quad i > j, \\ 1-p & \text{if} \quad i < j, \\ 1/2 & \text{if} \quad i = j. \end{cases}$$

This means that among any two different types of genes the one labelled with the higher number dominates with probability p. And this is regardless of the chosen pair. The associated heuristic map $T\colon \mathbb{R}^n \to \mathbb{R}^n$ can be computed from the equation (2.6) and in this case reduces to:

$$T_i(p_1, \dots, p_n) = p_i^2 + 2p_i \left(\sum_{k<i} p p_k + \sum_{k>i} (1-p) p_k \right).$$

Using Theorem 2.4 we can deduce which equivalence relations over $[n]$ are compatible with the dynamics of this weighted binary tournament.

If $p = 1/2$ it is clear that any equivalence relation will do as P is identically equal to $1/2$ and so (2.5) is vacuously satisfied.

If $p \neq 1/2$ then taking into account the definition of P the same condition can be rewritten as

$$i \equiv j \equiv k \qquad \text{whenever} \qquad i \equiv k \quad \text{and} \quad i < j < k \text{ or } i > j > k. \tag{2.7}$$

Above we noted that (2.5) is always going to be satisfied when either $i = j$, or $k = j$, or $i = k$; and we could safely assume that i, j, and k are all distinct.

The condition (2.7) can now be reinterpreted as saying that each block of a compatible equivalence relation must be an interval in $\{1, 2, \dots, n\}$ with the usual ordering. ◂

In this section we present a result showing that even for Markov Chains, i.e. linear maps, finding aggregations is a difficult task in general. More precisely we show

Theorem 2.6. *Deciding if there exists a non-trivial aggregation for a Markov chain is* NP*-complete.*

Recall that a Markov chain defined by a stochastic $n \times n$ matrix $\mathcal{M}$ possesses a non-trivial aggregation if there exist a (non-trivial) surjective partitioning function $\pi\colon [n] \to [m]$ where $1 < m < n$, and an $m \times m$ matrix R such that

$$\Xi \mathcal{M} = R \Xi$$

where $\Xi = \Xi_\pi$ is the change of variable transformation associated to π.

In order for this decision problem to be well defined we clearly need to restrict the entries of transition matrices to be in a countable domain, say in $\mathbb{Q}$. Note however that by multiplying each entry of the matrix $\mathcal{M}$ by a sufficiently large number we can obtain a matrix with integer entries and constant column sums which clearly has a non-trivially aggregation if and only if $\mathcal{M}$ does.

Further note that this problem is indeed in NP as given $\mathcal{M}$, Ξ, and R verifying if this is a solution amounts to multiplying and comparing matrices. It will thus suffice to show that some well-know NP-complete problem, say the subset sum problem (SSP), is *polynomially reducible* to the problem of deciding if an integer matrix with constant column sums has a non-trivial aggregation. For more details on these notions we refer the reader to any textbook dealing with complexity, e.g. [TW06].

Recall that SSP (originally the KNAPSACK problem in Karp's list [Kar72]) is an NP-complete decision problem that asks, given a set of positive integers $\{a_1, \dots, a_n\}$ and an integer $0 < K < \sum_{i=1}^{n} a_i$, whether there exists a subset $I \subseteq [n]$ such that $\sum_{i \in I} a_i = K$.

Proof of Theorem 2.6. Let an input for SSP $(\{a_1, \dots, a_n\}, K)$ be given. Set $L = \sum_{i=1}^{n} a_i - K$ and let $\mathcal{M}$ be an $(n+2) \times (n+2)$ matrix as below

$$\mathcal{M} = \begin{bmatrix} a_1 & 2a_1 & 3a_1 & \dots & (n+2)a_1 \\ a_2 & 2a_2 & 3a_1 & \dots & (n+2)a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & 2a_n & 3a_1 & \dots & (n+2)a_n \\ (n+2)K & (n+1)K & na_1 & \dots & K \\ (n+2)L & (n+1)L & na_1 & \dots & L \end{bmatrix}.$$

Note that $0 < L < \sum_{i=1}^{n} a_i$. We claim that $\mathcal{M}$ has a non-trivial aggregation if and only if the given SSP instance has a solution. This will be enough to finish the proof as this reduction is clearly polynomial in the size of the input.

First assume that $I \subseteq [n]$ solves the SSP, i.e. $\sum_{i \in I} a_i = K$. Let I^c denote the complement of I in $[n]$. Thus $\sum_{i \in I^c} a_i = L$.

Define $\pi \colon [n+2] \to [2]$ by

$$\pi(i) = \begin{cases} 1 & \text{if} \quad i \in I \quad \text{or} \quad i = n+1, \\ 2 & \text{otherwise,} \end{cases}$$

and let $\Xi = \Xi_\pi$ be the associated change of variables transformation. One readily checks that

$$\Xi \mathcal{M} = R \Xi,$$

where

$$R = \begin{bmatrix} (n+3)K & (n+3)K \\ (n+3)L & (n+3)L \end{bmatrix}$$

and thus $\mathcal{M}$ has a non-trivial aggregation.

Conversely, assume that $\pi \colon [n+2] \to [m]$ is a non-trivial aggregation compatible with $\mathcal{M}$, where $1 < m < n+2$. Let $\Xi = \Xi_\pi$ and R be such that

$$\Xi \mathcal{M} = R \Xi \tag{2.8}$$

holds. By the pigeon-hole principle and because of non-triviality there must exist $k \in [m]$ such that $1 < |\pi^{-1}(k)| < n+2$. Choose two different integers $s, t \in \pi^{-1}(k)$. Inspecting the elements on positions (k, s) and (k, t) in matrices $\Xi\mathcal{M}$ and $R\Xi$ and using (2.8) we conclude that they are the same and they equal

$$s \sum_{i \in I} a_i + (n + 3 - s)(\varkappa K + \lambda L) = t \sum_{i \in I} a_i + (n + 3 - t)(\varkappa K + \lambda L) = R_{(k,k)},$$

where $I = \pi^{-1}(k) \cap [n]$ and $\varkappa$ and λ are 0-1 indicators depending on whether $(n + 1)$ and $(n + 2)$ respectively are in $\pi^{-1}(k)$. From here we get

$$(s - t) \sum_{i \in I} a_i = (s - t)(\varkappa K + \lambda L)$$

and as $s \neq t$

$$\sum_{i \in I} a_i = \varkappa K + \lambda L.$$

The only possibilities are now:

- $\varkappa = \lambda = 1$ and $I = [n]$,
- $\varkappa = 1, \lambda = 0$ and $I \subsetneq [n]$,
- $\varkappa = 0, \lambda = 1$ and $I \subsetneq [n]$,
- $\varkappa = \lambda = 0$ and $I = \emptyset$.

The first can be discarded as it would imply that $\pi^{-1}(k) = [n + 2]$ and the aggregation given by π is trivial. Similarly, the last is impossible as it would imply that $\pi^{-1}(k) = \emptyset$ contradictory to our choice of $k \in [m]$. If the second holds then I is the subset solving our SSP $\sum_{i \in I} a_i = K$, and in case the third one holds, the complement I^c of I in $[n]$ solves the SSP $\sum_{i \in I^c} a_i = K$, as we know that $\sum_{i \in I} a_i = L$. This completes the proof of the theorem. □

To aid the understanding, we illustrate the proof above with an example.

Example 2.7. Take two instances of SSP $(\{5, 6, 1, 8\}, 12)$ and $(\{9, 4, 12, 1\}, 15)$. The construction from the proof gives matrices

$$\mathcal{M}_1 = \begin{bmatrix} 5 & 10 & 15 & 20 & 25 & 30 \\ 6 & 12 & 18 & 24 & 30 & 36 \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 8 & 16 & 24 & 32 & 40 & 48 \\ 72 & 60 & 48 & 36 & 24 & 12 \\ 48 & 40 & 32 & 24 & 16 & 8 \end{bmatrix}, \qquad \mathcal{M}_2 = \begin{bmatrix} 9 & 18 & 27 & 36 & 45 & 54 \\ 4 & 8 & 12 & 16 & 20 & 24 \\ 12 & 24 & 36 & 48 & 60 & 72 \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 90 & 75 & 60 & 45 & 30 & 15 \\ 66 & 55 & 44 & 33 & 22 & 11 \end{bmatrix}.$$

The first system has a valid aggregation $\{\{1, 2, 3, 5\}, \{4, 6\}\}$ and one can check that $\Xi_1 \mathcal{M}_1 = R_1 \Xi_1$ where

$$\Xi_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad R_1 = \begin{bmatrix} 84 & 84 \\ 56 & 56 \end{bmatrix}.$$

This agrees with the fact that the first instance of SSP has a solution $a_1 + a_2 + a_3 = 5 + 6 + 1 = 12$. Here $n = 4$ and the set of indices giving the solution is $I = \{1, 2, 3\} = \{1, 2, 3, 5\} \cap [4]$.

The other instance of SSP does not have a solution and the consequence is, as we have proved, that the Markov chain defined via transition probabilities in the normalised matrix $\frac{1}{182}\mathcal{M}_2$ has no non-trivial aggregations. ◂

It is worth noting that the problem of finding a non-trivial aggregation can be efficiently solved for certain classes of systems. For example, weighted binary tournaments discussed in Section 2.2 form one such a class. The polynomial algorithm that checks for non-trivial aggregations in that case is given below.

The algorithm relies heavily on the fact that we have an efficient way of checking whether a partition gives a compatible aggregations by means of the contiguity test (2.5), see also Theorem 2.4. In particular this means that given any non-trivial compatible partition $\{S_1, \ldots, S_k\}$ of $[n]$, and assuming that $|S_1| > 1$, the refined partition $\{S_1\} \cup \{\{i\} \mid i \in [n] \setminus S_1\}$ is also a non-trivial aggregation for the same weighted binary tournament. This reduces the number of aggregations we need to check greatly, but still leaves exponentially many of them to be considered. The final trick making this work in polynomial time is that by virtue of (2.5) at each step we can either verify that the current lump S is a block of a compatible aggregation, or we can increase it by at least one element $S \longleftarrow S \cup Q$ and be certain we are not omitting any solutions by doing so.

input : Rational $n \times n$ matrix P satisfying $\mathsf{P}_{ij} + \mathsf{P}_{ji} = 1$ and $\mathsf{P}_{ii} = 1/2$
output : Yes if the tournament given by P has an aggregation, No otherwise

```
foreach s, t ∈ [n], s ≠ t do
    S ⟵ {s, t}
    repeat
        Q ⟵ ∅
        foreach j ∈ [n] \ S do
            j_is_not_compatible ⟵ No
            foreach i, k ∈ S, i ≠ k do
                if P_ij ≠ P_kj then
                    j_is_not_compatible ⟵ Yes
                end
            end
            if j_is_not_compatible then
                Q ⟵ Q ∪ {j}
            end
        end
        S ⟵ S ∪ Q
    until Q = ∅
    if S ≠ [n] then
        return Yes
    end
end
return No
```

Let us explain in plain words what this algorithm does. It starts by considering all possible pairs $s, t \in [n], s \neq t$ and then attempts to prove that there is a compatible aggregation lumping those two variables together. S is the current candidate for a block of a compatible aggregation. Using criterion (2.5) the algorithm picks all the elements $Q \subseteq [n] \setminus S$ that falsify the contiguity property. If $Q = \emptyset$ then we have a certificate of S being a lump of a non-trivial compatible aggregation, otherwise $S \cup Q$ is taken to be a new candidate. If in the end this results with S being everything (S=$[n]$) we conclude that no non-trivial aggregation lumps s and t together. Should this be the case for all the pairs s and t, we have a proof that no non-trivial aggregations exist.

This discussion thus proves

Theorem 2.8. *Determining whether a non-trivial aggregation for a weighted binary tournament exists is in* P.

2.4 ARTIFICIAL CHEMISTRIES

We shall now show that the *selection map* model considered in Section 2.2 can further be generalised to include an even larger class of quadratic maps. A natural way to interpret these is through artificial chemistries.

Example 2.9 (Chemical reactions of degree 2)**.** Let each number in $[n]$ represent a different chemical. Assume that for each choice of two chemicals c_1 and c_2 (not necessarily different) we are given a distribution $\tau_v = (\tau_{1,v}, \tau_{2,v}, \ldots, \tau_{n,v})$ over $[n]$ whose entries are interpreted as the proportions of each of the chemicals produced by a chemical reaction involving the particles c_1 and c_2 as reactants. The vector $v \in X_2^n$ is, as before, used to represent the chosen pair by setting $v = e_{c_1} + e_{c_2}$. We can also write this as a set of $\binom{n+1}{2}$ equations of the form

$$c_1 + c_2 \to \tau_{1,v} \cdot 1 + \tau_{2,v} \cdot 2 + \cdots + \tau_{n,v} \cdot n.$$

The evolution of such a system (assuming the chemical solution contains a large number of particles and is well-mixed) is given by $T = (T_1, \dots, T_n) : \mathbb{R}^n \to \mathbb{R}^n$ where

$$T_i(p) = \sum_{v \in X_2^n} \frac{2}{v_1! \cdots v_n!} \tau_{i,v} p_1^{v_1} \cdots p_n^{v_n} = \sum_{v \in X_2^n} \frac{2}{v!} \tau_{i,v} p^v.$$

The factor $2/v!$ accounts for the number of different ways to choose the reactants represented by v. Here it is either 2, if the reactants are different, or 1, if they are not.

One can show that for an aggregation Ξ to be compatible with this dynamics $\Xi(\tau_{1,v}, \dots, \tau_{n,v}) = \Xi(\tau_{1,w}, \dots, \tau_{n,w})$ must hold true whenever $\Xi(v) = \Xi(w)$. In other words, taking the blocks of the partition associated with Ξ as meta-chemicals, the distributions over these must be well-defined for all the pairings of meta-chemicals as reactants. This is a direct consequence of Theorem 2.11 we prove later on. ◁

We could interpret any system given by a selection map as in (2.6), as a chemistry of degree 2. But the chemistries are still more general since they also incorporate systems as in the following example.

Example 2.10. Consider a second order chemistry on $\mathbb{Z}_n = \{0, 1, 2, \dots n-1\}$ with the reaction rules

$$i + j \to 1 \cdot (i + j \bmod n), \qquad \text{for all } i, j \in \mathbb{Z}_n.$$

Following Example 2.9, a partition $\pi : \mathbb{Z}_n \to \mathcal{M}$ will be a valid aggregation if and only for any two of its blocks, it is well defined which block they produce. Let us fix one of the blocks $\pi^{-1}(m)$ for $m \in \mathcal{M}$. Then for any element $d \in \mathbb{Z}_n$ the set $(\pi^{-1}(m) + d \bmod n)$ must be contained within one block. From here it is not hard to see that all the blocks are translates (or cosets) of the set $\{0, l, 2l, \dots, (n-l)\}$, where l is some positive divisor of n. Thus the range of π in $\mathcal{M}$ can be given a group structure that makes it isomorphic to $\mathbb{Z}_l$.

This establishes a correspondence between the compatible aggregations and the divisors of n. In fact, both the compatible aggregations and the set of divisors of n come equipped with a natural partial order making them into lattices. It is not hard to see that these lattices are isomorphic and the refinement relation in the former corresponds to the relation 'is divisible by' in the latter. ◁

In this section we give a general framework incorporating, amongst others, models from Sections 2.2 and 2.4. Our results in this section apply to maps that are given by an absolutely convergent series $T = (T_1, \dots, T_n) \colon \mathbb{R}^n \to \mathbb{R}^n$

$$T_i(p) = \sum_{v_1,\dots,v_n \in \mathbb{N}_0} \frac{\alpha_{i,v}}{v_1! \cdots v_n!} p_1^{v_1} \cdots p_n^{v_n} = \sum_{v \in \mathbb{N}_0^n} \frac{\alpha_{i,v}}{v!} p^v \tag{2.9}$$

where $\alpha_{i,v} \in \mathbb{R}$ for $1 \leq i \leq n$, $v \in \mathbb{N}_0^n$ is a family of parameters. Our goal is to derive a sufficient and necessary condition for an aggregation $\pi \colon [n] \to [m]$ to be compatible with T solely in terms of these parameters. Indeed, we prove

Theorem 2.11. *Let $T = (T_1, \dots, T_n) \colon \mathbb{R}^n \to \mathbb{R}^n$ be an absolutely convergent series as in (2.9) defining an analytic function on $\mathbb{R}^n$. An aggregation $\Xi = \Xi_\pi \colon \mathbb{R}^n \to \mathbb{R}^m$ (associated to $\pi \colon [n] \to [m]$) is a valid coarse graining if and only if*

$$\Xi(\alpha_{1,v}, \dots, \alpha_{n,v}) = \Xi(\alpha_{1,w}, \dots, \alpha_{n,w}) \text{ whenever } v, w \in \mathbb{N}_0^n \text{ with } \Xi(v) = \Xi(w). \tag{2.10}$$

In particular, for all $u \in \mathbb{N}_0^m$ the coefficients $(\beta_{1,u}, \dots, \beta_{m,u}) = \Xi(\alpha_{1,v}, \dots, \alpha_{n,v})$, where $v \in \mathbb{N}_0^n$ is chosen such that $\Xi(v) = u$, are well-defined and the coarse grained system is a convergent series on $\mathbb{R}^m$ given by $Q = (Q_1, \dots, Q_m) \colon \mathbb{R}^m \to \mathbb{R}^m$ where

$$Q_i(q) = \sum_{u_1,\dots,u_m \in \mathbb{N}_0} \frac{\beta_{i,u}}{u_1! \cdots u_m!} q_1^{u_1} \cdots q_m^{u_m} = \sum_{u \in \mathbb{N}_0^m} \frac{\beta_{i,u}}{u!} q^u.$$

Note the similarity between condition (2.10) above and (2.5). Indeed, it is not hard to see that the former generalises the latter when T is a heuristic induced by a two point selection, and hence the theorem above generalises Theorem 2.4.

Our proof of this result is elementary and relies on the well-known multi-variable multinomial theorem and the fact that aggregations compatible with T must also be compatible with any of the

partial derivatives of T. We also use a criterion for coarse grainings from [RVW06] specialised for the case of linear maps, Proposition 2.13 below.

Recall that, as before $p = (p_1, \dots, p_n) \in \mathbb{R}^n$, and e_i is i-th vector of the canonical basis in $\mathbb{R}^n$. Variables v, w, etc. are reserved for labelling vectors in $\mathbb{N}_0^n$ and $|v| := v_1 + v_2 + \cdots + v_n$. Following the multi index notation due to Schwartz, the factorial function naturally extends over integral vectors via $v! = v_1! \cdots v_n!$ and so do integral powers $p^v := p_1^{v_1} p_2^{v_2} \dots p_n^{v_n}$. It is also convenient to introduce a shorthand $C(n, r) := \binom{r+n-1}{r}$, and for the space of (real) $m \times n$ matrices $\mathcal{M}_{m\times n}$. We also use vectors to denote partial derivatives as in $\partial_v T(p) = \frac{d^{|v|}}{dp_1^{v_1} \dots dp_n^{v_n}} T(p_1, \dots, p_n)$.

The multinomial theorem is extremely easy to state using the introduced notation. The proof is classical and is left to the reader.

Proposition 2.12 (The vector multinomial theorem)**.** *Let $\Xi \in \mathcal{M}_{m\times n}$ be a stochastic* 0-1 *matrix ($m \leq n$) and $\pi\colon [n] \to [m]$ the associated aggregation, and let $u \in \mathbb{N}_0^m$. For any $p \in \mathbb{R}^n$ we have*

$$\big(\Xi p\big)^u = \prod_{j=1}^{m} \left(\sum_{\pi(i)=j} p_i \right)^{u_j} = \sum_{y \in \mathbb{N}_0^n \cap \Xi^{-1}(u)} \frac{u!}{y!} p^y.$$

Proposition 2.13 (Rowe et al. [RVW06, Theorem 1])**.** *Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable map and let $\Xi\colon \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$, be a linear transformation. Then Ξ is a coarse graining of the system T if and only if the kernel of Ξ is invariant under the differential of T at any point of the domain $\mathbb{R}^n$, formally*

$$(DT)_x \cdot \ker \Xi \subseteq \ker \Xi \quad \text{for all } x \in \mathbb{R}^n. \tag{2.11}$$

Lemma 2.14. *Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be a smooth function of class C^d and let $\Xi\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. If Ξ is a coarse graining of the system T then it coarse grains $\partial_v T\colon \mathbb{R}^n \to \mathbb{R}^n$ for any $v \in \mathbb{N}_0^n$ for which $|v| \leq d$.*

Proof. It is enough to prove that $\partial_i T = \partial_{e_i} T = \frac{d}{dp_i} T$ is coarse grained via Ξ for any $1 \leq i \leq n$ as the general result will then follow by induction. We denote the coarse grained system by $\tilde{T}$ and calculate

$$\Xi \frac{d}{dp_i} T(p) = \Xi \lim_{h \to 0} \frac{1}{h} \Big(T(p + he_i) - T(p) \Big) = \lim_{h \to 0} \frac{1}{h} \Big(\Xi T(p + he_i) - \Xi T(p) \Big) =$$
$$= \lim_{h \to 0} \frac{1}{h} \Big(\tilde{T}(\Xi p + h \Xi e_i) - \tilde{T}(\Xi p) \Big) = (D\tilde{T})_{\Xi p} \Xi e_i$$

which is a function of Ξp as can be seen from the formula. □

Note that the linearity of both the derivative and Ξ is crucial here as it has allowed us to swap the operators ∂_{e_i} and Ξ. We also note in passing that the calculation above proves that a surjective linear coarse graining of a smooth map produces a system with the same degree of smoothness. In fact, as Tóth et al. prove in [TLRT97, Theorem 2.1], the Lipschitz property is also preserved in the coarse grained system, and with the same order. We now proceed to prove the main result.

Proof of Theorem 2.11. To simplify notation we write α_v for the vector of parameters $(\alpha_{1,v}, \ldots, \alpha_{n,v})$ and similarly for α_w. This allows us to fully exploit the vector notation and write (2.9) as

$$T(p) = \sum_{v \in \mathbb{N}_0^n} \frac{\alpha_v}{v!} p^v.$$

Assume first that $\Xi(v) = \Xi(w)$ implies $\Xi(\alpha_v) = \Xi(\alpha_w)$. We need to show that $\Xi(T(p))$ is a function of $\Xi(p)$. We calculate

$$\Xi(T(p)) = \Xi \left(\sum_{v \in \mathbb{N}_0^n} \frac{1}{v!} \alpha_v p^v \right) = \sum_{v \in \mathbb{N}_0^n} \frac{1}{v!} \Xi(\alpha_v) p^v = \sum_{u \in \mathbb{N}_0^m} \sum_{v : \Xi(v) = u} \frac{1}{v!} \Xi(\alpha_v) p^v =$$
$$= \sum_{u \in \mathbb{N}_0^m} \frac{1}{u!} \beta_u \sum_{v \colon \Xi(v) = u} \frac{u!}{v!} p^v = \sum_{u \in \mathbb{N}_0^m} \frac{1}{u!} \beta_u (\Xi p)^u = Q(\Xi p)$$

where we used the multinomial theorem.

Conversely, assume that the aggregation Ξ is a coarse graining of T. It suffices to show that $\Xi(v) = \Xi(w)$ implies $\Xi(\alpha_v) = \Xi(\alpha_w)$ when $v - w = e_i - e_j$, i.e. when the vectors v and w differ by 1 at exactly two coordinates. For any other choice of an equivalent pair v and w we can construct a chain of

intermediate equivalent vectors transforming v to w by changing one coordinate at the time. Applying the base case to each link in the chain will produce a chain of equalities giving the general case. For example, if $n = 4$ and the aggregation corresponds to the partition $\{\{1, 2, 3\}, \{4\}\}$ then

$$(3, 0, 0, 1) \rightsquigarrow (2, 1, 0, 1) \rightsquigarrow (1, 2, 0, 1) \rightsquigarrow (0, 2, 1, 1)$$

is a valid sequence of intermediate transformations.

Assume, therefore, that for some $y, v, w \in \mathbb{N}_0^n$ we have $y = v - e_i = w - e_j$, and $\Xi(v) = \Xi(w)$. Using Lemma 2.14 we get that Ξ coarse grains $\partial_y T$. Note that $e_i - e_j = v - w \in \ker \Xi$, and the necessary condition for coarse graining (2.11) implies that

$$(D(\partial_y T))_{(0,\dots,0)}(e_i - e_j) = \partial_{e_i}\partial_y T(0, \dots, 0) - \partial_{e_j}\partial_y T(0, \dots, 0) = \\ = \partial_v T(0, \dots, 0) - \partial_w T(0, \dots, 0) \in \ker \Xi.$$

Hence,

$$\Xi(\partial_v T(0, \dots, 0)) = \Xi(\partial_w T(0, \dots, 0)).$$

But the series for T is given as a Taylor series about $(0, \dots, 0)$ and therefore $\partial_v T(0, \dots, 0) = \alpha_v$ for all $v \in \mathbb{N}_0^n$. Thus $\Xi(\alpha_v) = \Xi(\alpha_w)$, as required. □

Observation. Note that the proof works even if the series expansion for T is given about a point other than 0. This allows to extend the result to any analytic function T just by requiring that the condition (2.10) holds locally for a series expansion of T about any point in the domain.

Corollary 2.15. *Let $T = (T_1, \dots, T_n) \colon \mathbb{R}^n \to \mathbb{R}^n$ be an analytic function as in (2.9). An aggregation $\Xi \colon \mathbb{R}^n \to \mathbb{R}^m$ is a valid coarse graining if and only if it coarse grains each of the homogeneous components*

of T which are defined for each $d \in \mathbb{N}_0$ as the projection $T^{(d)} = (T_1^{(d)}, \dots, T_n^{(d)}) \colon \mathbb{R}^n \to \mathbb{R}^n$ on the space of homogeneous polynomials of degree d where

$$T_i^{(d)}(p) = \sum_{\substack{v_1,\dots,v_n \in \mathbb{N}_0 \\ v_1+\cdots+v_n=d}} \frac{\alpha_{i,v}}{v_1! \cdots v_n!} p_1^{v_1} \cdots p_n^{v_n} = \sum_{v \in \mathbb{N}_0^n : |v|=d} \frac{\alpha_{i,v}}{v!} p^v.$$

Proof. It suffices to note that $|v| = |\Xi(v)|$ and similarly for w, and hence $|v| = |w|$ whenever $\Xi(v) = \Xi(w)$. Applying Theorem 2.11 twice will now give the result. □

2.6 CONCLUDING REMARKS

In the present chapter we gave a criterion for an aggregation of variables to be compatible with a map that is given by its series expansion; and showed how it can be applied to artificial chemistries. Despite the attractive and deceptively simple looking characterisations presented here, we also showed that the existence of a non-trivial compatible aggregation is a difficult problem in general. This means that for some systems, an algorithm searching for compatible aggregations cannot essentially be better then the one that exhaustively tests all the partitions of the set of variables.

For certain classes of problems, as we saw in the case of binary tournaments, it is possible to give an efficient algorithm searching for aggregations. The ideas we used here are somewhat similar in flavour to those used in other model reduction algorithms for DTMCs, one of such being 'partition minimisation' introduced by Paige and Tarjan in [PT87], and adapted to Markov chains in [DHS03].

We also looked into different ways of modelling finite number particles systems. Given an infinite population model, one possible way of simulating the system was using Random Heuristic Search framework. The main characteristic of the RHS approach is that it is generational. Each new generation, although dependent on the previous one, consists of an entirely new collection of entities. Sometimes, other approaches may be more appropriate. Indeed, if our algorithm is supposed to simulate a real biological system, it is natural to allow the members of different generations to coexist. Similarly, if we are to simulate a chemical process, we would like to keep most of the particles for the new generation,

and only those few that bump into another particle with which they can react, will be replaced with a product of the reaction in the next time step.

Artificial chemistries are a fruitful area and it is no surprise that different authors (Fontana et al. [FB94], Tóth et al. [TLRT97], Dittrich et al. [DDF07] to name a few) model chemical reactions in different ways. What seems promising is that all of them incorporate a polynomial model similar to the one we had above. This should mean that the results on compatible aggregations can be easily transferred across the modelling paradigms, and this is something we would like to explore in future.

Chapter 3

REACHING CONSENSUS ON A CONNECTED GRAPH

In this chapter we study a simple random process by means of which vertices of a connected graph can reach a mutual decision only through pairwise interactions with their neighbours. Each vertex of the graph is initially assigned a label (we shall call them strategies) from the set $[n] = \{1, 2, \dots, n\}$ *uniformly at random. Then at each time step one of the edges of the graph is selected uniformly at random and the two vertices it connects get relabelled as follows: with probability* p *both adopt the higher strategy and with probability* $1 - p$ *both adopt the lower one. Almost surely this process has to terminate in finite time with all the vertices playing the same strategy.*

In the special case when the underlying graph is the complete graph over r *vertices* K_r *it is not too difficult to see that we can lump together any two isomorphic labellings of* K_r*. This means that we consider two labellings the same if the number of vertices playing each of the strategies is the same. This gives a coarse graining onto a Markov chain with the underlying state space* X_r^n *(as defined in Section 2.1). Indeed this chain is exactly the same as the one obtained by considering the heuristic map* T *from Example 2.5 and the chain from Section 2.1 induced by the transition function* F_2 *defined in* (2.3).

If the underlying graph is not complete then it is no longer the case that labellings with the same counts are necessarily isomorphic and thus the models considered in this chapter are truly a generalisation of those studied in Chapter 2.

For these models one can compute probabilities that a certain strategy gets chosen and these, surprisingly, do not depend on the network structure between the vertices. For the class of regular graphs — those in which each vertex is adjacent to the same number of neighbours — and with the restriction to two strategies one can also show that the expected waiting time until the decision has been made is the shortest for the complete graph. These two results are due to Haslegrave [HP17].

In order to obtain the formula for the mean of the waiting time we also study a version of the classical gambler's ruin process with delays. We obtain explicit expressions for the expected time until absorption and also prove that this monotonically increases with the parameter $p \in [0, 1/2]$ *in case of symmetric delays. As a direct application we obtain the monotonicity result for the mean of the waiting time until consensus on the complete graph when the number of strategies is* 2.

In the last section we briefly mention two results, again due to Haslegrave, that give upper bounds on the waiting time for $p = 0$.

The results in this chapter form joint work with John Haslegrave and the paper [HP17] *containing this material has been accepted for publication.*

The reader should be warned that in order to maintain consistency with [HP17] the notation is slightly altered below. From this point onwards we denote the number of strategies by m, and the number of vertices in the connected graph by n.

Let us reiterate. We consider the evolution of a system on a connected graph G with n vertices. Each vertex has a strategy taken from $\{1, \dots, m\}$ (we shall frequently write $[m]$ for this set). The starting strategies of the vertices are chosen independently and uniformly at random. At each time step an edge is chosen uniformly at random, and both vertices are updated to have the same strategy, which is the higher of the two with probability p and the lower with probability $1 - p$. This simple model covers a broad range of real-life scenarios where a consensus is reached via pairwise interactions among the individual agents, whether we are interested in modelling an infectious disease spread or the process by which a certain gene became prevalent in the human genome.

The model was inspired by the well-studied tournament games in the theory of genetic algorithms (see Rowe, Vose and Wright [RVW06, RVW05a] and Vose [Vos99]) and indeed, it is a generalisation of these as we have seen in the introduction. The idea is that the underlying connected graph allows for modelling a spatial aspect of the problem at hand. It therefore comes as a surprise when in Section 3.1 we prove that the probability that a certain strategy prevails does not depend on the network structure of the vertices. This is achieved by reducing the problem to the study of the two-strategy case by looking at contiguous partitions of the strategy set. Validity of these coarse grainings was previously checked only for the complete graph case in [RVW05a] (and see also Example 2.5).

The expected time to reach a consensus will, of course, depend on the graph structure and we are able to give explicit formula for this mean only for the case of the complete graph and two strategies. This is done in Section 3.2 by relating the problem to a version of the gambler's ruin process with delays. By elementary means we show that this expression is monotonic in parameter $p \in [0, 1/2]$ in the case of symmetric delays, which translates to a monotonicity result for the expected decision time of our process on the complete graph. Computer simulations using the PRISM model-checking software [KNP11] seem to indicate that this holds true more generally for any fixed connected graph but the proof remains elusive.

It seems natural to conjecture that for a fixed parameter p the process over the complete graph, on average, reaches consensus most quickly. This is again supported by the computer simulations but has been only proven it in the class of regular graphs where each vertex is adjacent to the same number of neighbours and with the restriction to two strategies, see Section 3.1.

3.1 Absorption probabilities; the optimality of the complete graph

Since G is connected, eventually the process will, with probability 1, reach a state where only one strategy remains. Let S be the random variable denoting the strategy that is left. Below we reproduce an argument due to Haslegrave showing that the distribution of S depends only on parameters n, m, and p. In particular, it does not depend on the structure of G, only its order.

Theorem 3.1 (Haslegrave)**.** *For any graph G with n vertices, if the initial state is chosen uniformly at random from $[m]^n$ then $\mathbb{P}(S = l)$ does not depend on the structure of G.*

Proof. Trivially if $p = 1/2$ all strategies are equivalent, and each is equally likely to remain to the end, so we may assume $p \neq 1/2$. We first compute $\mathbb{P}(S \leqslant l)$ by coarse graining the strategies into those at most l and those exceeding l; call these sets of strategies A and B respectively. This is a coarse graining in the sense that when we consider the vertices as playing strategies in $\{A, B\}$ nothing changes unless the edge chosen has one vertex with strategy A and one with strategy B (call this a "significant edge"), in which case they will both adopt B with probability p and both adopt A with probability $1 - p$. Thus

the coarse grained process we obtain on strategies $\{\mathcal{A}, B\}$ is exactly the same as the original process for $m = 2$, save that the distribution of starting states is different. We shall have $S \leqslant l$ if and only if the coarse grained process reaches consensus with all vertices playing $\mathcal{A}$.

Write a_0 for the number of vertices initially playing $\mathcal{A}$, and let a_r be the number playing $\mathcal{A}$ after the r^{th} time a significant edge is chosen. The evolution of a_r is a random walk with absorbing states at 0 and n, where $a_r = a_{r-1} + 1$ with probability $1 - p$ and $a_r = a_{r-1} - 1$ with probability p, independent of which edges are chosen, and indeed independent of G. So the probability that a_r reaches n before 0, i.e. $\mathbb{P}(S \leqslant l)$, does not depend on G, only on n, m and l. Therefore $\mathbb{P}(S = l) = \mathbb{P}(S \leqslant l) - \mathbb{P}(S \leqslant l - 1)$ is also independent of G. □

Remark 3.2. Note that the process $(a_r)_{r \in \mathbb{N}_0}$ above is just an instance of the classical gambler's ruin process (precisely defined in the following section) and using the formulae from say [Fel68] one can obtain the probability that a_r reaches n before 0 conditionally on $a_0 = k$:

$$\mathbb{P}(S \leq l \mid a_0 = k) = \frac{(1-p)^n - p^k(1-p)^{n-k}}{(1-p)^n - p^n}, \quad \text{for } 0 \leq k \leq n.$$

Recall that the initial distribution of strategies is uniform and thus the probability that any vertex starts with a strategy in $\mathcal{A}$ which is less than or equal to l is $\frac{l}{m}$. As there are n vertices in G, the probability that exactly k of them are initially playing $\mathcal{A}$ is:

$$\mathbb{P}(a_0 = k) = \binom{n}{k} \left(\frac{l}{m}\right)^k \left(\frac{m-l}{m}\right)^{n-k}, \quad \text{for } 0 \leq k \leq n.$$

Now using

$$\mathbb{P}(S \leq l) = \sum_{k=0}^{n} \mathbb{P}(S \leq l \mid a_0 = k)\, \mathbb{P}(a_0 = k)$$

and

$$\mathbb{P}(S = l) = \mathbb{P}(S \leq l) - \mathbb{P}(S \leq l - 1)$$

one obtains the following expression for these probabilities:

$$\mathbb{P}(S = l) = \begin{cases} 1/m & \text{if } p = 1/2, \\ \frac{(2pl-2p+m-mp-l+1)^n-(2pl+m-mp-l)^n}{(m-mp)^n-(mp)^n} & \text{otherwise.} \end{cases}$$

The probability of a particular strategy remaining at the end does not depend on the structure of G, but the time taken until this point is reached will do. It is natural to conjecture that the graph which has the quickest expected time is K_n. Haslegrave showed this for the special case where G is known to be regular and $m = 2$. We quote the result below and omit the proof as it does not employ any coarse graining ideas.

Theorem 3.3 (Haslegrave)**.** *For $m = 2$ and any values of n and p, K_n has the shortest expected time to completion of any n-vertex regular graph.*

3.2 GAMBLER'S RUIN WITH DELAYS

In this section we consider a special case of our problem of reaching a consensus on the complete graph with n vertices where $m = 2$. We shall say that the vertices playing strategy 1 are active and those playing strategy 2 are inactive. Because of the symmetries of the complete graph, from the probabilistic point of view, it is easily seen that the evolution of this system is isomorphic to a random walk over the set of states $\{0, 1, \ldots, n\}$ with 0 and n being absorbing states. More precisely, given that k vertices are active and the remaining $n - k$ are inactive, the probability of sampling a significant edge (see Section 3.1) is $\gamma_k = \frac{2k(n-k)}{n(n-1)}$ and conditionally on choosing a significant edge the probability of activating yet another vertex is $1 - p$, and with probability p a previously active vertex is deactivated. We remark in passing that the probability of sampling a significant edge is symmetric under swapping the strategies, $\gamma_{n-k} = \gamma_k$. Below, we recall some of the theory on random walks relevant to our problem.

Gambler's ruin (GR) is a classical problem in probability theory. Given fixed parameters $p \in [0, 1/2]$ and $n \in \mathbb{N}$, a Markov chain $(X_t)_{t\in\mathbb{N}_0}$ over the state space $\{0, 1, \ldots, n\}$ is defined as follows.

The states 0 and n are set to be absorbing and the remaining transition probabilities for states $0 < k < n$ are given by

$$p_{k,k-1} = \mathbb{P}(X_{t+1} = k-1 \mid X_t = k) = p,$$
$$p_{k,k+1} = \mathbb{P}(X_{t+1} = k+1 \mid X_t = k) = 1-p.$$

This Markov chain models the situation where a gambler enters a casino with £X_0 in his pocket and plays a sequence of games in which his odds of winning are $p : 1-p$ and each time he bets £1 on his win. This continues until he either hits his goal £n, or until he bankrupts, whichever occurs first. The time of this happening is represented by the random variable $T = \min\{t \in \mathbb{N}_0 : X_t \in \{0, n\}\}$ and is usually called the absorption time, which is known to be almost surely finite.

There are a few interesting quantities to investigate in this setting: the probability of gambler's ruin and how it depends on the initial capital, $\mathbb{P}(X_T = 0 \mid X_0 = k)$, the expected time for this to happen, $\mathbb{E}(T \mid X_0 = k, X_T = 0)$, etc. It turns out that for the classical GR many of these quantities can be explicitly computed. This is usually done by employing martingale theory (see e.g. Williams [Wil91]), or, more elementarily, by solving certain recurrence relations (as in [Fel68]).

In the present chapter we seek to analyse the more general problem of gambler's ruin with delays (DGR). Given p and n as before, and a sequence of parameters $(\gamma_1, \ldots, \gamma_{n-1}) \in (0,1]^{n-1}$ we define a new Markov chain $(X_t)_{t\in\mathbb{N}_0}$ over $\{0, 1, \ldots, n\}$ with 0 and n still being absorbing states and the following transition probabilities for $0 < k < n$:

$$p_{k,k-1} = \mathbb{P}(X_{t+1} = k-1 \mid X_t = k) = p\gamma_k,$$
$$p_{k,k+1} = \mathbb{P}(X_{t+1} = k+1 \mid X_t = k) = (1-p)\gamma_k,$$
$$p_{k,k} = \mathbb{P}(X_{t+1} = k \mid X_t = k) = 1-\gamma_k.$$

This modifies the previous model by allowing a draw outcome of a game with probability $1-\gamma_k$ in which case our gambler's fortune is unchanged, and conditioned on winning or losing £1 the probabilities are the same as before. It may seem artificial as a game of chance to allow the probability of the draw

outcome to depend on the current fortune of the gambler. However, for our purposes, this is exactly what was needed, as the number of significant edges (and hence the probability to sample one) at any time depends only on the number of currently active vertices.

Note that setting $\gamma_0 = 0, \gamma_n = 0$ the formulae above extend to $0 \leq k \leq n$. In the special case $\gamma_1 = \cdots = \gamma_{n-1} = 1$ we recover the classical GR.

There is a vast amount of literature dealing with gambler's ruin and its extensions. This ranges from classical textbooks on probability such as Feller's [Fel68] to recent papers generalising the original problem in various directions. Engel [Eng93], Stirzaker [Sti94, Sti06], Bruss, Louchard and Turner [BLT03], and Swan and Bruss [SB06] all look at the problem of $N > 2$ gamblers playing each other at random and compute probabilities of each player being ruined and various other associated quantities depending on the initial wealth distribution. Some authors refer to this as the N-tower problem as the process can be visualised by N towers of stacked coins where at each step a coin is taken from the top of a tower chosen at random and placed on another tower amongst the others chosen again at random. The game stops when one of the towers becomes empty.

Other variations include two players (a casino and a gambler) with multiple currencies [KP02] by Kmet and Petkovšek. Lengyel in [Len09] and Gut in [Gut13] allow ties, and more generally Katriel in [Kat14, Kat13] studies absorption time for a game in which the pay-off is a random variable with range $[-\nu, +\infty) \cap \mathbb{Z}$ for a positive integer ν. Common to all these is that they assume identically distributed increments, whereas we allow that these depend on the given state.

In Gut's paper [Gut13], a particular instance of DGR when all the delays are the same is investigated. One can recover all of his results (with slightly different notation) by setting $\gamma_1 = \cdots = \gamma_{n-1} = 1 - r$. El-Shehawey [ES09] allows all the probabilities to win, lose or draw to depend on the player's current fortune. This is indeed a more general setting then ours but only absorption (i.e. ruin) probabilities are provided there and the expected waiting time until absorption is not considered. The procedure to compute these expectations was, however, described in Parzen's book [Par99, pp. 239 – 241] and we apply it to our setting below.

We shall now derive the formula for the expected time of absorption of DGR. As before, T is the time of absorption. To simplify notation, for each $0 \leq k \leq n$ we denote $\mathbb{E}(T \mid X_0 = k)$ by E_k. Note

that the ratio $\frac{p_{k,k-1}}{p_{k,k+1}} = \frac{p}{1-p} \in [0,1]$ is fixed and we denote it by λ. In order to calculate the expected time of absorption, we need to solve the following recurrence relation

$$\gamma_k E_k = 1 + \gamma_k(pE_{k-1} + (1-p)E_{k+1}), \tag{3.1}$$

for $0 < k < n$, with the boundary conditions $E_0 = 0$, and $E_n = 0$. Note that the associated homogeneous equation

$$E_k = pE_{k-1} + (1-p)E_{k+1} \tag{3.2}$$

whose solutions yield the probabilities for the chain to be absorbed in 0 or n, depending on which boundary conditions are imposed, is the same as in the case of classical GR. In other words, since the equation (3.2) above does not depend on the lagging parameters γ_k, the probability that the gambler bankrupts before earning £n is the same for both DGR and GR.

It is not hard to see that for any $a, b \in \mathbb{R}$ the expression $a + b\lambda^k$ solves the homogeneous equation above and finding the solution is therefore just a matter of fitting the constants a and b. In order to find all the solutions to (3.1) it therefore suffices to find just one particular solution to it. One way to solve this is by assuming a series expansion $\sum_i a_i \lambda^i$ of the solution. After a somewhat tedious computation which we deliberately skip, one can obtain the solution:

$$E_k = \frac{1+\lambda}{1-\lambda}\left(S_n \frac{1-\lambda^k}{1-\lambda^n} - \sum_{i=1}^{k-1} \frac{1}{\gamma_i}(1-\lambda^{k-i})\right), \text{ for } 0 \le k \le n, \tag{3.3}$$

where

$$S_n = \sum_{i=1}^{n-1} \frac{1}{\gamma_i}(1-\lambda^{n-i}).$$

The reader is invited to check that this indeed satisfies both the recurrence relation (3.1) and the boundary conditions. Setting $\gamma_1 = \cdots = \gamma_{n-1} = 1 - r$ gives

$$S_n = \frac{1}{1-r}\left(n - \frac{1-\lambda^n}{1-\lambda}\right),$$
$$E_k = \frac{1}{1-r} \cdot \frac{1+\lambda}{1-\lambda}\left(n\frac{1-\lambda^k}{1-\lambda^n} - k\right),$$

which coincides with the result in Gut's aforementioned paper [Gut13].

Note that plugging in the values $\gamma_i = \frac{2i(n-i)}{n(n-1)}$ into (3.3) will give the explicit formula for the expected time of reaching a consensus on the complete graph assuming we start with k supporters of the first (more persuasive if $p < 1/2$) and $n - k$ of the second (weaker) option.

Remark 3.4. Note that all the formulae have a removable singularity at 1 and hence are well defined by continuity at $\lambda = 1$ which corresponds to $p = 1/2$.

3.2.1 *Monotonicity of the mean absorption time*

We now wish to show that as p increases from 0 to 1/2 (or λ from 0 to 1) the mean absorption time monotonically increases as well. We could try to prove that each E_k is monotonic in p but this clearly is not true even in the case with no delays. One can easily compute that, for example, E_1 when $n = 3$ attains a global maximum at $\lambda = (-1 + \sqrt{3})/2$. For this reason we shall be considering symmetric sums $E_k + E_{n-k}$.

Unfortunately, these symmetric sums are also not in general monotonic. It turns out, however, that for a fixed $0 < k < n$ the symmetric term $E_k + E_{n-k}$ is indeed increasing with λ, as long as we assume that the parameters γ_i are symmetric, i.e. if $\gamma_i = \gamma_{n-i}$ for $0 < i < n$. Note that for the application we have in mind this suffices, as the starting distribution of strategies over the graph is usually chosen in a way that makes it symmetric under swapping the strategies, and also the probability of sampling a significant edge (which is interpreted as a delay parameter γ_i) only depends on the number of vertices currently playing one or the other strategy, and, as we noted before, is independent under swapping the two strategies.

We shall first give the proof of this fact for the classical GR which immediately extends to the case where all the parameters γ_i are the same. We shall need the following lemma.

Lemma 3.5. *Let $\alpha > 1$. The function*

$$f(\lambda) = \frac{1-\lambda}{1+\lambda} \cdot \frac{1+\lambda^\alpha}{1-\lambda^\alpha}$$

is a decreasing (non-negative) function of λ on $[0,1]$.

Proof. The function is continuously differentiable on $(0,1)$ with $f'(0) = -2$, hence it suffices to show that f' does not have zeros in $(0,1)$.

The zeros of f', if they were to exist, would have to satisfy the following equation

$$\alpha\lambda^{\alpha-1}(1-\lambda^2) + \lambda^{2\alpha} - 1 = 0,$$

or the equivalent one

$$g(\lambda) = \alpha\left(\frac{1}{\lambda} - \lambda\right) + \left(\lambda^\alpha - \frac{1}{\lambda^\alpha}\right) = 0.$$

Setting $\lambda = e^{-t}$ we get

$$g(\lambda) = h(t) = 2\alpha\left(\sinh(t) - \sinh(\alpha t)\right)$$

and it suffices to show that $h(t)$ does not have zeros in $t \in (0, +\infty)$. But this is clear since sinh is increasing on $(0, +\infty)$ and hence $t < \alpha t$ implies $\sinh(t) - \sinh(\alpha t) < 0$. □

Theorem 3.6. *In the classical GR setting, the symmetrised expectation*

$$E_k + E_{n-k} = n \cdot \frac{1+\lambda}{1-\lambda} \cdot \frac{(1-\lambda^k)(1-\lambda^{n-k})}{1-\lambda^n}$$

is a (non-negative) increasing function of λ on $[0,1]$ for each $0 \le k \le n$.

Proof. Note,

$$1 - \lambda^n = \frac{1}{2}(1-\lambda^k)(1+\lambda^{n-k}) + \frac{1}{2}(1+\lambda^k)(1-\lambda^{n-k}).$$

Hence,

$$\frac{1}{E_k + E_{n-k}} = \frac{1}{2n}\left(\frac{1-\lambda}{1+\lambda}\cdot\frac{1+\lambda^{n-k}}{1-\lambda^{n-k}} + \frac{1-\lambda}{1+\lambda}\cdot\frac{1+\lambda^{k}}{1-\lambda^{k}}\right)$$

and applying Lemma 3.5 twice yields the result. □

We would now like to prove the same result for the general symmetric DGR. The expression for the symmetric term is

$$E_k + E_{n-k} = \frac{1+\lambda}{1-\lambda}\Bigg[\sum_{i=1}^{n-1}\frac{1}{\gamma_i}(1-\lambda^{n-i})\frac{2-\lambda^k-\lambda^{n-k}}{1-\lambda^n} - \sum_{i=1}^{k-1}\frac{1}{\gamma_i}(1-\lambda^{k-i}) - \sum_{i=1}^{n-k-1}\frac{1}{\gamma_i}(1-\lambda^{n-k-i})\Bigg]. \quad (3.4)$$

For a fixed $0 < i < n$ letting $\gamma_i = \gamma_{n-i}$ tend to 0 whilst keeping the rest of the parameters bounded away from zero, the terms containing $\frac{1}{\gamma_i} = \frac{1}{\gamma_{n-i}}$ will become dominant which means that the expression above increases with λ if and only if each of those terms increases with $\lambda \in [0,1]$. It now remains to collect the like terms involving $\frac{1}{\gamma_i} = \frac{1}{\gamma_{n-i}}$, and to show that these are increasing with λ.

Let us fix k and i. We may assume that $0 < k, i \leq n/2$ as it is assumed that γ_is are invariant under changing i with $n-i$, and as the expression under consideration $E_k + E_{n-k}$ is also symmetric. The term multiplying $\frac{1}{\gamma_i} = \frac{1}{\gamma_{n-i}}$ in (3.4) is

$$\frac{1+\lambda}{1-\lambda}\left[(2-\lambda^{n-i}-\lambda^i)\frac{(2-\lambda^k-\lambda^{n-k})}{1-\lambda^n} - (1-\lambda^{k-i}) - (1-\lambda^{n-k-i})\right] \quad (3.5)$$

if $i < k$, and hence $n-i > n-k$; and

$$\frac{1+\lambda}{1-\lambda}\left[(2-\lambda^{n-i}-\lambda^i)\frac{(2-\lambda^k-\lambda^{n-k})}{1-\lambda^n} - (1-\lambda^{n-k-i}) - (1-\lambda^{i-k})\right] \quad (3.6)$$

if $i > k$, and hence $n-i < n-k$. If $k = i$ both expressions are valid. Notice that swapping i with k transforms one into another and therefore it suffices to prove that the expression in (3.5) is an increasing function of $\lambda \in [0,1]$ for fixed $0 < i \leq k \leq n/2$.

Remark 3.7. Note that in case n is even and $i = n/2$, we have $i = n - i$ and also $k \leq i$, so in order not to double-count, the expression we should be considering is not (3.6) but rather

$$\frac{1+\lambda}{1-\lambda}\left[(1-\lambda^{n/2})\frac{(2-\lambda^k-\lambda^{n-k})}{1-\lambda^n} - (1-\lambda^{n/2-k})\right]$$

which is exactly a half of (3.6). It therefore still suffices to show monotonicity of (3.6), or equivalently (3.5).

Let us denote by $G(\lambda)$ the expression inside the square brackets in (3.5):

$$G(\lambda) = (2-\lambda^{n-i}-\lambda^i)\frac{(2-\lambda^k-\lambda^{n-k})}{1-\lambda^n} - (1-\lambda^{k-i}) - (1-\lambda^{n-k-i}).$$

Then in order to show that

$$\lambda \mapsto \frac{1+\lambda}{1-\lambda}\cdot G(\lambda)$$

is increasing on $[0,1]$ it is sufficient to show that

$$H(\lambda) = \frac{1+\lambda^i}{1-\lambda^i}\cdot G(\lambda)$$

is increasing and non-negative on the same domain, as by virtue of Lemma 3.5 we know that

$$\lambda \mapsto \frac{1+\lambda}{1-\lambda}\cdot\frac{1-\lambda^i}{1+\lambda^i}$$

is non-negative and increasing, and so will be the product of the two. We calculate,

$$\begin{aligned}
G(\lambda) &= \frac{(2-\lambda^{n-i}-\lambda^i)(2-\lambda^k-\lambda^{n-k}) - (2-\lambda^{k-i}-\lambda^{n-k-i})(1-\lambda^n)}{1-\lambda^n} \\
&= \frac{2-2\lambda^k-2\lambda^{n-k}-2\lambda^{n-i}-2\lambda^i+\lambda^{k+i}+\lambda^{n-(k-i)}+\lambda^{k-i}+\lambda^{n-(k+i)}+2\lambda^n}{1-\lambda^n} \\
&= \frac{(1-\lambda^i)\left[2-2\lambda^{n-i}+\lambda^{k-i}+\lambda^{n-(k+i)}-\lambda^k-\lambda^{n-k}\right]}{1-\lambda^n} \\
&= \frac{(1-\lambda^i)\left[2(1-\lambda^{n-i})+(1-\lambda^i)(\lambda^{k-i}+\lambda^{n-(k+i)})\right]}{1-\lambda^n}
\end{aligned}$$

and hence

$$H(\lambda) = \frac{1+\lambda^i}{1-\lambda^i} \cdot G(\lambda) = \frac{(1+\lambda^i)\left[2(1-\lambda^{n-i}) + (1-\lambda^i)(\lambda^{k-i} + \lambda^{n-(k+i)})\right]}{1-\lambda^n}.$$

After introducing a substitution $\lambda = e^{-2t}$,

$$\begin{aligned} F(t) = \frac{1}{2}H(e^{-2t}) &= \frac{2\cosh(it)\left[\sinh((n-i)t) + \sinh(it)\cosh((n-2k)t)\right]}{\sinh(nt)} \\ &= \frac{2\cosh(it)\sinh((n-i)t) + \sinh(2it)\cosh((n-2k)t)}{\sinh(nt)} \end{aligned}$$

it suffices to show that F is non-negative and decreasing on $[0, +\infty)$. Using addition formulae we can rearrange the numerator of the previous expression to read

$$\begin{aligned} &[2\cosh(it)\sinh(nt)\cosh(it) - \sinh(nt)] + \sinh(nt) - 2\cosh(it)\sinh(it)\cosh(nt) \\ &\qquad + \sinh(2it)\cosh((n-2k)t) \\ &= \sinh(nt)\cosh(2it) + \sinh(nt) - \sinh(2it)\cosh(nt) + \sinh(2it)\cosh((n-2k)t) \\ &= \sinh((n-2i)t) + \sinh(2it)\cosh((n-2k)t) + \sinh(nt). \end{aligned}$$

Therefore

$$F(t) = 1 + \frac{\sinh((n-2i)t) + \sinh(2it)\cosh((n-2k)t)}{\sinh(nt)}.$$

To make things cleaner, we introduce yet another substitution

$$\begin{aligned} Q(t) = F(t/n) - 1 &= \frac{\sinh((1-\alpha)t) + \sinh(\alpha t)\cosh((1-\beta)t)}{\sinh(t)} \\ &= \frac{2\sinh((1-\alpha)t) + \sinh((\alpha+\beta-1)t) + \sinh((1-\beta+\alpha)t)}{2\sinh(t)} \end{aligned}$$

where $\alpha = \frac{2i}{n}$, $\beta = \frac{2k}{n}$, and since $0 < i \leq k \leq n/2$ we have $0 < \alpha \leq \beta \leq 1$. It is clear now from the formula that $F(t) \geq 1$ on $[0, \infty)$ and in particular F is non-negative on the positive reals. It therefore

remains to show that Q is decreasing on $[0, \infty)$, or equivalently (since the denominator of $Q'(t)$ is $4\sinh^2(t)$) that

$$4\sinh^2(t)Q'(t) \leq 0, \text{ for } t \in [0, \infty).$$

We calculate

$$\begin{aligned}
W(t) &= 4\sinh^2(t)Q'(t) \\
&= 2\sinh(t)\big[2(1-\alpha)\cosh((1-\alpha)t) + (\alpha+\beta-1)\cosh((\alpha+\beta-1)t) \\
&\quad + (1-\beta+\alpha)\cosh((1-\beta+\alpha)t)\big] - 2\cosh(t)\big[2\sinh((1-\alpha)t) \\
&\quad + \sinh((\alpha+\beta-1)t) + \sinh((1-\beta+\alpha)t)\big] \\
&= 2(1-\alpha)\big[\sinh((2-\alpha)t) + \sinh(\alpha t)\big] + (\alpha+\beta-1)\big[\sinh((\alpha+\beta)t) \\
&\quad + \sinh((2-\alpha-\beta)t)\big] + (1-\beta+\alpha)\big[\sinh((2-\beta+\alpha)t) + \sinh((\beta-\alpha)t)\big] \\
&\quad - 2\big[\sinh((2-\alpha)t) - \sinh((\alpha)t)\big] - \big[\sinh((\alpha+\beta)t) - \sinh((2-\alpha-\beta)t)\big] \\
&\quad - \big[\sinh((2-\beta+\alpha)t) - \sinh((\beta-\alpha)t)\big] \\
&= -2\alpha\sinh((2-\alpha)t) + 2(2-\alpha)\sinh(\alpha t) - (2-\alpha-\beta)\sinh((\alpha+\beta)t) \\
&\quad + (\alpha+\beta)\sinh((2-\alpha-\beta)t) - (\beta-\alpha)\sinh((2-\beta+\alpha)t) \\
&\quad + (2-\beta+\alpha)\sinh((\beta-\alpha)t)
\end{aligned}$$

This last expression for $W(t)$ clearly evaluates to zero at $t = 0$ and therefore it is enough to show that this is decreasing on $t \in [0, \infty)$, in other words it suffices to show $W'(t) \leq 0$ for $t \geq 0$. We calculate again,

$$\begin{aligned}
W'(t) &= 2(2-\alpha)\alpha\left[\cosh(\alpha t) - \cosh((2-\alpha)t)\right] \\
&\quad + (\alpha+\beta)(2-\alpha-\beta)\left[\cosh((2-\alpha-\beta)t) - \cosh((\alpha+\beta)t)\right] \\
&\quad + (2-\beta+\alpha)(\beta-\alpha)\left[\cosh((\beta-\alpha)t) - \cosh((2-\beta+\alpha)t)\right] \\
&= 4(2-\alpha)\alpha\sinh(t)\sinh((\alpha-1)t) \\
&\quad + 2(\alpha+\beta)(2-\alpha-\beta)\sinh(t)\sinh((1-\alpha-\beta)t) \\
&\quad + 2(2-\beta+\alpha)(\beta-\alpha)\sinh(t)\sinh((\beta-\alpha-1)t) \\
&= -\sinh(t)\left[4(2-\alpha)\alpha\sinh((1-\alpha)t) + 2(2-\beta+\alpha)(\beta-\alpha)\sinh((1-\beta+\alpha)t)\right. \\
&\quad \left. - 2(\alpha+\beta)(2-\alpha-\beta)\sinh((1-\alpha-\beta)t)\right]
\end{aligned}$$

In the case $\alpha + \beta > 1$ the claim easily follows as the minus sign in front of the third term can be used to change the argument of that sinh function to $(\alpha + \beta - 1)t$. Recalling that $0 < \alpha \leq \beta \leq 1$ it is easy to check that all the other constant factors appearing in the expression are non-negative.

In the case $\alpha + \beta \leq 1$, the claim follows from the facts that

$$\begin{gathered}
4(2-\alpha)\alpha + 2(2-\beta+\alpha)(\beta-\alpha) \geq 2(\alpha+\beta)(2-\alpha-\beta), \\
1-\alpha \geq 1-\alpha-\beta, \\
1-\beta+\alpha \geq 1-\alpha-\beta,
\end{gathered}$$

and the following lemma.

Lemma 3.8. *Let $a_1, a_2, a_3, b_1, b_2, b_3$ be non-negative real numbers such that $b_1 \geq b_3$, $b_2 \geq b_3$, and $a_1 + a_2 \geq a_3$. Then for all $t \geq 0$*

$$a_1\sinh(b_1 t) + a_2\sinh(b_2 t) - a_3\sinh(b_3 t) \geq 0.$$

Proof. We rewrite the left hand side as

$$a_1 [\sinh(b_1 t) - \sinh(b_3 t)] + a_2 [\sinh(b_2 t) - \sinh(b_3 t)] + (a_1 + a_2 - a_3) \sinh(b_3 t).$$

Since sinh is an increasing function, each of the terms above is non-negative. □

This completes the proof of the following theorem.

Theorem 3.9. *For each $0 \leq k \leq n$ the symmetric sum of the mean absorption times $\mathbb{E}(T \mid X_0 = k) + \mathbb{E}(T \mid X_0 = n - k)$ of gambler's ruin with symmetric delays is monotonically increasing with $p \in [0, 1/2]$.*

In particular, we have proved the following result.

Theorem 3.10. *For $m = 2$ and $G = K_n$ the complete graph with n vertices, and if the initial state is chosen symmetrically with respect to swapping strategies (e.g. uniformly at random), then the expected time until reaching consensus increases monotonically with $p \in [0, 1/2]$.*

3.3 A few words on upper bounds on the time to completion

Haslegrave also considered which graphs give the longest expected time to completion. In the special case when $p = 0$ and $m = 2$, he proves the following upper bound.

Theorem 3.11 (Haslegrave)**.** *For any connected graph G with n vertices, $\mathbb{E}(T) < n^2 \log n + n$.*

This bound is close to best possible, as he also shows that jellyfish graphs attain the bound up to a factor of $1 - o(1)$. Recall that the *jellyfish* graph consists of a clique (that is, a complete subgraph) with several shorter pendant paths attached as evenly as possible.

Theorem 3.12 (Haslegrave)**.** *Let J_n be the jellyfish graph consisting of a clique of size $n - 2n/\log_2 n$, with $n/(\log_2 n)^2$ pendant paths of length $2 \log_2 n$ each. Then the expected time of the process on J_n is $(1 - o(1))n^2 \log n$.*

The proofs of both of these results involve bounding the expected time to completion in a variant of the coupon collector's problem and are beyond the scope of this thesis.

Chapter 4

INDUCED DYNAMICS ON THE HYPERSPACE OF COMPACT SUBSETS

TUDYING the dynamics of a system (X, T) can be approached from many different angles. The most direct strategy, by studying individual orbits, can indeed be very fruitful, and most of the topological dynamics has built upon this premise. In reality, however, one rarely knows precisely the state of a physical system. The observables usually come with some uncertainty and it therefore makes sense to study probability distributions on X and how they evolve. As Bauer and Sigmund nicely put it in their paper [BS75] *from 1975:*

> *"The elements of $M(X)$ [the space of probability measures on X] can be viewed as statistical states, representing imperfect knowledge of the system. The elements of X are imbedded in $M(X)$ as the pure states."*

To a topologist, it seems natural to dismiss probabilities and study how the support of a probability distribution evolves over time, i.e. to study the induced dynamics on 2^X, the hyperspace of compact non-empty subsets of X. This transition from probability measures to their supports is in fact a coarse graining.

In this chapter we study the interplay between the original system (X, T) and its associated system $(2^X, 2^T)$. The focus is on periodic points and how knowing which periods appear in one system can help to find those appearing in the other. We give some necessary conditions on admissible sets of periods for these maps. Seemingly unrelated to this, we construct an almost totally minimal homeomorphism of the Cantor set. We also apply our theory to give a full description of admissible period sets for induced maps of the interval maps, and maps over the reals. The description of admissible periods is also given for maps induced on symmetric products.

The results in this chapter are joint with Leobardo Fernández and Chris Good — the paper [FGPR15] *containing this material is currently under review.*

Let us first expand on the motivational construction from the introduction. Given a compact metric space X and a continuous function $T\colon X \to X$, recall that

$$2^X = \{A \subseteq X \mid A \text{ is non-empty and compact}\}$$

denotes the *hyperspace of compact non-empty subsets of* X. The topology on 2^X is induced by the *Hausdorff distance*:

$$d_H(A, B) = \inf\{\varepsilon > 0 \mid A \subseteq N_\varepsilon(B) \text{ and } B \subseteq N_\varepsilon(A)\},$$

for any $A, B \in 2^X$, where $N_\varepsilon(A) = \{x \in X \mid (\exists y \in A)\, d(x, y) < \varepsilon\}$ denotes the ε-neighbourhood of the set A. It turns out that this topology on 2^X is compact and coincides with the abstractly defined *Vietoris' topology* given by the basis

$$\mathcal{B} = \{\langle U_1, U_2, \ldots, U_m\rangle \mid U_i \text{ is open for each } i \in \{1, 2, ..., m\}, m \in \mathbb{N}\},$$

where

$$\langle U_1, U_2, ..., U_m\rangle = \left\{A \in 2^X \mid A \subseteq \bigcup_{i=1}^{m} U_i \text{ and } A \cap U_i \neq \emptyset, i \in \{1, 2, ..., m\}\right\}.$$

For the proof of these and related results see [Mac05] and [Nad06, Theorem 0.11 and 0.13].

On this compact space one can define a natural induced map $2^T\colon 2^X \to 2^X$ given by:

$$2^T(A) = T(A) = \{T(a) \mid a \in A\}, \text{ for any } A \in 2^X,$$

which is continuous with respect to the Hausdorff distance. Intuitively, if T is taken to represent an update function of some system, then tracking the orbits of 2^T amounts to tracking a collection of points of the original system as they evolve over time.

From the measure theoretic perspective computing the probability of seeing a system in a certain state at a certain time is more important then trying to compute the exact state of the system given that the initial state is usually not accurately known and only probabilities can be assigned to regions of the state space from which the system was likely to have started. This point of view has lead to the rich study of the space of measures over X and the associated *push-forward operator* $\mu \mapsto T\#\mu$ where $T\#\mu$ is a measure defined by:

$$(T\#\mu)(S) = \mu(T^{-1}(S)), \text{ for any measurable } S \subseteq X.$$

If X is for example an interval $[0, 1]$, T a smooth piecewise monotone map with $T' \neq 0$ everywhere except possibly at a finite number of points, and if the measure μ has a density f with respect to the Lebesgue measure; by the change of variables formula one can see that the push-forward measure $T\#\mu$ also has a density g given by the formula:

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

which is precisely the defining relation of the *transfer operator*.

Dispensing with probabilities, one can ask how supports of these measures evolve over time. Recall that the support of a measure can intuitively be understood as a closed set of full measure that is minimal with respect to inclusion. The set, denoted by $\text{supp}(\mu)$, outside of which our system is unlikely to be found if μ represents the probability of finding our system in some state. It is immediate from the definition that $\text{supp}(T\#\mu) = 2^T(\text{supp}(\mu))$ hence the evolution of the support is precisely described by the hyperspace dynamics and the map $\text{supp} \colon \mathcal{M}(X) \to 2^X$ is a coarse graining.

Clearly one is interested in which properties of the dynamical system (X, T) transfer to the system $(2^X, 2^T)$ and there have been a number of results in this direction. For example, Banks [Ban05] proves that T is weakly mixing if and only if 2^T is weakly mixing which is further equivalent to 2^T being

transitive. Bauer and Sigmund in [BS75] show that if T has positive entropy, then the entropy of 2^T is infinite. Kwietniak and Oprocha go on to show [KO07] that if T does not have a dense set of recurrent points, then the entropy of 2^T is at least $\log 2$. Guirao, Kwietniak, Lampart, Oprocha, and Peris in [GKL+09] show that each of the usual notions of chaos except for the classical Devaney's definition is transferred from T to 2^T, whereas the converse does not hold for any of them. Fernández and Good in [FG16] show that 2^T has shadowing if and only if T does. Recently, Fernández, Good, and Ramírez jointly with the author in [FGPR15] have proved that 2^T is chain transitive if and only if T is chain weakly mixing. The results from this paper have not been included in this thesis.

Prior to our work, little has been done on the periodic points of 2^T other than some results on periodicity in symmetric products by Gómez, Illanes, and Méndez in [GRIM12] which we extend in Section 4.7. It turns out that the theory here is of some interest. Moreover, the techniques that seem to be necessary to construct hyperspaces with various periods are of independent interest.

The rest of the chapter is organised as follows. In Section 4.1 we briefly introduce the notation. Then in Section 4.2, Example 4.7 we show that the induced system $(2^X, 2^T)$ can contain periodic points of any period even when T has no periodic points at all. A complete description of the situation for interval maps is given in Theorem 4.8 where we show that if T is an interval map, then the set of periods of 2^T is either $\{1\}$ or $\{1, 2\}$ or $\mathbb{N}$. A similar result holds for the maps over $\mathbb{R}$, see Theorem 4.31, except that it is now possible that the map has no periods at all — as is the case e.g. for the map $x \mapsto x + 1$.

In Section 4.3 we introduce minimal building blocks of periodic sets in hyperspaces and derive some fundamental properties related to these. And in Section 4.4 we seek to characterise admissible periods for induced maps. Trivially, if x is a point of period n under T, then $\{x\}$ is a point of period n under 2^T. Since finite unions of points are again closed sets, it follows, for example, that if T has a period 4 point and a period 6 point, then 2^T has points of period 1, 2, 3, 4, 6, and 12. It is possible to give a complete description of admissible periods in 2^T that arise in this fashion, i.e. those that are formed exclusively of points that were periodic in the original system (X, T), and this is done in Section 4.7. This section also contains results on periods in symmetric products. More generally if 2^T has points of period 4 and 6 consisting of pairwise disjoint subsets of X, then it would again have

points of period 1, 2, 3, 4, 6, and 12. It is reasonable to ask then, whether this is a general property of induced maps on hyperspaces.

It turns out that this is not the case. In Section 4.6 we construct a dynamical system (Z, R) on the Cantor set such that the induced system $(2^Z, 2^R)$ only has periodic points of periods 1, 2, and 3. In particular it does not have a period 6 point. In Theorem 4.26 we generalise this construction to an induced system which admits only those periods that are divisors[1] of two positive integers p and q.

To obtain those results we first in Section 4.5 construct an *almost totally minimal system* (ATM) over the Cantor set which is of interest in its own right, see Theorem 4.18. Recall that a system (X, T) having a fixed point $x_0 \in X$ is almost totally minimal if after removing its (unique) fixed point the remaining non-compact system $(X^* = X \setminus \{x_0\}, T)$ is totally minimal, meaning that the full orbit with respect to any iterate of T of any point is dense in X^* (see Section 4.5 for a precise definition). This is done via *graph covers*, a tool first devised by Gambaudo and Martens in [GM06] to give a combinatorial description of minimal systems over the Cantor set. This theory has proved useful not just for describing the algebraic structure of such systems but also for constructing maps on the Cantor set with particular properties. Shimomura [Shi16b], for example, uses it to construct a transitive, completely scrambled 0-dimensional system. Essentially the same method but formalised in a slightly different way was used by Akin, Glasner, and Weiss in [AGW08] to give a generic self homeomorphism of the Cantor set. Bernardes and Darji extend some of their results in [BD12] and again make use of the same method. There is a close link between these graph covers, Bratteli-Vershik diagrams and Kakutani-Rokhlin towers and it would be possible to pass from one representation to another. As an example, in Section 4.9 we give a Bratteli-Vershik representation of that almost totally minimal system.

Section 4.7 contains results on periods in symmetric products, and towards the end in Section 4.10 we list some problems naturally arising from our considerations that are still without a satisfying resolution.

[1] Here and elsewhere in text by divisors we actually mean *positive* divisors.

All the spaces under consideration will be compact and metric unless specified otherwise. Given a continuous self-map $T\colon X \to X$ of such a space the *inverse limit* $\varprojlim (X, T)$ is defined as the set of full orbits of this system

$$\varprojlim (X, T) = \left\{ \underline{x} = (x_i)_{i\in\mathbb{Z}} = (\dots, x_{-1}, x_0, x_1, x_2, \dots) \in \prod_{i\in\mathbb{Z}} X \mid T(x_i) = x_{i+1} \right\}.$$

Note the slightly unconventional enumeration of indices above. We shall say that any $\underline{x} \in \varprojlim (X, T)$ with $x_0 = x$ is a *full orbit* of the point $x \in X$. There exists a natural homeomorphism of $\varprojlim (X, T)$ called (right) shift $\sigma\colon \varprojlim (X, T) \to \varprojlim (X, T)$ given by $(\sigma(\underline{x}))_i = x_{i+1}$, for all $i \in \mathbb{Z}$. For more on inverse limits see [IM12].

Recall that to a full orbit $\underline{x} \in \varprojlim (X, T)$ one can associate two limit sets, ω- and α-limit set which are the accumulation sets of the forward and backward orbit of x_0 respectively.

$$\omega(\underline{x}) = \bigcap_{m=0}^{\infty} \overline{\bigcup_{n=m}^{\infty} \{x_n\}},$$
$$\alpha(\underline{x}) = \bigcap_{m=0}^{\infty} \overline{\bigcup_{n=m}^{\infty} \{x_{-n}\}}.$$

Both of these are closed, non-empty, and strongly T-invariant meaning that $T(\omega(\underline{x})) = \omega(\underline{x})$ and $T(\alpha(\underline{x})) = \alpha(\underline{x})$, see e.g. [BC92, BGL13]. Note that we could equivalently say that they are fixed points of the induced map 2^T. As a shorthand we denote their union $\omega(\underline{x}) \cup \alpha(\underline{x})$ by $\lambda(\underline{x})$ which is again a closed and strongly T-invariant set

$$\lambda(\underline{x}) = \alpha(\underline{x}) \cup \omega(\underline{x}) = \bigcap_{m=0}^{\infty} \overline{\bigcup_{n=m}^{\infty} \{x_{-n}, x_n\}}.$$

Remark 4.1. Note that the definition of the ω-limit set involves only the forward orbit and hence does not depend on the choice of the full orbit. Thus it makes sense to define $\omega(x)$ for any point $x \in X$.

As we shall constantly be dealing with the sets of periods of different functions it is convenient to introduce a symbol $\mathrm{Per}(T)$ for the subset of natural numbers such that $k \in \mathrm{Per}(T)$ if and only if there exists a point $x \in X$ with the fundamental period k. $\mathrm{Per}(2^T)$ will consequentially be the set of all the fundamental periods of points in 2^X.

We shall also be interested in restrictions of the map 2^T to a few 2^T-invariant subsets of 2^X. We introduce a special symbol for each of these restrictions:

$$T_n = 2^T|_{F_n(X)},$$
$$T^{<\omega} = 2^T|_{F(X)},$$

where $F_n(X) = \{A \in 2^X \mid A \text{ has at most } n \text{ points}\}$ is the *n-fold symmetric product of* X and $F(X) = \bigcup_{n=1}^{\infty} F_n(X)$ is the collection of all finite subsets of X. Occasionally we shall write just T for any of the above maps (including 2^T itself) as this does not lead to any confusion, and is useful to keep the notation simple, especially when we need to refer to the n^{th} iterate of the map 2^T which we simply denote by T^n.

The usual *n-fold Cartesian product* will also be of interest as the n-fold symmetric product can be seen as a quotient of this space. Somewhat unconventionally in this chapter we denote the product space $\underbrace{X \times X \times \cdots \times X}_{n-times}$ by $X^{(n)}$ and the induced map by $T^{(n)}$. This is not to be confused with T^n which as mentioned above is simply the n^{th} iterate of T.

Let us define a few more notions we use below.

Definition 4.2 (Pseudo-orbit)**.** A sequence $\langle x_0, x_1, x_2, \dots \rangle$ is said to be a *δ-pseudo-orbit* for some $\delta > 0$ provided that $d(f(x_i), x_{i+1}) < \delta$ for each $i \in \mathbb{N}_0$. A *finite δ-pseudo-orbit of length $l \geq 1$* is a finite sequence $\langle x_0, x_1, x_2, \dots, x_l \rangle$ satisfying $d(f(x_i), x_{i+1}) < \delta$ for $0 \leq i < l$. We also say that it is a *δ-pseudo-orbit between x_0 and x_l*.

Definition 4.3 (Recurrence)**.** A point $x \in X$ is said to be *recurrent* in (X, T) if $x \in \omega(x)$. A point that is not recurrent is said to be *non-recurrent*.

For notational convenience, we now introduce a stronger notion than non-recurrence which we call *full non-recurrence*. A point $x \in X$ is said to be *fully non-recurrent* if there exists a full orbit

$\underline{x} \in \varprojlim (X, T)$ of $x = x_0$ such that $x \notin \lambda(\underline{x})$. Note that it is, thus, possible for a point to be neither recurrent nor fully non-recurrent. The motivation for introducing this notion is given in Remark 4.11.

Definition 4.4 (Chain recurrence)**.** A point $x \in X$ is said to be *chain recurrent* in (X, T) provided that for any $\varepsilon > 0$ there exists an ε-pseudo-orbit $\langle x = x_0, x_1, \dots, x_l = x \rangle$ of positive length $l \geq 1$.

Definition 4.5 (Turbulence)**.** The interval map $T\colon I \to I$ is said to be *turbulent* if there exist compact subintervals J, K with at most one common point such that $J \cup K \subseteq T(J) \cap T(K)$.

Definition 4.6 (0-dimensional space)**.** A compact metric space is 0*-dimensional* if it has a base consisting of clopen sets, or equivalently, if it is totally disconnected.

4.2 TWO SIMPLE RESULTS

To put our results into perspective, we start with two related results.

Example 4.7. Let $C = \mathbb{S}^1 \times [0, 1]$ be a cylinder where $\mathbb{S}^1 = [0, 1]/_\sim$ denotes the unit circle obtained from the interval $[0, 1]$ with its endpoints identified. Let $T\colon C \to C$ be an irrational rotation by $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ about the central axis combined with an upward displacement that preserves the bases, e.g. $(\varphi, z) \overset{T}{\mapsto} (\varphi + \alpha \bmod 1, 2z - z^2)$. Note that this is a homeomorphism of C. Let $X \subset C$ be the set consisting of the two bases $\mathbb{S}^1 \times \{0, 1\}$ and a full orbit $\{\dots, z_{-1}, z_0, z_1, z_2, \dots\}$ of a point z_0 on a generating line, say $z_0 = (0, 1/2)$, where we set $z_k = T^k(z_0)$ for all $k \in \mathbb{Z} \setminus \{0\}$ (see Figure 4.1). Then clearly $T|_X$ has no periodic points and at the same time, for any $k \in \mathbb{N}$ the set $\mathbb{S}^1 \times \{0, 1\} \cup \{\dots, z_{-2k}, z_{-k}, z_0, z_k, z_{2k}, \dots\}$ is periodic under 2^T with period k. ◂

Block and Coven [BC86] prove that if T is a continuous map from a compact interval to itself and every point is chain recurrent, then either T^2 is the identity map or T is turbulent. From this, one can easily deduce the following.

Theorem 4.8. *Let T be a continuous map of a compact interval to itself. Then* $\mathrm{Per}(2^T)$ *is either* $\{1\}$ *or* $\{1, 2\}$ *or* $\mathbb{N}$.

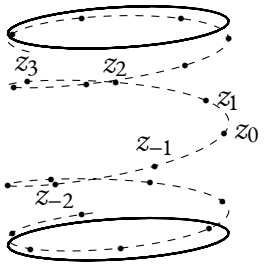

Figure 4.1: The set X from Example 4.7.

Proof. If T^2 is the identity, then $\mathrm{Per}(2^T)$ is either $\{1\}$ or $\{1, 2\}$. If T^2 turbulent, then T must have a periodic point with period which is not a power of 2 (see p33 in [BC92]). But if T has a point of period $2^n(2k+1)$, then by Šarkovs′kiĭ's Theorem, it has points of period $2^{n+1}m$ for any positive integer m, from which it follows that 2^T has points of period m, i.e. that $\mathrm{Per}(2^T) = \mathbb{N}$.

So suppose that there exists a point x_0 that is not chain recurrent.[2] Clearly, x_0 is not periodic. Also, we may assume that T is onto, as otherwise one can take a restriction to the surjective core of the map T and repeat the same reasoning. We can therefore take a full orbit $\underline{x} \in \varprojlim (X, T)$ of x_0 and consider the limit set

$$\lambda(\underline{x}) = \alpha(\underline{x}) \cup \omega(\underline{x}) = \bigcap_{m=0}^{\infty} \overline{\bigcup_{n=m}^{\infty} \{x_{-n}, x_n\}}.$$

We claim $x_0 \notin \lambda(\underline{x})$, i.e. that x_0 is fully non-recurrent. Firstly note that $x_0 \notin \omega(x_0)$ as otherwise it would be recurrent and hence chain recurrent. The other possibility is that there exists an increasing subsequence $(p_k)_{k\in\mathbb{N}}$ such that $x_{-p_k} \to x_0$ as $k \to \infty$. But then $x_{-p_k+1} \to x_1$ and for any $\varepsilon > 0$ one can choose $k_0 \in \mathbb{N}$ large enough so that $x_{-p_{k_0}+1}$ is ε close to x_1. Thus, $\langle x_0, x_{-p_{k_0}+1}, x_{-p_{k_0}+2}, \ldots, x_0\rangle$ is an ε-pseudo-orbit from x_0 to x_0 making x_0 a chain recurrent point. A contradiction.

As x_0 is fully non-recurrent, by Proposition 4.10, we immediately obtain all periods in $\mathrm{Per}(2^T)$ proving our corollary. □

[2] Compare this with Proposition 4.10. Also note that whether a point is *(chain) recurrent* depends only on its forward orbit, as is common in the literature; but, for convenience, our notion of *fully non-recurrent* point involves the full λ-limit set of the point. As a results, it is possible for a point to be neither recurrent nor fully non-recurrent.

We shall first describe the most basic type of periodic points that appear in the induced dynamics on 2^X. It captures both the periods arising from cycles in (X, T) via $x \mapsto \{x\}$ embedding as well as the periods as in Example 4.7.

Given a point $\underline{x} = (\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots) \in \varprojlim (X, T)$ we define the *set of periods of* $\underline{x}$ by

$$\mathrm{Per}(\underline{x}) = \{k \in \mathbb{N} \mid \overline{\{x_{mk} \mid m \in \mathbb{Z}\}} \in 2^X \text{ is periodic with period } k\}.$$

We use the same symbol as before but the meaning will be clear from the context. Note that this set does not depend on the choice of the starting point, i.e. $\mathrm{Per}(\underline{x}) = \mathrm{Per}(\sigma(\underline{x})) = \mathrm{Per}(\sigma^{-1}(\underline{x}))$. It may however depend on the chosen backward orbit of x_0. For example, if x_0 is a fixed point which also has a history of infinitely many isolated points then $\mathrm{Per}((\dots, x_{-2}, x_{-1}, x_0, x_0, \dots)) = \mathbb{N}$ but $\mathrm{Per}((\dots, x_0, x_0, x_0, \dots)) = \{1\}$.

Remark 4.9. If $\{x_0, x_1, \dots, x_{p-1}\}$ is a p-cycle in (X, T) then it is not hard to check that

$$\mathrm{Per}((\dots, x_{p-1}, x_0, \dots, x_{p-1}, x_0, \dots))$$

is the set of all divisors of p.

The situation akin to that in Example 4.7 occurs whenever there exists a *fully non-recurrent point*, i.e. a point whose full orbit does not accumulate at the point itself. Formally $x \in X$ is *fully non-recurrent* if there exists a full orbit $\underline{x} \in \varprojlim (X, T)$ of $x = x_0$ such that $x \notin \lambda(\underline{x})$.

Proposition 4.10. *Let $T\colon X \to X$ be a continuous map and assume that there exists a fully non-recurrent point x_0 with a full orbit $\underline{x} = (\dots, x_{-1}, x_0, x_1, \dots) \in \varprojlim (X, T)$. Then $\mathrm{Per}(\underline{x}) = \mathbb{N}$ and therefore also $\mathrm{Per}(2^T) = \mathbb{N}$.*

Proof. Let $k \in \mathbb{N}$ be arbitrary and set $S = \overline{\{x_{mk} \mid m \in \mathbb{Z}\}} \in 2^X$. S is clearly mapped to itself under k iterates of 2^T. It is therefore periodic with a fundamental period that divides k. To show that this is in fact k it suffices to see that $x_0 \notin (2^T)^n(S)$ for any $0 < n < k$.

To that end take any such n and note that $(2^T)^n(S) = \overline{\{x_{mk+n} \mid m \in \mathbb{Z}\}}$. Recall that $x_0 \notin \lambda(\underline{x})$ and so if x_0 was in $(2^T)^n(S)$ it could only be equal to x_{m_0k+n} for some $m_0 \in \mathbb{Z}$ and thus x_0 would have to be periodic. This would then imply that $x_0 \in \omega(\underline{x}) \subseteq \lambda(\underline{x})$ which gives a contradiction. □

Remark 4.11. It is worth noting that this result does not hold if x_0 above is only required to be non-recurrent. Indeed, the system (X, T) we construct in Theorem 4.18 below has points whose forward orbits in the limit approach the unique fixed point and are, thus, non-recurrent in the traditional sense. Their backward orbit, however, must be densely filling the space as we know that the full orbit of each point, save the fixed one, is dense. Those points are thus non-recurrent but are not fully non-recurrent and one can easily check that $\mathrm{Per}(2^T) = \{1\}$.

Remark 4.12. The result holds even if x itself is recurrent but has a point x_{-k} for some $k > 0$ in the backward part of its orbit that is fully non-recurrent. This is because $\mathrm{Per}(\underline{x}) = \mathrm{Per}(\sigma^{-k}(\underline{x})) = \mathbb{N}$.

In light of this remark it is useful to extend the notion of non-recurrence to full orbits. We say that $\underline{x} = (\dots, x_{-1}, x_0, x_1, \dots) \in \varprojlim (X, T)$ is *fully non-recurrent* if $x_k \notin \lambda(\underline{x})$ for some $k \in \mathbb{Z}$.

We have already used the obvious fact that $\mathrm{Per}(\underline{x}) \subseteq \mathrm{Per}(2^T)$ and one might suspect that by taking the union $\bigcup_{\underline{x} \in \varprojlim (X,T)} \mathrm{Per}(\underline{x})$ one could obtain the whole of $\mathrm{Per}(2^T)$. But this is not the case as is easy to see by looking at a system that consists of five points, two of which form a 2-cycle and the other three a 3-cycle. Then clearly $6 \in \mathrm{Per}(2^T)$ but no point in the inverse limit has 6 in its set of periods.

$\mathrm{Per}(\underline{x})$, however, has a nice structure. Below we show that it is closed under taking divisors and least common multiples. This in particular implies that in case it is finite, $\mathrm{Per}(\underline{x})$ is simply the set of divisors of its largest element. But what if it is infinite, does it have to be $\mathbb{N}$? A negative answer to this provides the odometer, a classical example of a transitive system over the Cantor set.

Example 4.13. Let $X = \Sigma_2 = \{0, 1\}^{\mathbb{N}}$. The 2-adic odometer $T \colon X \to X$ is defined recursively by

$$T(\xi_0, \xi_1, \dots) = \begin{cases} (1, \xi_1, \dots), & \text{if } \xi_0 = 0, \\ (0, f(\xi_1, \xi_2, \dots)), & \text{otherwise.} \end{cases}$$

It is not hard to see that $\mathrm{Per}(\underline{x}) = \{1, 2, 2^2, 2^3, \dots\}$ is the set of all powers of 2 for any $\underline{x} \in \varprojlim (X, T)$, and that $\mathrm{Per}(2^T)$ is the same set.

We note in passing that this is also an example of a system where 2^T-periodic points are dense in 2^X. This is simply because any closed set in Σ_2 can be approximated by a finite union of clopen cylinders to an arbitrary precision and these cylinders are clearly periodic under 2^T and so is their finite union. We remind the reader that a cylinder in Σ_2 is a set of the form

$$\{(\xi_0, \xi_1, \dots) \in \Sigma_2 \mid \xi_i = a_i \text{ for } 0 \leq i \leq n\},$$

for some $n \in \mathbb{N}$ and a choice of $a_i \in \{0, 1\}$, and that all such sets form a clopen basis of the Cantor topology on Σ_2. ◁

To prove the aforementioned structural result for elementary sets of periods, we shall need two lemmata.

Lemma 4.14. *If $S \in 2^X$ is a k-periodic then none of its $k-1$ iterates under 2^T can be a subset of S.*

Proof. Otherwise, let us assume that $(2^T)^j(S) \subseteq S$ for some $0 < j < k$. Using the usual notation one would write $T^j(S) \subseteq S$ and from there

$$S = T^{kj}(S) \subseteq T^{(k-1)j}(S) \subseteq \cdots \subseteq T^j(S) \subseteq S.$$

Thus $T^j(S) = S$ and S periodic with a period strictly less than k, a contradiction. □

Lemma 4.15. *Let $\underline{x} \in \varprojlim (X, T)$. Then*

$$\mathrm{Per}(\underline{x}) = \{\ k \in \mathbb{N} \mid \exists N_0 \in \mathbb{N} \quad s.t. \quad x_{-N_0 k} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } k \nmid l\}}\ \}. \tag{4.1}$$

Also

$$\mathrm{Per}(\underline{x}) = \{\ k \in \mathbb{N} \mid \exists N_0 \in \mathbb{N} \quad s.t. \quad \forall N \geq N_0 \quad x_{-Nk} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } k \nmid l\}}\ \} \tag{4.2}$$

and in particular

$$\mathrm{Per}(\underline{x}) = \bigcup_{N=0}^{\infty} \{\ k \in \mathbb{N} \mid x_{-Nk} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } k \nmid l\}}\ \}. \tag{4.3}$$

Proof. (4.2) follows easily from (4.1) if one recalls that $x_{-Nk} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } k \not| l\}}$ implies $x_{-(N+1)k} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } k \not| l\}}$. And from there it is clear that the sets under the union sign in (4.3) form a monotonically increasing family of sets converging to the set in (4.1).

We now turn to proving that the first claim holds. We first show that any k satisfying the defining statement of the set in (4.1) is necessarily in $\mathrm{Per}(\underline{x})$. Let such a $k \in \mathbb{N}$ be fixed and let N_0 be as in the definition of the set. It is clear that $x_{-N_0 k}$ is a "distinguishing feature", an element contained in $\overline{\{x_{mk} \mid m \in \mathbb{Z}\}}$ but which cannot be in any of its $k-1$ forward iterates under 2^T. The set $\overline{\{x_{mk} \mid m \in \mathbb{Z}\}}$ is thus truly a k-period point.

Conversely, take a $k \in \mathbb{N}$ which does not satisfy the statement in (4.1). Hence, there are infinitely many points in the k-step backward orbit of x_0 that are also in $\overline{\{x_l \mid l \in \mathbb{Z} \text{ and } k \not| l\}}$. In particular, infinitely many of them are in $\overline{\{x_{km+j} \mid m \in \mathbb{Z}\}}$ for some $0 < j < k$. Similar reasoning as in the first paragraph of this proof allows us to conclude that, not just infinitely many, but all of the elements of the k-step orbit $\{x_{mk} \mid m \in \mathbb{Z}\}$ are contained in $\overline{\{x_{km+j} \mid m \in \mathbb{Z}\}}$ and hence their closure as well. But this implies that $\overline{\{x_{km} \mid m \in \mathbb{Z}\}}$ is not a k-periodic point by Lemma 4.14. □

Proposition 4.16. *Let* $\underline{x} \in \varprojlim (X, T)$. *Then* $\mathrm{Per}(\underline{x})$ *is non-empty (always contains at least* 1*) and closed under taking divisors and least common multiples.*

Proof. Clearly $1 \in \mathrm{Per}(\underline{x})$. The rest is proved by invoking Lemma 4.15. Let $k \in \mathrm{Per}(\underline{x})$, $d|k$ and let N_0 be chosen as in (4.1), i.e. $x_{-N_0 k} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } k \not| l\}}$ and hence $x_{-(N_0 k/d)d} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } k \not| l\}} \supseteq \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } d \not| l\}}$. Therefore $d \in \mathrm{Per}(\underline{x})$.

As we have already seen that this set is closed under taking divisors it will suffice to show that for any two co-prime $m, n \in \mathrm{Per}(\underline{x})$ their product mn is in there as well. Chose N_0 to be the greater of the two integers associated to m and n in the context of (4.1). Then $x_{-(N_0 m)n} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } n \not| l\}}$ and $x_{-(N_0 n)m} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } m \not| l\}}$. As m and n are co-prime we have that $nm \not| l$ if and only if $n \not| l$ or $m \not| l$, hence $x_{-N_0(mn)} \notin \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } n \not| l\}} \cup \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } m \not| l\}} = \overline{\{x_l \mid l \in \mathbb{Z} \text{ and } mn \not| l\}}$. Therefore $mn \in \mathrm{Per}(\underline{x})$. □

A few facts about $\mathrm{Per}(2^T)$ are obvious. By considering singleton sets we see that $\mathrm{Per}(2^T) \supseteq \mathrm{Per}(T)$ and also $1 \in \mathrm{Per}(2^T)$ as $X \in 2^X$ is a fixed point. In fact $\mathrm{Per}(T^{<\omega}) = [\mathcal{D}(\mathrm{Per}(T))]$[3] $\subseteq \mathrm{Per}(2^T)$ since $F(X) \subseteq 2^X$. It turns out that $\mathrm{Per}(2^T)$ itself is closed under taking prime power divisors. Whether it must be closed under taking any divisor remains an open question. Surprisingly it needs not to be closed under taking least common multiples. The construction of such an example occupies Section 4.6. Intuitively one might think that if $A_0, A_1, \dots, A_{n-1}$ is an n-cycle and $B_0, B_1, \dots, B_{m-1}$ is an m-cycle in 2^X, and $d|n$ then $A_0 \cup A_d \cup \cdots \cup A_{n-d}$ should be a d-periodic point and $A_0 \cup B_0$ should be a $[m, n]$-periodic point, where $[m, n]$ denotes the least common multiple of m and n. But this does not hold in general as it could happen that their fundamental period is smaller than expected. As a trivial example consider a map over $X = \{0, 1, 2, 3\}$ given by $T(x) = x + 1 \bmod 4$. Then $A_0 = \{0, 1, 2\} \in 2^X$ is 2^T-periodic with period 4 but $A_0 \cup T^2(A_0) = X$ is a fixed rather than a period 2 point in 2^X.

Theorem 4.17. *Given a continuous map $T\colon X \to X$, the set of periods $\mathrm{Per}(2^T)$ of the induced map on 2^X contains $\mathrm{Per}(T) \cup \{1\}$ and is closed under taking prime power divisors.*

Proof. We have already seen that $\mathrm{Per}(T) \cup \{1\} \subseteq \mathrm{Per}(2^T)$. Given an $n \in \mathrm{Per}(2^T)$ and its prime factorisation $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$ where $r \geq 1$, for the second part of the claim it will suffice to find a full orbit $\underline{x} \in \varprojlim (X, T)$ for which $\mathrm{Per}(\underline{x})$ contains $p_1^{\alpha_1}$ as the result will then follow from Proposition 4.16.

To that end we set $k = p_1^{\alpha_1}$ and $l = n/p_1 = p_1^{\alpha_1 - 1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$, and let $A_0, \dots, A_{n-1}$ be a periodic orbit for 2^T of period n. By Lemma 4.14 there exists $x_0 \in A_0 \setminus A_l$, and as A_is are mapped surjectively onto each other, it is possible to find a full orbit $\underline{x} = (\dots, x_{-1}, x_0, x_1, x_2, \dots)$ of x_0 such that $x_{mn+i} \in A_i$ for all $m \in \mathbb{Z}$. We claim that $k \in \mathrm{Per}(\underline{x})$, i.e. that $\overline{\{x_{mk} \mid m \in \mathbb{Z}\}}$ is k-periodic under 2^T.

Firstly note that $x_{-mn+i} \in A_i \setminus A_{i+l \bmod n}$ for $m \in \mathbb{N}$, as otherwise if $x_{-mn+i} \in A_{i+l \bmod n}$ then after mapping it forward by T^{mn-i} we would have a contradiction with $x_0 \in A_0 \setminus A_l$. Similar reasoning allows us to conclude that, as we go backwards along the orbit, the pre-images $x_0, x_{-n}, x_{-2n}, x_{-3n}, \dots$ belong to, possibly, more and more complements of different A_is and, as there are only finitely many of those, this number must stabilise. The backward iterate at which this happens is then taken to be x_0

[3] This notation is defined in Remark 4.29 below.

and all the other indices are shifted accordingly. Note that this does not affect the claim we wish to prove as $\mathrm{Per}(\underline{x}) = \mathrm{Per}(\sigma^t(\underline{x}))$ for any $t \in \mathbb{Z}$. Also note that for this new x_0 we still have $x_0 \in \mathcal{A}_0 \setminus \mathcal{A}_l$. This modification will however allow us that from $x_0 \in \mathcal{A}_j$ for some j we infer $x_{mn+i} \in \mathcal{A}_{i+j \bmod n}$ for all $m \in \mathbb{Z}$ where before we could conclude this only for positive ms. Essentially the same trick was used previously in the proof of Lemma 4.15.

After we altered the enumeration in $\underline{x}$ we are ready to conclude the proof by showing that (now modified) $x_0 \notin \overline{\{x_t \mid t \in \mathbb{Z} \text{ and } k \nmid t\}}$. Note that

$$\overline{\{x_t \mid t \in \mathbb{Z} \text{ and } k \nmid t\}} = \bigcup_{i=1}^{k-1} \overline{\{x_{mk+i} \mid m \in \mathbb{Z}\}}$$

and for the sake of getting a contradiction we assume that $x_0 \in \overline{\{x_{mk+i} \mid m \in \mathbb{Z}\}}$ for some $0 < i < k$. This means that either $x_0 = x_{mk+i}$ for some $m \in \mathbb{Z}$ and therefore $x_0 \in \mathcal{A}_{\tilde{i}}$ where

$$\tilde{i} \equiv mk + i \pmod{n}, \qquad \text{hence} \qquad \tilde{i} \equiv i \pmod{k},$$

or x_0 is a limit of such points (x_{mk+i}s) out of which infinitely many must fall within the same congruence class with respect to n, say $\tilde{i}$, and thus also in the same $\mathcal{A}_{\tilde{i}}$ for some $0 \leq \tilde{i} < n$, which must also satisfy $\tilde{i} \equiv i \pmod{k}$. As $\mathcal{A}_{\tilde{i}}$ is closed, in both cases we get $x_0 \in \mathcal{A}_0 \cap \mathcal{A}_{\tilde{i}}$.

From here we can conclude that in the periodic case $x_0 \in \mathcal{A}_{\tilde{i}} \cap \mathcal{A}_{2\tilde{i} \bmod n}$ as $x_0 = x_{mk+i} = x_{2mk+2i}$. Also, if x_0 is a limit point of some subsequence in $\{x_{mn+\tilde{i}} \mid m \in \mathbb{Z}\} \subseteq \mathcal{A}_{\tilde{i}}$, using the reasoning described above and in light of the fact that $x_0 \in \mathcal{A}_{\tilde{i}}$, this subsequence is also in $\mathcal{A}_{2\tilde{i} \bmod n}$ and so must be x_0 as its limit. In both cases we henceforth conclude $x_0 \in \mathcal{A}_0 \cap \mathcal{A}_{\tilde{i}} \cap \mathcal{A}_{2\tilde{i} \bmod n}$. Continuing in this fashion we see that $x_0 \in \mathcal{A}_j$ for any j that satisfies

$$j \equiv s\tilde{i} \pmod{n} \text{ for some } s \in \mathbb{N}_0.$$

To obtain a contradiction with $x_0 \notin \mathcal{A}_l$ it now only remains to show that there exist $s \in \mathbb{N}_0$ for which $s\tilde{i} \equiv l \pmod{n}$. It is an elementary fact from number theory that there exists $\tilde{s} \in \mathbb{N}_0$ such that $\tilde{s}\tilde{i} \equiv (\tilde{i}, n) \pmod{n}$, where (a, b) stands for the greatest common divisor of integers a and b. Note

that $\tilde{i} \not\equiv 0 \pmod{k}$ and hence $p_1^{\alpha_1} = k \nmid \tilde{i}$. Thus $(\tilde{i}, n) = (\tilde{i}, n/p_1) = (\tilde{i}, l) \mid l$ and setting $s = \frac{l}{(\tilde{i},n)}\tilde{s}$ gives the desired conclusion. □

We now turn to showing that $\mathrm{Per}(2^T)$ is not always closed under taking least common multiples. But for this we need first to revisit the notion of *almost minimal systems*.

4.5 Almost totally minimal Cantor system

Recall that any two non-compact, locally compact, totally disconnected, separable, metrizable spaces with no isolated points are homeomorphic (see e.g. [Dano1, Proposition 1.1]). These are essentially equal to the Cantor set without a point, which is in turn homeomorphic to a countable union of Cantor sets. Let us denote such a set by X^*. In [Dano1] Danilenko gives a direct limit construction of a class of minimal systems on X^*. The system (X^*, T) is said to be *minimal* if the only non-empty, closed, strongly T-invariant subset of X^* is X^* itself, or equivalently if any full orbit[4] of any point is dense in X^*. In the compact case this is equivalent to asking that the ω-limit set of any point is whole of the state space, but here one has to be careful with the definition of minimality as some points can have their ω-limit sets empty due to non-compactness even if their full orbits are dense. In fact Danilenko proved that for any invertible minimal map on X^* the set of points with their forward orbit dense (those points x for which $\omega(x) = X^*$) will be a dense, G_δ set with empty interior (see [Dano1, Theorem 1.2]).

The class of maps constructed there conveniently extends to a class of continuous homeomorphisms of the Cantor set X which is obtained as a one point compactification of X^* with point at infinity being mapped onto itself. Such maps, with one fixed point and all the other points having dense full orbits are called *almost minimal systems* and in particular they are examples of *essentially minimal systems* which are defined as those systems which have one unique minimal subsystem (see [HPS92]). To make our proofs in Section 4.6 work we need a map that is *almost totally minimal* (ATM) meaning that the system (X, T^m) is almost minimal for any iterate T^m of the original map T where $m \in \mathbb{N}$, in

[4] If T is non-invertible then there will be points with more than one full orbit.

other words, that the system $(X^*, T|_{X^*})$ with the fixed point removed is totally minimal in the usual sense.

Theorem 4.18. *There exists an almost totally minimal homeomorphism $T : X \to X$ of the Cantor set X. This T has exactly one fixed point $x_0 \in X$ and the full T^m-orbit of every other point is dense in X for every $m \in \mathbb{N}$ i.e.*

(i) $(\exists!\, x_0 \in X) \quad T(x_0) = x_0$,

(ii) $(\forall y \in X \setminus \{x_0\})\ (\forall m \in \mathbb{N}) \quad \overline{\{T^{mk}(y) \mid k \in \mathbb{Z}\}} = X$.

Remark 4.19. We remark in passing that, as X here is a Baire space without isolated points, the set $\overline{\{T^k(y) \mid k \in \mathbb{Z}\}}$ is equal to X if and only if $\lambda(\underset{\sim}{y}) = X$ where $\underset{\sim}{y} \in \varprojlim (X, T)$ is the full orbit of y.

To prove this theorem we revisit Gambaudo and Martens' combinatorial approach for constructing self-maps of the Cantor set (or indeed any 0-dimensional compact metric space) via *graph covers* (see [GM06]).

4.5.1 *Graph covers*

It is a well-known fact that any totally disconnected, compact and Hausdorff space (these are sometimes called *Stone spaces*) can be obtained as an inverse limit of a (countable) system of discrete finite spaces. This property actually characterises Stone spaces via Stone duality (see [Nag70, Proposition 8-5]) which is why these are also occasionally called *profinite spaces*.

It turns out, using similar ideas, that it is possible to give a complete description not just for Stone spaces, but also for the self-maps on them by adding arrows to the discrete spaces forming the inverse system and, thus, creating an inverse limit of directed graphs. We shall now briefly recall the main results of this theory following Shimomura's treatment in [Shi14].

A *graph*[5] is a pair $G = (V, E)$ where V is a finite set of vertices and $E \subseteq V \times V$ is a set of directed edges. We additionally require that each vertex has at least one outgoing and one incoming edge. A

[5] All graphs we consider are directed.

graph homomorphism between graphs (V, E) and (V', E') is a vertex map $\phi\colon V \to V'$ which respects the edges, i.e. for any pair $(v, w) \in E$ it must be $(\phi(v), \phi(w)) \in E'$. A graph homomorphism is said to be *+-directional* if $\phi(w_1) = \phi(w_2)$ whenever both $(v, w_1) \in E$ and $(v, w_2) \in E$. If additionally $(v_1, w) \in E$ and $(v_2, w) \in E$ implies $\phi(v_1) = \phi(v_2)$ it is said that ϕ is *bidirectional*. A *graph cover* is a +-directional homomorphism of graphs that is also *edge-surjective* meaning that the map which ϕ naturally induces on the set of edges $\phi\colon E \to E'$ is surjective.

Given a sequence of graph covers $G_0 \xleftarrow{\phi_0} G_1 \xleftarrow{\phi_1} G_2 \xleftarrow{\phi_2} \cdots$ one forms a Stone space as the inverse limit

$$G_\infty = \varprojlim G_i = \left\{ (v^{(i)})_{i\geq 0} \in \prod_{i=0}^{\infty} V_i \mid v^{(i)} = \phi_i(v^{(i+1)}) \text{ for all } i \in \mathbb{N}_0 \right\}.$$

It is possible to define a self-map $\phi_\infty\colon G_\infty \to G_\infty$ by setting

$$\phi_\infty\left((v^{(i)})_{i\geq 0}\right) = \left(\phi_i(w^{(i+1)})\right)_{i\geq 0}$$

where $w^{(i)}$ is any vertex for which the edge $(v^{(i)}, w^{(i)})$ is in E_i. Intuitively, the map is given by the rule 'follow the arrows' except that if there are more than one outgoing arrow from the chosen vertex at any given level, we might need to peek at one level below whose finer resolution will help us decide which of the arrows to follow. +-directionality of the covers ensures that this process, or indeed ϕ_∞ is a well-defined continuous self-map of the Stone space G_∞. If additionally each of the bonding maps in the sequence is bidirectional then the map ϕ_∞ is a homeomorphism, see [Shi14, Lemma 3.5].

The full correspondence is given by

Theorem 4.20 ([Shi14, Theorem 3.9])**.** *A topological (compact, metric, surjective) dynamical system is 0-dimensional if and only if it is topologically conjugate to G_∞ for some sequence of covers $G_0 \xleftarrow{\phi_0} G_1 \xleftarrow{\phi_1} G_2 \xleftarrow{\phi_2} \cdots$.*

We remark that these results have a flavour of those by Mioduszewski in [Mio63] but are more elaborate as here we are dealing only with 0-dimensional spaces.

We are now ready to construct the system described in Theorem 4.18. The inspiration for this came from the construction in [Shi16b]. Very recently similar techniques appeared in [Shi16a].

4.5.2 *Proof of Theorem 4.18*

Let $G_0 = (\{v_0^{(0)}\}, \{(v_0^{(0)}, v_0^{(0)})\})$ be one vertex with a self-loop. Given G_i we shall inductively define G_{i+1}. The graph G_{i+1} will consist of $|V_{i+1}| = (|V_i|+C_i)\cdot(i+1)!+1$ vertices[6] $V_{i+1} = \{v_0^{(i+1)}, \dots, v_{|V_{i+1}|-1}^{(i+1)}\}$ which make a cycle in that order, and additionally E_{i+1} also contains a self-loop at $v_0^{(i+1)}$, see Figure 4.2. The constants $C_i \in \mathbb{N}$ are inductively chosen so that $|V_i| + C_i$ and $(i+1)!$ are co-prime. It remains to specify the cover $\phi_i \colon V_{i+1} \to V_i$ by the formula

$$\phi_i(v_k^{(i+1)}) = v_{I(i,k)}^{(i)}$$

where

$$I(i,k) = \begin{cases} 0, & \text{if } k \equiv 0, 1, 2, \dots, \text{ or } C_i \pmod{|V_i| + C_i}, \\ l, & \text{otherwise, where } l \equiv k - C_i \pmod{|V_i| + C_i}. \end{cases}$$

It is perhaps easier to see what is going if we look at how ϕ_i maps the edges of G_{i+1} onto the edges of G_i. Let us denote the self loops in G_i and G_{i+1} by e_0 and e_0' respectively, and let $\{f_1, \dots, f_{|V_i|}\}$ and $\{f_1', \dots, f_{|V_{i+1}|}'\}$ be successive edges in their bigger cycles respectively. The formulae above capture the fact that the cycle $\{f_1', \dots, f_{|V_{i+1}|}'\}$ on the $(i+1)^{\text{st}}$ level is wound $(i+1)!$ times over the full circuit (with the first edge being repeated C_i times) in the graph G_i in the following succession

$$\underbrace{\underbrace{e_0 \dots e_0 f_1 \dots f_{|V_i|}}_{C_i \text{ times}} \quad \underbrace{e_0 \dots e_0 f_1 \dots f_{|V_i|}}_{C_i \text{ times}} \quad \cdots \quad \underbrace{e_0 \dots e_0 f_1 \dots f_{|V_i|}}_{C_i \text{ times}}}_{(i+1)! \text{ times}} \quad e_0.$$

[6] Thus $|V_n|$ grows faster than superfactorials $n!(n-1)! \cdots 2!1!$.

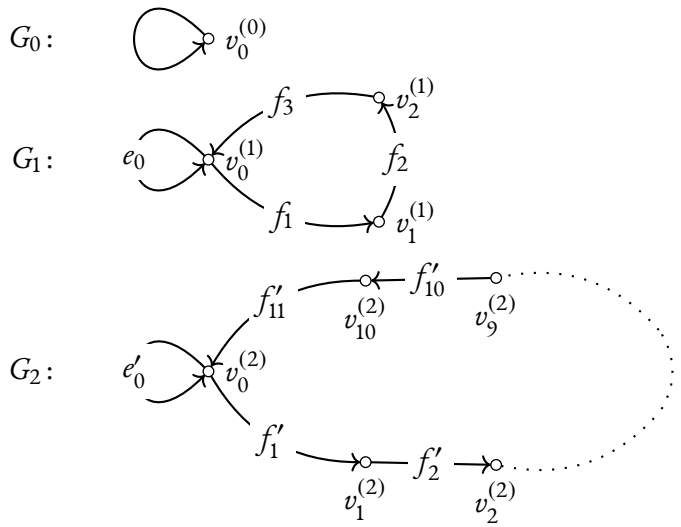

Figure 4.2: The first 3 steps of the construction, with $C_0 = 1$ and $C_1 = 2$.

In particular, looking at Figure 4.2, $\phi_1 \colon G_2 \to G_1$ maps denoted edges as follows:

$$\begin{array}{cccccccccccc} f'_1, & f'_2, & f'_3, & f'_4, & f'_5, & f'_6, & f'_7, & f'_8, & f'_9, & f'_{10}, & f'_{11} \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ e_0, & e_0, & f_1, & f_2, & f_3, & e_0, & e_0, & f_1, & f_2, & f_3, & e_0 \end{array}$$

Also note the additional twist over e_0 at the end which is needed to ensure that these are bidirectional covers.

Let G_∞ be the inverse limit of the just constructed inverse system $G_0 \xleftarrow{\phi_0} G_1 \xleftarrow{\phi_1} G_2 \xleftarrow{\phi_2} \cdots$. As each vertex in G_i is covered with at least 2 vertices of G_{i+1} one easily checks that G_∞ has no isolated points and is therefore homeomorphic to the Cantor set. As we mentioned earlier, the sequence consists of bidirectional covers and $\phi_\infty \colon G_\infty \to G_\infty$ is thus a homeomorphism of the Cantor set.

It remains to be seen that (G_∞, ϕ_∞) is almost totally minimal. One fixed point of the system is clearly $(v_0^{(i)})_{i\geq 0}$. Any other point $(v_{p_i}^{(i)})_{i\geq 0} \in G_\infty$ will have $p_i > 0$ for all i large enough. Let $x_\infty = (v_{p_i}^{(i)})_{i\geq 0}$ be one such a point. We wish to show that for any given $m \in \mathbb{N}$ the full orbit under ϕ_∞^m of this point is dense in G_∞.

We shall denote by $[v_k^{(i)}]$ the set of all the points in G_∞ with i^{th} coordinate equal to $v_k^{(i)}$. Recall that these sets $[v_k^{(i)}]$, where $i \in \mathbb{N}_0$ and $0 \leq k \leq |V_i| - 1$, form a clopen basis for the topology on G_∞. They

are neatly nested within each other and those on the level $i+1$ refine all those on levels less than i. In fact one has the relation $[v_k^{(i+1)}] \subseteq [\phi_i(v_k^{(i+1)})] = [v_{I(i,k)}^{(i)}]$. It will thus suffice to see that for any level $L \in \mathbb{N}_0$, one can find ϕ_∞^m-iterates of x_∞ that hit each of the sets $[v_k^{(L)}]$, for $0 \leq k \leq |V_L| - 1$.

Let L be one such level and choose $\mathcal{M} \geq \max\{L, m-1\}$ large enough so that $p_{\mathcal{M}+1} > 0$. We claim that $\phi_\infty^{(\mathcal{M}+1)!}$-iterates of x_∞ hit each of the sets $[v_k^{(\mathcal{M})}]$, for $0 \leq k \leq |V_{\mathcal{M}}| - 1$, which will suffice as $m|(\mathcal{M}+1)!$ and $\mathcal{M} \geq L$.

Note that as $p_{\mathcal{M}+1} > 0$ it is possible to infer where a large portion of iterates of x_∞ are mapped within level $\mathcal{M}+1$ without actually having to look at what happens at the level below[7]. In particular we know that $\phi_\infty^{-p_{\mathcal{M}+1}+k}(x_\infty) \in [v_k^{(\mathcal{M}+1)}]$ for each $0 \leq k \leq |V_{\mathcal{M}+1}| - 1$. This means that those same iterates hit each of the sets $[v_k^{(\mathcal{M})}]$, for $0 \leq k \leq |V_{\mathcal{M}}| - 1$, at least $(\mathcal{M}+1)!$ times, as each of the vertices in $G_{\mathcal{M}+1}$ covers those of $G_{\mathcal{M}}$ with at least multiplicity $(\mathcal{M}+1)!$. This, along with the choice of $C_{\mathcal{M}}$ (the number of successive repetitions of $v_0^{(\mathcal{M})}$ in the covering map), creates an apt offset implying that $\phi_\infty^{(\mathcal{M}+1)!}$-iterates of x_∞ end up in all of the sets $[v_k^{(\mathcal{M})}]$.

To put it precisely, let $p_{\mathcal{M}+1} = q(\mathcal{M}+1)! + r$ with $0 \leq r < (\mathcal{M}+1)!$ as in Euclidean division. Then if $k = s(\mathcal{M}+1)! + r$ with $0 \leq s \leq |V_{\mathcal{M}}| + C_{\mathcal{M}} - 1$ we have $\phi_\infty^{(s-q)(\mathcal{M}+1)!}(x_\infty) = \phi_\infty^{-q(\mathcal{M}+1)!-r+k}(x_\infty) = \phi_\infty^{-p_{\mathcal{M}+1}+k}(x_\infty) \in [v_k^{(\mathcal{M}+1)}] \subseteq [v_{I(\mathcal{M},k)}^{(\mathcal{M})}]$. This is justified by noting that k falls within the required range

$$0 \leq r \leq k \leq (\mathcal{M}+1)!(|V_{\mathcal{M}}| + C_{\mathcal{M}}) = |V_{\mathcal{M}+1}| - 1.$$

As $|V_{\mathcal{M}}| + C_{\mathcal{M}}$ and $(\mathcal{M}+1)!$ are co-prime it is an elementary number theoretic fact that $s(\mathcal{M}+1)!$, and hence also $k = s(\mathcal{M}+1)! + r$, will run through all the residue classes modulo $|V_{\mathcal{M}}| + C_{\mathcal{M}}$ as s runs through $\{0, 1, \ldots, |V_{\mathcal{M}}| + C_{\mathcal{M}} - 1\}$. Inspecting the definition of $I(\mathcal{M}, k) = I(\mathcal{M}, s(\mathcal{M}+1)! + r)$, we see that this attains all the values in $\{0, 1, \ldots, |V_{\mathcal{M}}| - 1\}$ when s goes through $\{0, 1, \ldots, |V_{\mathcal{M}}| + C_{\mathcal{M}} - 1\}$. This completes our proof. □

[7] By the level below we mean level $\mathcal{M}+2$, i.e. a level with a finer structure.

Using Theorem 4.18 we can now prove an interesting embedding result which will later help us to construct a map on the Cantor set for which the induced map on the hyperspace has period 2 and 3 points but none of the period 6.

Theorem 4.21. *Let (Y, S) be a 0-dimensional minimal system. There exist a system $(\hat{Y}, \hat{S})$ on the Cantor set $\hat{Y}$ such that:*

(i) (Y, S) dynamically embeds into $(\hat{Y}, \hat{S})$ as a nowhere dense set via $\iota\colon Y \to \hat{Y}$,

(ii) Every full $\hat{S}^m$-orbit of any point $y \in \hat{Y} \setminus \iota(Y)$ is dense in $\hat{Y}$ for every $m \in \mathbb{N}$.

Let (X, T) be the ATM system we constructed in Theorem 4.18. Let $X = \mathcal{A} \sqcup B$ be a separation into two disjoint clopen non-empty sets. Further assume that the fixed point x_0 of the ATM is in $\mathcal{A}$. For any $x \in X$ let

$$N(x, B) = \min\{k \in \mathbb{N}_0 \mid T^{-k}(x) \in B\}$$

be the time elapsed since x last visited B. If the minimum above does not exist, we set $N(x, B) = \infty$. Clearly $N(x, B) = 0$ for all $x \in B$, but it is also finite for all other $x \in X$ which have their backward orbit dense in X which we know is a G_δ dense set in X with empty interior. In fact, $N(x, B)$ is finite on a much larger (dense and open) set $U = \{x \in X \mid N(x, B) < \infty\}$. To see that U is open it suffices to notice that $N(x, B)$ is a locally constant function. What we mean by this is that for any point $x \in U$ there exists a (cl)open neighbourhood containing x on which the function $N(\,\cdot\,, B)$ is constant ($= N(x, B)$). This is easy to see as one just needs to ensure that the neighbourhood is small enough so that the $N(x, B)$ backward iterates of the points in it follow tightly those of x.

Going forward fix a point $y_0 \in Y$ to act as the origin and set

$$\hat{Y} = \overline{\{(x, S^{N(x,B)}(y_0)) \in X \times Y \mid N(x, B) < \infty\}}$$

where the closure is taken in $X \times Y$. Note that $\hat{Y}$ is still homeomorphic to the Cantor set since it is defined as the closure of a subset of the Cantor set with no isolated points. It is also immediately clear that $\hat{Y}$ can be written as

$$\hat{Y} = \overline{\bigcup_{k=0}^{\infty} \hat{S}^k(B \times \{y_0\})}$$

where $\hat{S}$ is the map on the product $X \times Y$ given by $\hat{S} = \pi \circ S \circ \pi$ and where π is a partial projection onto $X \times \{y_0\}$ defined by

$$\pi(x, y) = \begin{cases} (x, y), & \text{if } x \in A, \\ (x, y_0), & \text{if } x \in B. \end{cases}$$

To prove that this is the unique minimal closed strongly $\hat{S}$-invariant set containing $B \times \{y_0\}$ it only remains to be seen that $\hat{Y} \subseteq \hat{S}(\hat{Y})$. Actually, $B \times \{y_0\} \subseteq \hat{S}(\hat{Y})$ will suffice.

To see this let $\pi_X \colon X \times Y \to X$ be the projection to the first coordinate given by $\pi_X(x, y) = x$. This is in fact a semi-conjugacy as one can verify that $\pi_X \circ \hat{S} = T \circ \pi_X$. Using this we get $T(\pi_X(\hat{Y})) = \pi_X(\hat{S}(\hat{Y})) \subseteq \pi_X(\hat{Y})$ and so $\pi_X(\hat{Y})$ is a T-invariant set clearly containing B. Remembering that the set of points with their forward orbit dense under T is itself dense in X we conclude that $\pi_X(\hat{Y}) = X$.

Now given any $(x, y_0) \in B \times \{y_0\}$ set $x_{-1} = T^{-1}(x)$. Since $\pi_X(\hat{Y}) = X$, there exists a point $(x_{-1}, y) \in \hat{Y}$. Then clearly $\hat{S}(x_{-1}, y) = (x, y_0)$ which concludes the proof that $B \times \{y_0\} \subseteq \hat{S}(\hat{Y})$.

Another important thing to note is that the map π_X restricted to $\hat{Y}$ is almost a homeomorphism as it is injective everywhere except possibly at some points of $\pi_X^{-1}(X \setminus U) \cap \hat{Y}$. To see that π_X is injective on $\pi_X^{-1}(U) \cap \hat{Y}$ recall that for any $x \in U$ one can find a clopen neighbourhood V_x where $N(\,\cdot\,, B)$ is constant and hence $V_x \times \{S^{N(x,B)}(y_0)\}$ is contained in $\hat{Y}$ and this is the only intersection of $\hat{Y}$ with the rectangle $V_x \times Y$. In other words, π_X is invertible on the part of $\hat{Y}$ contained in the strip $V_x \times Y$.

This observation allows us to prove the following lemma.

Lemma 4.22. *For any fixed natural number $m \in \mathbb{N}$, $\hat{Y}$ is the least closed $\hat{S}^m$-invariant set that contains $B \times \{y_0\}$.*

Proof. Let $m \in \mathbb{N}$ be fixed, and let $F \subseteq \hat{Y}$ be a closed, $\hat{S}^m$-invariant subset of $\hat{Y}$ containing $B \times \{y_0\}$. We need to see that $F = \hat{Y}$. Reasoning similarly as before, we see that $\pi_X(F)$ is a closed T^m-invariant

subset of X containing B. Almost minimality of (X, T^m) then gives $\pi_X(F) = X$. If we now remember that π_X is injective on $\pi_X^{-1}(U) \cap \hat{Y} = \{(x, S^{N(x,B)}(y_0)) \mid N(x, B) < \infty\}$, we get that F contains $\{(x, S^{N(x,B)}(y_0)) \mid N(x, B) < \infty\}$ and from there immediately $F = \hat{Y}$. □

Remark 4.23. The statement of Lemma 4.22 could equivalently be written as

$$\hat{Y} = \overline{\bigcup_{k=0}^{\infty} \hat{S}^{km}(B \times \{y_0\})}, \qquad \text{for any } m \in \mathbb{N}.$$

Proof of Theorem 4.21. As we said before, $\hat{Y}$ is the closure of a subset of the Cantor set $X \times Y$ with no isolated point and is thus homeomorphic to the Cantor set itself.

We shall now see that $\{x_0\} \times Y$ is indeed contained in $\hat{Y}$. This follows from the fact that $\hat{Y}$ fully projects on the X and thus we can conclude that there exists $y \in Y$ such that $(x_0, y) \in \hat{Y}$. Since we know that $\hat{Y}$ is $\hat{S}$-invariant, that the dynamics of $\hat{S}$ coincides with that of S on $\{x_0\} \times Y$, and because (Y, S) is minimal we infer that $\{x_0\} \times Y \subset \hat{Y}$.

Since all the points in $\{x_0\} \times Y$ are limit points of those in $\{(x, S^{N(x,B)}(y_0)) \mid N(x, B) < \infty\}$ we can conclude that the interior of $\{x_0\} \times Y$ inside $\hat{Y}$ is empty. This concludes the proof of the statement (i) above where the map ι is given by $y \mapsto (x_0, y)$.

To see (ii), let $(x, y) \in \hat{Y}$ be a point with $x \neq x_0$ and let $m \in \mathbb{N}$ be fixed. We know that the closure of any full $\hat{S}^m$-orbit of (x, y) is a closed $\hat{S}^m$-invariant set. Since π_X is a semi-conjugacy the projection of this set to X is going to be the closure of some full obit of x under the map T^m. Almost total minimality allows us to conclude that this projection is the whole of X and thus reasoning as in Lemma 4.22 we conclude that the closure of this full orbit of (x, y) under $\hat{S}^m$ is the whole of $\hat{Y}$. □

From Theorem 4.21 one directly obtains the following corollary.

Corollary 4.24. *Theorem 4.21 holds for any* 0*-dimensional system* (Y, S) *that can be obtain as a power of a minimal system.*

Proof. Let (Y, R) be a minimal system such that $R^n = S$ and apply Theorem 4.21 to obtain the system $(\hat{Y}, \hat{R})$. Then the system which embeds (Y, S) as a nowhere dense set and in which any point outside this embedding has its full orbit dense under any iterate of the map is simply $(\hat{Y}, \hat{S} = \hat{R}^n)$. □

This poses a question:

Question 4.25. Which 0-dimensional systems cannot be represented as a power of a minimal system? Does Theorem 4.21 hold for those systems?

4.6.1 $2, 3 \in \mathrm{Per}(2^T) \not\Rightarrow 6 \in \mathrm{Per}(2^T)$

We can now use Theorem 4.21 to construct a map on the Cantor set for which the induced map on the hyperspace has period 2 and 3 points but none of the period 6.

Let (Y, S) be a two point cycle, i.e. let $Y = \{x_0, x_1\}$ and let $S\colon Y \to Y$ be a map swapping the two points in Y. Apply Theorem 4.21 to obtain the system $(\hat{Y}, \hat{S})$ into which (Y, S) is dynamically embedded. We identify Y with its image in $\hat{Y}$ so we can write $Y \subset \hat{Y}$. Now take a product of $(\hat{Y}, \hat{S})$ with a three cycle and quotient out the six cycle obtained from three copies of the cycle $\{x_0, x_1\}$. More precisely let $Z = \hat{Y} \times \{0, 1, 2\}/_\sim$ be a quotient space obtained by gluing the points in $\{(x_0, i) \mid i = 0, 1, 2\}$ together and likewise the points in $\{(x_1, i) \mid i = 0, 1, 2\}$. Let us call those two points z_0 and z_1 respectively. We define a map $R\colon Z \to Z$ by:

$$R(y, i) = (\hat{S}(y), i + 1 \bmod 3).$$

Note that this rule respects the quotient relation on $\hat{Y} \times \{0, 1, 2\}$ and therefore accounts for a well-defined map on Z.

We claim that (Z, R) is a system for which $\mathrm{Per}(2^R) = \{1, 2, 3\}$. The whole set $Z \in 2^Z$ is of course a fixed point. It is clear that $\{z_0\} \leftrightarrow \{z_1\}$ is a 2-cycle in 2^Z as z_0 is 2-periodic for R. It is also clear that the natural embedding of $\hat{Y}$ in Z as $\hat{Y} \times \{0\}$ produces a 3-cycle in 2^Z. It remains to show that no 6-cycle, or indeed any other cycle, in 2^Z exists. For if it did, the set $F \in 2^Z$ of period m other than 1, 2, or 3 would have to contain at least one point $(y, i) \in Z$ other than z_0 or z_1. Note that $y \in \hat{Y} \setminus Y$, and let $\underline{y} = (\dots, y_{-1}, y_0, y_1, y_2, \dots) \in \varprojlim \left(\hat{Y}, \hat{S}\right)$ be a full orbit of $y = y_0$ such that

$$\{(y_{mk}, i + mk \bmod 3) \mid k \in \mathbb{Z}\} \subseteq F.$$

In particular $\overline{\{(y_{mk}, i + mk \bmod 3) \mid k \in 3\mathbb{Z}\}} = \overline{\{(y_{3mk}, i) \mid k \in \mathbb{Z}\}} \subseteq F$. Now part (*ii*) of Theorem 4.21 guaranties that the full orbit $(\dots, y_{-3m}, y_0, y_{3m}, y_{6m}, \dots) \in \varprojlim \left(\hat{Y}, \hat{S}^{3m}\right)$ is dense in $\hat{Y}$. Thus $\hat{Y} \times \{i\}/_\sim \subseteq F$. Since $i \in \{0, 1, 2\}$ was arbitrary, the previous reasoning shows that in fact $F = \hat{Y} \times I/_\sim$ where $I \subseteq \{0, 1, 2\}$ is non-empty. But then clearly the period k of F is either 1 or 3. This contradiction concludes the proof that $\mathrm{Per}(2^R) = \{1, 2, 3\}$.

4.6.2 *General p and q*

It is easy to see how the construction above generalises to obtain a system (Z, R) for which $\mathrm{Per}(2^R)$ is the set of divisors of p or q where p and q are two positive integers.

As before let (Y, S) be a p-cycle embedded into $(\hat{Y}, \hat{S})$ by virtue of Theorem 4.21. We set $Z = \hat{Y} \times \{0, 1, \dots, q-1\}/_\sim$ where $\sim$ identifies q distinct sets of points coming from the p-cycle into one p-periodic orbit $\{z_0, \dots, z_{p-1}\} \subset Z$. The map $R\colon Z \to Z$ given by:

$$R(y, i) = (\hat{S}(y), i + 1 \bmod q)$$

is once again well-defined and clearly satisfies $p, q \in \mathrm{Per}(2^R)$. We claim that any other $m \in \mathrm{Per}(2^R)$ must be a divisor of either p or q.

Firstly, if we assume that the periodic point $F \in 2^Z$ of period m is completely contained in $\{z_0, \dots, z_{p-1}\}$ then it is not hard to see that $m|p$. If otherwise there exists a point $(y, i) \in F$ other than any of $z_0, \dots, z_{p-1}$ then arguing as before we can show that $\hat{Y} \times \{i\}/_\sim \subseteq F$. Thus $F = \hat{Y} \times I/_\sim$ for some non-empty $I \subseteq \{0, 1, \dots, q-1\}$. From here we immediately get that $m|q$. Therefore we have proved the following theorem.

Theorem 4.26. *Let $p, q \in \mathbb{N}$. There exists a continuous onto map of the Cantor set $R\colon Z \to Z$ for which the periods appearing in the induced map are exactly divisors of either p or q, i.e.*

$$\mathrm{Per}(2^R) = \{m \in \mathbb{N} \mid m|p \text{ or } m|q\}.$$

In this section we shall be concerned only with periodic points in 2^X made up entirely of periodic points for (X, T). We shall see that in this setting the anomalies such as those in Theorem 4.26 are not possible.

Firstly note that one can reduce the problem to studying periodic points of $T^{<\omega}$. Namely, let $F \in 2^X$ be a periodic point with fundamental period $k = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ where each point of F is periodic under T. Then as in the proof of Theorem 4.17 we can find points $x_1, \ldots, x_r$ in F for which $p_i^{\alpha_i} \in \mathrm{Per}(\underline{x}_i)$ where each $\underline{x}_i$ is the unique full orbit of x_i that is periodic. It could happen that some of these x_is are represented by the same point or by points that are members of the same full orbit, which in this case means the same cycle. For a moment let us assume that this is not the case and that all of these r full orbits are mutually disjoint when considered as subsets of X. Then for each $1 \leq i \leq r$ let $F_i \in 2^X$ be the canonical $k_i = p_i^{\alpha_i}$-periodic set given by $F_i = \{x_i, T^{k_i}(x_i), T^{2k_i}(x_i), \ldots\}$. Note that F_is are finite and mutually disjoint, and thus $\bigcup_{i=1}^r F_i$ is a k-periodic point for 2^T but also for $T^{<\omega}$. This proves the reduction when distinct prime factors of k can be represented by mutually disjoint orbits.

Suppose now that x_1 and x_2 belong to the same orbit. As they are both periodic it must be that $\underline{x}_1$ is just shifted $\underline{x}_2$. We then apply Proposition 4.16 to get $l_1 = k_1 k_2 \in \mathrm{Per}(\underline{x}_1) = \mathrm{Per}(\underline{x}_2)$. Grouping the elements of the same orbits in this way and discarding all but one representative for each of the orbits we obtain a factorisation of $k = l_1 \cdots l_s$ into s co-prime factors where $s \leq r$ and where each l_i is an element of some $\mathrm{Per}(\underline{x}_{t(i)})$ with $x_{t(1)}, \ldots, x_{t(s)}$ all belonging to distinct orbits. After setting $F_i = \{x_{t(i)}, T^{l_i}(x_{t(i)}), T^{2l_i}(x_{t(i)}), \ldots\}$ the union $\bigcup_{i=1}^s F_i$ will again be a k-periodic point for both 2^T and $T^{<\omega}$. We thus obtained the reduction in general.

We shall now give explicit formulae for $\mathrm{Per}(T^{(n)})$, $\mathrm{Per}(T_n)$, and $\mathrm{Per}(T^{<\omega})$ in terms of $\mathrm{Per}(T)$. Recall that $T^{(n)}$ is an induced map on the n-fold Cartesian product $X^{(n)} = X \times \cdots \times X$, that T_n is the restrictions of 2^T on $F_n(X)$ — the space of subsets of X of cardinality at most n, and that $T^{<\omega}$ is the restriction of 2^T on $F(X)$ — the space of subsets of X of finite cardinality.

Proposition 4.27. *We have the following identities*

$$\mathrm{Per}(T^{(n)}) = \Big\{[m_1, \dots, m_n] \mid m_i \in \mathrm{Per}(T) \text{ for all } 1 \le i \le n\Big\}, \tag{4.4}$$

$$\mathrm{Per}(T_n) = \bigcup_{l=1}^{n} \Big\{[d_1, \dots, d_l] \mid d_i | m_i \in \mathrm{Per}(T) \text{ for all } 1 \le i \le l, \text{ and } \frac{m_1}{d_1} + \cdots + \frac{m_l}{d_l} \le n\Big\}, \tag{4.5}$$

$$\mathrm{Per}(T^{<\omega}) = \bigcup_{l=1}^{\infty} \Big\{[d_1, \dots, d_l] \mid d_i | m_i \in \mathrm{Per}(T) \text{ for all } 1 \le i \le l\Big\}, \tag{4.6}$$

where $[k_1, \dots, k_m]$ *stands for the least common multiple of* $k_1, \dots, k_m$.

Proof. The statement (4.4) is easy. Given points $x_1, \dots, x_n$ with fundamental periods $m_1, \dots, m_n$ it is clear that the point $(x_1, \dots, x_n) \in X^{(n)}$ has fundamental period $[m_1, \dots, m_n]$. And conversely any periodic point of $T^{(n)}$ must arise in this fashion.

We shall now prove (4.5). Let $Q = \{x_1, \dots, x_n\}$ be a periodic point in $F_n(X)$. Clearly each x_i must be a periodic point of T. Then Q can be naturally partitioned into sets $Q_1, \dots, Q_l$ where each Q_i contains points belonging to the same cycle under T, and no two points from distinct Q_is belong to the same cycle. All the points in Q_i have the same period m_i under T. Let us denote the periods of $Q, Q_1, \dots, Q_l \in F_n(X)$ by $d, d_1, \dots, d_l$ respectively. Clearly $d = [d_1, \dots, d_l]$ and since $T_n^{m_i}(Q_i) = Q_i$ we also have $d_i | m_i$. By the construction we have $l \le n$ as every Q_i must contain at least one point from Q.

Lastly, we claim that $\frac{m_i}{d_i} \le |Q_i|$ where $|.|$ denotes the cardinality of a set. This is because $T_n^{d_i}$ acts on Q_i as a permutation over set of cardinality $|Q_i|$ of order at most $\frac{m_i}{d_i}$. But a permutation can not have order larger than the cardinality of the set it acts upon and, thus, we get the claim. From this it follows $\sum_{i=1}^{l} \frac{m_i}{d_i} \le |\bigcup_{i=1}^{l} Q_i| = |Q| = n$ which finishes the proof of the inclusion $\subseteq$ in (4.5).

The other inclusion follows by noting that any period of the form $[d_1, \dots, d_l]$ where $1 \le l \le n$, $d_i | m_i$ for some $m_i \in \mathrm{Per}(T)$, and $\sum_{i=1}^{l} \frac{m_i}{d_i} \le n$ can be realised by taking points $x_1, \dots, x_l$ of periods $m_1, \dots, m_l$ respectively and then setting $Q = \bigcup_{i=1}^{l} Q_i$ where for each $1 \le i \le l$ the set Q_i is formed of $\frac{m_i}{d_i}$ equispaced iterates of x_i. More precisely $Q_i = \{x_i, T^{d_i}(x_i), T^{2d_i}(x_i), \dots, T^{(\frac{m_i}{d_i}-1)d_i}(x_i)\}$. Thus (4.5) holds.

Lastly, note that $\mathrm{Per}(T^{<\omega}) = \bigcup_{n\in\mathbb{N}} \mathrm{Per}(T_n)$, therefore (4.6) follows immediately from (4.5). □

Remark 4.28. An important observation that implicitly drives the proof of the equation (4.5) above was that no periods are lost if one only considers points in $F_n(X)$ formed of equispaced[8] iterates of points in X. To take a simple example, consider a point x which is periodic with fundamental period 6 in some system (X, T). Even though $\{x, T(x)\}$, $\{x, T(x), T^2(x)\}$, and many others are period 6 points for T_3, there is a simple, canonical, equispaced point of period 6 for T_3, namely $\{x\}$.

Remark 4.29. Note that the statement (4.6) could equally be written as

$$\mathrm{Per}(T^{<\omega}) = [\mathcal{D}(\mathrm{Per}(T))]$$

where $\mathcal{D}(S) = \{d \in \mathbb{N} \mid d|n \text{ for some } n \in S\}$ is the set of divisors of S, and $[S] = \{[k_1, \dots, k_l] \mid l \in \mathbb{N}, k_1, \dots, k_l \in S\}$ is the set of least common multiples of a set $S \subseteq \mathbb{N}$. In particular, $\mathrm{Per}(T^{<\omega})$ is the smallest set containing $\mathrm{Per}(T)$ that is closed both under taking least common multiples and divisors.

Let us show how Proposition 4.27 can be used to compute the periods of induced maps on hyperspaces.

Example 4.30. Let $T\colon [0,1] \to [0,1]$ be an interval map of type 2^r ($r > 1$), i.e. a map for which $\mathrm{Per}(T) = \{1, 2, 2^2, \dots, 2^{r-1}, 2^r\}$. One possible construction of such a map is described in [BC92, Example I.13]. For any positive integer n, all the sets $\mathrm{Per}(T_n)$, $\mathrm{Per}(T^{(n)})$, $\mathrm{Per}(T^{<\omega})$ clearly contain $\mathrm{Per}(T)$. Note that in our special case, if $d_i|m_i$ and $m_i \in \mathrm{Per}(T)$ then $d_i \in \mathrm{Per}(T)$, and hence also $[d_1, \dots, d_l] = \max\{d_1, \dots, d_l\} \in \mathrm{Per}(T)$. Therefore $\mathrm{Per}(T_n) = \mathrm{Per}(T^{(n)}) = \mathrm{Per}(T^{<\omega}) = \mathrm{Per}(T)$. Note that using Theorem 4.8 and the fact $\mathrm{Per}(T) \subseteq \mathrm{Per}(2^T)$ we can also conclude $\mathrm{Per}(2^T) = \mathbb{N}$. ◁

[8] Formally, we say that $Q = \{x_1, \dots, x_m\}$ is an equispaced point in $F_n(X)$ if $\min\{k \geq 1 \mid T^k(x_i) \in Q\}$ does not depend on $1 \leq i \leq m$.

In this section we prove an extension of Theorem 4.8 to a non-compact setting where the underlying set is taken to be $\mathbb{R}$. Recall that in this case $2^{\mathbb{R}}$ denotes the set of all compact subsets of the reals and that $2^{\mathbb{R}}$ itself is no longer compact.

Theorem 4.31. *Let $T\colon \mathbb{R} \to \mathbb{R}$ be a continuous map over the reals. Then $\mathrm{Per}(2^T)$ is either $\emptyset$ or $\{1\}$ or $\{1, 2\}$ or $\mathbb{N}$.*

Proof. We shall consider 3 cases depending on the number of fixed points of the map T.

Case 1. If T has no fixed points then either $T(x) > x$ or $T(x) < x$ for all $x \in \mathbb{R}$ which means that $\mathrm{Per}(T) = \mathrm{Per}(2^T) = \emptyset$.

Case 2. If T fixes at least two points then T is either the identity (in which case $\mathrm{Per}(2^T) = \{1\}$) or there exists a non-degenerate interval $[a, b]$ whose endpoints are fixed by T but no other point in (a, b) is. Therefore either $T(x) > x$ or $T(x) < x$ for all the points $x \in (a, b)$.

Let us assume that the former is the case as the other possibility is treated analogously. Let $z = \min\{x \in (a, b] \mid T(x) = b\}$. If $z = b$ then $T((a, b)) = (a, b)$ and any point in (a, b) has a bounded (contained in $[a, b]$) fully non-recurrent full orbit with its forward iterates converging to b and its backward orbit converging to a. If on the other hand $a < z < b$ then any backward orbit (within $[a, b]$) of z is still converging to a and as $T(z) = b$ the point z is (boundedly) fully non-recurrent. In both cases Proposition 4.10 gives $\mathrm{Per}(2^T) = \mathbb{N}$.

Case 3. If T fixes exactly one point (say $T(0) = 0$) and further assume that $\mathrm{Per}(2^T) \supsetneq \{1\}$ as otherwise we are done. Let $F \in 2^{\mathbb{R}}$ be a compact set with period $p > 1$. Then $D = \bigcup_{k=0}^{p-1} T^k(F)$ is a compact set which is fixed by T and is not $\{0\}$. It then follows using the same reasoning as in [BC86, Lemma 3] that T^2 fixes at least two points. For convenience of the reader we reproduce the argument.

Clearly T^2 fixes the point 0 thus we need to find one additional T^2-fixed point. Set $m = \min D$ and $M = \max D$. As 0 is the only fixed point of T it is not hard to see that m and M must be on

the opposite sides of 0, i.e. $m < 0 < M$. Further we see that $T([m, 0]) \supseteq [0, M]$ and similarly $T([0, M]) \supseteq [m, 0]$. This means that there exists a point $z \in [m, 0)$ such that $T^2(z) = m$. As $T^2(m) \geq m$ and $T^2(z) = m \leq z$ there must exist a point in $[m, z]$ which is fixed by T^2. This point is clearly an additional fixed point for T^2 as $z < 0$.

Now that we know that T^2 fixes at least two points we can argue as in Case 2. Either T^2 is the identity map in which case $\mathrm{Per}(2^T) = \{1\}$ or $\{1, 2\}$, or there exist a point with a bounded fully non-recurrent full orbit under the map T^2. For definiteness let us call it $\underline{x}^{(2)} = (\dots, x_{-2}, x_0, x_2, \dots)$. Then $\underline{x} = (\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots)$ where $x_{2i+1} = T(x_{2i})$ still has bounded orbit and we claim it is fully non-recurrent. We shall show that $x_{-2k} \notin \lambda(\underline{x})$ for some $k \in \mathbb{N}$. Otherwise, for each $k \in \mathbb{N}$ there would exist a monotone (increasing to $+\infty$ or decreasing to $-\infty$) subsequence $(p_n^k)_n$ such that $x_{2p_n^k+1} \to x_{-2k}$. Note that we know that $x_{-2k} \notin \lambda(\underline{x}^{(2)})$ where this set denotes the limit set of $\underline{x}^{(2)} \in \varprojlim \left(\mathbb{R}, T^2\right)$. But then $x_{2p_n^k+2} \to x_{-2k+1}$ and thus $x_{-2k+1} \in \lambda(\underline{x}^{(2)})$. As $\lambda(\underline{x}^{(2)})$ is T^2 invariant we get that $x_{2k+1} \in \lambda(\underline{x}^{(2)})$ for all $k \in \mathbb{Z}$, and since it is closed we get that x_0 is in it too which contradicts the fact that $x_0 \notin \lambda(\underline{x}^{(2)})$ as x_0 is fully non-recurrent under T^2.

This completes the proof that for some $k \in \mathbb{N}$ the point x_{-2k} is fully non-recurrent under T and we can once more use Proposition 4.10 to conclude that $\mathrm{Per}(2^T) = \mathbb{N}$.

□

One could now use this result to deduce Theorem 4.8 as follows. Given an interval map $T\colon [0, 1] \to [0, 1]$ we first extend it over the whole $\mathbb{R}$ by setting

$$\tilde{T}(x) = \begin{cases} T(x), & \text{if } x \in [0, 1], \\ T(1), & \text{if } x \geq 1, \\ T(0), & \text{if } x \leq 0. \end{cases}$$

Clearly the option $\mathrm{Per}(2^{\tilde{T}}) = \emptyset$ is not viable any more as $\tilde{T}$ has to fix a point inside $[0, 1]$ and furthermore any $2^{\tilde{T}}$-periodic compact set is contained in $[0, 1]$ and thus 2^T-periodic with the same period.

We have mentioned before that it would be possible to give rewrite rules for translating a description of a systems given by graph covers to those using Bratteli-Vershik diagrams or Kakutani-Rokhlin towers. In this section we give an example of how this is done by providing another, equivalent construction of the system (X, T) from Theorem 4.18 using the former.

There are many references explaining the theory behind Bratteli-Vershik diagrams, e.g. [HPS92, Do003], but to us most relevant was a recent paper [AEG15] by Amini, Elliott, and Golestani where they give a category theory treatment of the matter highlighting the connection between these diagrams and essentially minimal systems.

Consider the infinite graph in Figure 4.3, where all the edges are oriented downwards even if this is not depicted there. Additionally, for each node other than the root L_0 which has no incoming edges, an ordering is given on the set of incoming edges into that node. This order is depicted in the figure as the order in which the edges connect to that node going from left to right.

We have yet to explain how this infinite graph is constructed. Recall that in the proof of Theorem 4.18 we obtained a sequence of numbers $(C_i)_{i\in\mathbb{N}_0}$ which was of some significance in the construction of the map T. Here, any node L_{i+1} will have only one incoming edge coming from the node L_i for any $i \in \mathbb{N}_0$. The incoming edges into the node R_{i+1} will be: C_{i+1} edges coming form the node L_i, then one edge from the node R_i, then another C_{i+1} edges form L_i, and another one from R_i, and this sequence of $C_{i+1} + 1$ edges is to be repeated in total $(i + 1)!$ times, after which the last one in comes the edge from L_i. Note that we could simply say that there are $(i + 1)! \cdot (C_{i+1}) + 1$ edges coming from L_i and $(i + 1)!$ from R_i into R_{i+1}, but by listing them as above, we implicitly specified the order on those edges incoming into the node R_{i+1} for any $i \in \mathbb{N}$.

The space we shall be considering is the space of all infinite paths starting from L_0 and then following the sequence of edges down through the vertices L_i/R_i. Encoding these paths as sequences of edges allows one to see this space as the Cantor set of symbolic sequences. It remains to specify the map on this space of infinite paths.

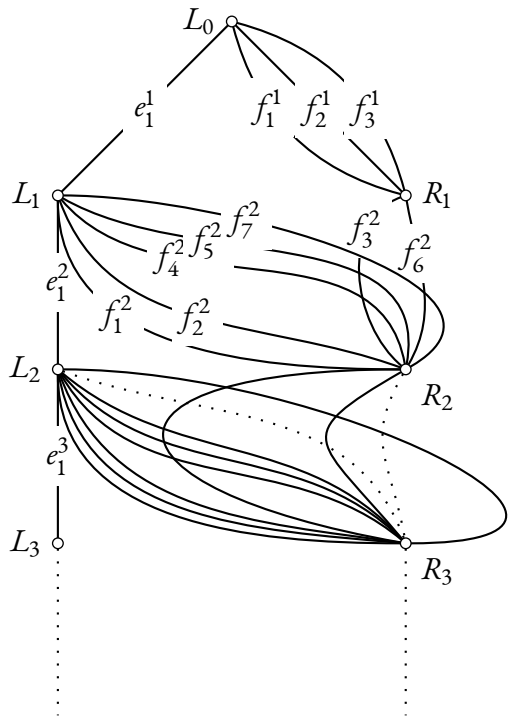

Figure 4.3: Bratteli-Vershik representation of the system from Theorem 4.18.

Recall that for each node other than the root there exists the minimal and the maximal ingoing edge into that node. A path is said to be minimal (resp. maximal) if it consists entirely of minimal (resp. maximal) edges. It is not hard to check that in our graph the path $e_1^1 e_1^2 e_1^3 \dots$ that is always staying on the left side is both minimal and maximal and no other minimal nor maximal path exits. This path will be the fixed point of our map. For any other path $p^1 p^2 p^3 \dots$ there must exist a $k \in \mathbb{N}_0$ such that the edge p^k is not maximal. Choose this k to be the smallest possible and then we let our function map this path to the path $q^1 q^2 \dots q^k p^{k+1} p^{k+2} p^{k+3} \dots$ where q^k is the successor of the edge p^k and q^is for $i < k$ are chosen in such a way that $q^1 q^2 \dots q^{k-1}$ forms the (unique) minimal path connecting L_0 and the source vertex of the edge q^k. Equivalently, $q^1 \dots q^k$ is chosen to be the successor of the path $p^1 \dots p^k$ with respect to the natural induced order on the sequence of paths (of length k) terminating at the source of p^{k+1}.

We leave to the reader to check that this does give the same (or formally, conjugated) system as the method given in Section 4.5. For this it is helpful to see how this representation was derived from Figure 4.2.

Each vertex above stands for a cycle in the graph covers representation. To be precise, the self-loop of the graph G_0 is represented by the root L_0, the self-loop of any G_i is represented by the vertex L_i, and the bigger cycle $v_0^{(i)} v_1^{(i)} \dots v_{|V_i|}^{(i)}$ in G_i corresponds to the right side vertex R_i. Finally the edges and the order in which they connect vertices L_i and R_i to the vertices L_{i+1} and R_{i+1} are to be inferred from the way the cycles in the graph G_{i+1} wrap around through the cycles of G_i. In particular, the loop e'_0 in G_2 just goes once through the loop e_0 in G_1 and hence only one edge from L_1 to L_2. On the other hand, the cycle $f'_1 f'_2 \dots f'_{11}$ in G_2 wraps firstly two times around e_0, then once around the bigger cycle in G_1, then repeats this, and then finally winds the last time over e_0. This means that the edges incoming into R_2 are: two edges from L_1, followed by one from R_1, then again two from L_1 and one from R_1, and lastly one from L_1.

At the very end we mention that the corresponding representation using Kakutani-Rokhlin towers is obtained by associating the cycles in the graphs G_i with the eponymous towers and the vertices within those cycles correspond to the levels of the towers.

4.10 CONCLUDING REMARKS

Nearly all the results contained in this chapter were driven by the problem of finding a complete description of all admissible pairs $(\mathrm{Per}(T), \mathrm{Per}(2^T))$. We say that a pair (A, B) of subsets of positive integers $A, B \subseteq \mathbb{N}$ is admissible if there exist a compact metric dynamical system (X, T) with $(A, B) = (\mathrm{Per}(T), \mathrm{Per}(2^T))$. The first step would be to find constraints on $\mathrm{Per}(2^T)$ given $\mathrm{Per}(T)$ and this already seems involved. We have shown that $\mathrm{Per}(2^T)$ must be closed under taking prime divisors, but the question remains if it is closed under taking divisors in general. Whatever turns out to be the truth, one will still be faced with the laborious task of constructing maps attaining all the admissible periodicity sets in-between, if one seeks for such a full characterisation. Some of these will be easy, but we have already gone through much trouble in order to construct examples having two co-prime periods in $\mathrm{Per}(2^T)$ but not their product. We conjecture that, more generally, a map with $\mathrm{Per}(2^T) = \{m \mid m|p \text{ for } p \in P\}$ exists for any finite $P \subset \mathbb{N}$.

One plausible strategy for obtaining such results easily would be to find a way to extend the embedding result from Theorem 4.21 to any, and not just minimal, Cantor set dynamics (Y, S). We have already remarked that this is possible for the systems which are a powers of a minimal system but anything more general than this is still out of our reach. We thus pose a question:

Question 4.32. Does Theorem 4.21 hold for any 0-dimensional system (Y, S)?

Should the answer to this question be yes, one could construct a system for which the set of periods in the hyperspace is $\{m \mid m|p \text{ for some } p \in P\}$ as follows. Let $p_1, p_2 \in P$ and let (Y, S) be the system with $\mathrm{Per}(2^S) = \{m \mid m|p_1 \text{ or } m|p_2\}$ as in Theorem 4.26. To add another period $p_3 \in P$ we simply imitate the proof of Theorem 4.26. We firstly embed (Y, S) within $(\hat{Y}, \hat{S})$ and then define $Z = \hat{Y} \times \{0, 1, \dots, p_3 - 1\}/_\sim$ where $\sim$ identifies p_3 copies of $Y \subset \hat{Y}$ into just one copy. The map over Z is given by $R\colon (y, i) \mapsto (\hat{S}(y), i + 1 \bmod p_3)$. It is not hard to check that $\mathrm{Per}(2^Z) = \{m \mid m|p_1 \text{ or } m|p_2 \text{ or } m|p_3\}$. This procedure can now be repeated with the system we have just obtained (Z, R) as a starting point. Continuing in this way we shall eventually add all the periods from P, building along the way systems with increasingly complex hierarchies of subsystems nested within each other. The final system obtained in this way will have $\{m \mid m|p \text{ for some } p \in P\}$ for the set of periods of its induced system.

Another possible direction for further research is investigating almost totally minimal systems and the spaces that admit them. This is related to a prominent question in topological dynamics to characterise spaces which admit minimal maps. One usually asks this question in the class of compact systems, but as we have seen on the example of the Cantor set, asking for almost minimality on compact spaces is related to minimality on non-compact spaces.

Let us elaborate, if (X^*, T) is a system on a non-compact X^*, and if furthermore T is a (totally) minimal homeomorphism, then there is a well-defined extension of T on the one point compactification of X^* that fixes the point at infinity and that is again a homeomorphism which is almost (totally) minimal.

Chapter 5

TWO ADDITIONAL RESULTS

The two additional results included in this chapter continue our study of the Cantor set dynamics. The first part is concerned with a beautiful result proved by Sherman in his PhD thesis [She12] *where he completely characterised the orbit structure of homeomorphisms of the Cantor set. A part of his work relies on a surprising result that a compact system consisting entirely of periodic points must either have a finitely based spectrum, a notion defined below, or otherwise have a non-trivial connected component. He proves this using somewhat involved point-set topology argument. In the first part of this chapter we show how by choosing an appropriate coarse graining one can obtain this as a consequence of a structural result about shift spaces.*

In the second part of this chapter using similar symbolic methods we answer a question posed by Barwell in his thesis [Bar11, BDG12]. *He asked whether in a system with the shadowing property each internally chain transitive set is necessarily an ω-limit set of some point in that system. We show that unexpectedly the answer is no. The terminology will be defined below.*

5.1 ON SHERMAN'S RESULT

Up until now, whenever we had a dynamical system (X, T) we assumed that X was a nice (Hausdorff, metric and often compact) topological space and that T was a continuous map on it. Let us for a moment forget about the topology and continuity — and just think of X as a set on which a map T is given. We can think of the space X as having the discrete topology where each point is isolated and where each $x \in X$ is connected by an arrow to its image $T(x)$.

If we think of these arrows as links in a chain then the whole space X breaks apart in a number of disjoint chains each of which is a *full orbit* of the map T. These orbits can be categorised in three main groups. Those which contain a loop of n points $z_1 \to z_2 \to \cdots \to z_n \to z_1$ are called *n-cycles*, those with a chain stretching infinitely in both directions $\cdots \to z_{-1} \to z_0 \to z_1 \to z_2 \to \dots$ are *$\mathbb{Z}$-orbits*, and those remaining are *$\mathbb{N}$-orbits*. Note that if T is a bijection then all the orbits are either cycles or $\mathbb{Z}$-obits. The cardinalities of each type of those orbits are denoted by σ_n, ζ and ν respectively. This information written as a sequence $(\nu, \zeta, \sigma_1, \sigma_2, \sigma_3, \dots)$ is called the *orbit spectrum* of (X, T) and it is denoted by $\sigma(T)$.

It is interesting to see how much the topology of X affects the orbit spectrum of a system. In particular one can ask which orbit spectra are admissible for a class of dynamical systems over, say, compact Hausdorff spaces. This very question was answered by Good, Greenwood, Knight, McIntyre, and Watson in [GGK+06] — and they further proved that knowing the orbit spectrum alone is enough to determine whether one can put a compact Hausdorff topology on an, a priori topologyless, system (X, T) which makes the map T continuous.

Under the additional assumption that T is a bijection, they obtain a similar characterisation of the admissible orbit spectra for the class of compact metric spaces. In this case there are no $\mathbb{N}$-orbits, hence $\nu = 0$.

Theorem 5.1 (Good et al. [GGK+06]). *Let $T : X \to X$ be a bijection. Then X admits a compact metric topology with respect to which T is a homeomorphism if and only if ζ and each σ_n, $n \in \mathbb{N}$ is either countable or has cardinality $\mathfrak{c}$, and either:*

(1) $|X| = \mathfrak{c}$; or

(2) $\zeta \neq 0$ and $\sigma_n \neq 0$ for some $n \in \mathbb{N}$; or

(3) $\sigma(T)$ is finitely based.

Recall that $\sigma(T)$ is said to be *finitely based* if the set $N = \{n \in \mathbb{N} \mid \sigma_n \neq 0\}$ is finitely generated — meaning that there exist a finite subset $G \subseteq N$ such that each $j \in N$ is a multiple of a number in G.

This theorem was further refined by Sherman in [She12] where he yields a criterion to make X the Cantor set. As an initial step towards his classification he obtains the following result:

Theorem 5.2 (Sherman [She12]). *If X is a compact Hausdorff space and $T\colon X \to X$ is a homeomorphism with $\zeta = 0$, i.e. all points are periodic, and if $\sigma(T)$ is not finitely based, then X has a non-degenerate connected component.*

We shall now show how to obtain this result using the coarse graining approach. Recall that the connected component of a point $x \in X$ is the union of all connected subspaces of X containing it. In a compact Hausdorff space this is easily seen to be the intersection of all clopen sets containing x, see [Eng89, Theorem 6.1.23]. This allows us to prove the following lemma.

Lemma 5.3. *Let X be a compact Hausdorff space and $x_1, x_2, \dots, x_n$ a finite number of points in it, no two of which belong to the same connected component of X. Then there exists a continuous map $\pi\colon X \to \{1, 2, \dots, n\}$ so that $\pi(x_i) = i$, for each $1 \leq i \leq n$.*

Proof. Since no two of the points $x_1, x_2, \dots, x_n$ share a connected component and since components can be written as intersections of clopen sets, there must exist clopen sets $U_1, U_2, \dots, U_n$ in X such that each U_i contains x_i and none of the other $n-1$ points.

We can easily ensure that the sets are disjoint just by taking $U_i' = U_i \setminus \bigcup_{j \neq i} U_j$. The sets U_i' satisfy the same property as those without dashes and are also pairwise disjoint.

Finally note that the remaining part of the space $R = X \setminus \bigcup_{i=1}^{n} U_i'$ is also clopen and we can define a continuous map satisfying the lemma as follows:

$$\pi(x) = \begin{cases} i & \text{if } \; x \in U_i' \quad \text{for some } \; 1 \leq i \leq n, \\ 1 & \text{if } \; x \in R. \end{cases}$$

□

Before we can proceed with the proof of Theorem 5.2 we have to recall some facts about shift spaces. The *full shift* over a finite discrete set of symbols $\mathcal{A}$ is the set $\mathcal{A}^{\mathbb{N}}$ consisting of all right-infinite sequences of elements in $\mathcal{A}$. The product topology on this set can be induced by the following metric: two distinct points $\xi = (\xi_1, \xi_2, \dots)$ and $\eta = (\eta_1, \eta_2, \dots)$ are 2^{-i} apart if and only if their i^{th} entries differ $\xi_i \neq \eta_i$ while the ones preceding them match $\xi_j = \eta_j$, for $1 \leq j < i$. This topology makes $\mathcal{A}^{\mathbb{N}}$

into a Cantor set and it is common to denote it by $\Sigma_{\mathcal{A}}$ or in case $\mathcal{A} = \{0, 1, \dots, n-1\}$ by Σ_n. There is a natural left shift map on this space $\sigma\colon \Sigma_{\mathcal{A}} \to \Sigma_{\mathcal{A}}$ given by:

$$\xi = (\xi_1, \xi_2, \xi_3, \dots) \overset{\sigma}{\mapsto} (\xi_2, \xi_3, \xi_4, \dots) = \sigma(\xi).$$

Any compact, σ-invariant subset $X \subseteq \Sigma_{\mathcal{A}}$ of the full shift is called a *shift space*. Here by σ-invariant we mean $\sigma(X) = X$ so that the shift map σ restricted to X is well-defined and surjective. More on this topic can be found for example in a textbook by Lind and Marcus [LM95].

The shift space systems are one of the best understood dynamical systems around. One can use their symbolic structure to effectively compute a number topological invariants such as their entropy, number of periodic points, structure of limit sets etc. This is why, quite often, they are used as toy models to test conjectures — and why finding a coarse graining from, or to a shift space leads to a better understanding of the system one studies.

Nevertheless, their dynamics can be quite restrictive. As we are about to see, a shift space consisting only of periodic points must be finite.

Lemma 5.4. *Let X be a shift space, i.e. a σ-invariant closed subset of the full shift over $n \in \mathbb{N}$ symbols. Then all points in (X, σ) are periodic if and only if X is a finite set.*

Proof. We shall first show that periodicity implies finiteness. For the sake of getting a contradiction assume that X is not finite and yet all points are periodic. Note that in the full shift over n symbols, and thus in any of its subshifts, there are at most n^k points of period k. Since X is infinite it must contain points with arbitrarily large periods.

Let $(x_i)_{i\in\mathbb{N}}$ be a sequence of points whose corresponding periods $(p_i)_{i\in\mathbb{N}}$ form a strictly increasing sequence of positive integers. By passing onto a subsequence, if necessary, we may assume that $\lim_{i\to\infty} x_i = x$. This limit is again a periodic point with some period $m \in \mathbb{N}$. By discarding an initial portion of the sequence we may assume that $p_1 > m$ and in particular no point of $(x_i)_{i\in\mathbb{N}}$ is equal to x. We can write x as

$$x = (\overline{\xi_1, \xi_2, \dots, \xi_m},\ \overline{\xi_1, \xi_2, \dots, \xi_m},\ \dots).$$

For each $i \in \mathbb{N}$ let $t(i) \in \mathbb{N}$ be the first position in which the sequences x_i and x differ. Since $\lim_{i\to\infty} x_i = x$ it must be that $\lim_{i\to\infty} t(i) = \infty$. By passing onto a subsequence yet again, if necessary, we may assume that $t(i)$ gives the same remainder $0 \leq r < m$ when divided by m for all $i \in \mathbb{N}$. Since the period of x is m this means that the symbol $\xi_{t(i)}$ in x at the position in which x and x_i differ is always ξ_r (in case $r = 0$ we set $\xi_0 = \xi_m$).

Let us now consider the sequence $(\sigma^{t(i)-1}(x_i))_{i\in\mathbb{N}}$. The initial coordinate of each of its elements is never ξ_r for the reasons given in the previous paragraph. By passing onto a subsequence for the third time, if necessary, we may assume that $(\sigma^{t(i)-1}(x_i))_{i\in\mathbb{N}}$ is again convergent and that the initial coordinate of its limit $\lim_{i\to\infty} \sigma^{t(i)-1}(x_i) = y$ is also not ξ_r. We can write

$$y = (\eta_1, \eta_2, \dots)$$

where $\eta_1 \neq \xi_r$.

Since $t(i)$ grows unboundedly the numbers $t(i) - 1 - m$ are positive for all large enough $i \in \mathbb{N}$. Therefore taking the limit $\lim_{i\to\infty} \sigma^{t(i)-1-m}(x_i)$ makes sense and it is clear from the construction that this limit is:

$$\lim_{i\to\infty} \sigma^{t(i)-1-m}(x_i) = (\overline{\xi_r, \xi_{r+1}, \dots, \xi_{r+m-1}},\ \eta_1, \eta_2, \dots)$$

where ξ_j for $j \geq m$ is defined as $\xi_{j \bmod m}$.

Similarly for any $k \in \mathbb{N}$ we have

$$\lim_{i\to\infty} \sigma^{t(i)-1-mk}(x_i) = (\underbrace{\overline{\xi_r, \xi_{r+1}, \dots, \xi_{r+m-1}},\ \dots\ , \overline{\xi_r, \xi_{r+1}, \dots, \xi_{r+m-1}}}_{k \text{ times}},\ \eta_1, \eta_2, \dots)$$

and $\eta_1 \neq \xi_r$.

By the initial assumption, all of these points we have constructed should be periodic under the shift map. But since all of them contain y in their forward orbit they must all belong to the same cycle consisting only of iterates of y. This is only possible if $y = (\overline{\xi_r, \xi_{r+1}, \dots, \xi_{r+m-1}},\ \overline{\xi_r, \xi_{r+1}, \dots, \xi_{r+m-1}},\ \dots)$ which is in contradiction with $\eta_1 \neq \xi_r$. This finishes the proof of one implication.

To prove the converse let us assume that X is finite. As σ is surjective on X it also must be bijective. This means that σ permutes the points of X and thus all of them are periodic. □

The following simple fact about coarse graining periodic points will also come in handy in the proof of Theorem 5.2.

Lemma 5.5. *Assume that (X, T) coarse grains onto (Y, S) via $\Xi\colon X \to Y$. Then any periodic point $x \in X$ coarse grains onto a periodic point $\Xi(x) \in Y$ and furthermore the period of $\Xi(x)$ divides that of x.*

Proof. Let us assume that the period of x is n. Then $T^n(x) = x$ and thus

$$\Xi(x) = \Xi(T^n(x)) = S^n(\Xi(x))$$

which means that $\Xi(x)$ is periodic. This however does not imply that the fundamental period of $\Xi(x)$ is n. Let $m \leq n$ be the least positive integer for which $S^m(\Xi(x)) = \Xi(x)$ and write $n = mq + r$ with $q \in \mathbb{N}$ and $0 \leq r < m$ as in Euclidean division. Then

$$\Xi(x) = S^n(\Xi(x)) = S^{mq+r}(\Xi(x)) = S^r(S^{mq}(\Xi(x))) = S^r(\Xi(x))$$

where in the last equality we used the fact that $\Xi(x)$ is m-periodic.

This now implies that $r = 0$ as m was chosen to be the least positive integer satisfying $S^m(\Xi(x)) = \Xi(x)$. Thus $n = mq$ and $m|n$. □

Now we have all the ingredients to prove Sherman's theorem.

Proof of Theorem 5.2. Since $\sigma(T)$ is not finitely based, the set $\mathcal{N} = \{n \in \mathbb{N} \mid \sigma_n \neq 0\}$ is infinite. Enumerate its elements in the increasing order by a sequence $(p_i)_{i\in\mathbb{N}}$. As $\mathcal{N}$ is not finitely generated, one can find a sequence of points $(x_i)_{i\in\mathbb{N}}$ in X such that for each $i \in \mathbb{N}$ none of the integers $p_1, p_2, \ldots, p_i$ divides the period of x_i.

Using compactness we can extract a convergent subsequence $(x_{s(i)})_{i\in\mathbb{N}}$. Its limit $\lim_{i\to\infty} x_{s(i)} = x \in X$ must again be periodic with period $p_m \in \mathcal{N}$ for some $m \in \mathbb{N}$. If all the connected components of X

were degenerate then each iterate of x would lie in its own component and applying Lemma 5.3 we could obtain a map $\pi\colon X \to \{1, 2, \dots, p_m\}$ sending each point of the cycle $\{x, T(x), \dots, T^{p_m-1}(x)\}$ to a different integer. Indeed we can choose π so that $\pi(T^{i-1}(x)) = i$ for $1 \leq i \leq p_m$.

The map π naturally induces a coarse graining Ξ into the full shift over p_m symbols $\{1, 2, \dots, p_m\}$ which is given by:

$$\Xi(x) = (\pi(x), \pi(T(x)), \pi(T^2(x)), \dots).$$

This is just the itinerary assignment from Milnor and Thurston's kneading theory[MT88, CE80]. Each point gets a sequence of numbers in $\{1, 2, \dots, p_m\}$ depending on the area of set X in which its iterates land. The image $\Xi(X)$ of X in the full shift is easily seen to be a shift space, a σ-invariant compact subset of the full shift. As any coarse graining maps periodic points to periodic points and since all points in X are periodic, by Lemma 5.4 we get that $\Xi(X)$ is a finite set. This further implies that all but finitely many points in the sequence $(x_{s(i)})_{i\in\mathbb{N}}$ are mapped to the same point as x which due to the construction is

$$\Xi(x) = (\overline{1, 2, \dots, p_m},\ \overline{1, 2, \dots, p_m},\ \dots),$$

and is clearly σ-periodic with period p_m. Therefore, by Lemma 5.5, p_m divides the period of all but finitely many points in $(x_{s(i)})_{i\in\mathbb{N}}$. This is not possible as by the construction p_m does not divide the period of $x_{s(i)}$ for all i large enough (it suffices to take $i \geq m$). This contradiction means that at least two different iterates of x must lie in the same connected component proving that not all the components of X can be degenerate. □

As we are closing this section we remark that this forms a small part of Sherman's characterisation and that it looks like the coarse graining approach alone is not powerful enough to recover his complete result.

In this section we give an example of a dynamical system on a compact metric space having shadowing property for which $\omega_f \neq ICT(f)$, thus answering negatively on a question raised in [Bar11, BDG12]. Independently, Gareth Davies found a similar example which remains unpublished.

We warn the reader that in order to stay consistent with most of the literature on the topic — the map providing dynamics in this section is denoted by f rather than T. The underlying space X is always assumed to be compact and metric.

Let (X, f) be a dynamical system. To gain insight about its long term behaviour it is useful to study the structure of its *limit sets* or *attractors*. The most natural limit set to consider is probably the ω-limit set of a point $x \in X$ which is defined as the set of limit points of its orbit sequence $(f^i(x))_{i\in\mathbb{N}_0}$, or equivalently by:

$$\omega_f(x) = \bigcap_{i=1}^{\infty} \overline{\bigcup_{j=i}^{\infty} \{f^j(x)\}},$$

where, as is usual, $f^j = \underbrace{f \circ \cdots \circ f}_{j \text{ times}}$ stands for the j-fold iteration of f. We shall write just $\omega(x)$ when the mapping is clear from the context.

It can be shown that each ω-limit set is a non-empty, closed, and f-invariant subset of X, see e.g. [BC92, Chapter IV]. In particular, this means that all of them belong to 2^X — the hyperspace of closed subsets of X. By

$$\omega_f = \{\omega_f(x) \mid x \in X\} \subseteq 2^X,$$

we denote the set of all ω-limit sets in the system.

It is also known (see again [BC92, Chapter IV]) that for any given point x the set $\omega(x)$ satisfies a slightly technical property known as weak incompressibility (WI) introduced by Šarkovs′kiĭ in [Šar65].

Definition 5.6 (Weak Incompressibility)**.** An f-invariant and closed set $A \subseteq X$ is said to be *weakly incompressible* if for any proper, non-empty closed subset $F \subset A$ we have

$$F \cap \overline{f(A \setminus F)} \neq \emptyset,$$

or equivalently, any non-empty subset $G \subseteq A$ relatively open in A that satisfies $f(\overline{G}) \subseteq G$ is necessarily A itself.

Barwell, Good, Oprocha, and Raines in [BGOR13] showed that WI is equivalent to another well-known notion, namely that of internal chain transitivity (ICT). But before we can define this we need to recall what pseudo-orbits are.

Definition 4.2 (Pseudo-orbit)**.** A sequence $\langle x_0, x_1, x_2, \dots \rangle$ is said to be a *δ-pseudo-orbit* for some $\delta > 0$ provided that $d(f(x_i), x_{i+1}) < \delta$ for each $i \in \mathbb{N}_0$. A *finite δ-pseudo-orbit of length $l \geq 1$* is a finite sequence $\langle x_0, x_1, x_2, \dots, x_l \rangle$ satisfying $d(f(x_i), x_{i+1}) < \delta$ for $0 \leq i < l$. We also say that it is a *δ-pseudo-orbit between x_0 and x_l*.

Definition 5.7 (Internal Chain Transitivity)**.** An f-invariant and closed set $A \subseteq X$ is said to be *internally chain transitive* if given any $\delta > 0$ there exists a δ-pseudo-orbit between any two points of A that is completely contained inside A. By $ICT(f)$ we denote the set containing all ICT subsets of X. This is again a subset of 2^X.

As we mentioned before, any ω-limit set is WI/ICT. We thus have the following inclusion of sets in the hyperspace 2^X:

$$\omega_f \subseteq ICT(f).$$

For some systems this is a strict inclusion and it is not hard to find such examples. It is much more interesting to try to characterise systems in which ω_f and $ICT(f)$ coincide. This would be useful as it is easier to check if a given set is ICT than if it is an ω-limit set.

Bowen in [Bow75] proved that $\omega_f = ICT(f)$ holds for Axiom A diffeomorphisms. In a series of papers [BDG12, BR15, Bar10, BGO12, BGKR10, BGOR13] Barwell, Davies, Good, Knight, Oprocha, and Raines prove that for many other classes of systems, they also coincide. These include shifts of finite type and Julia sets for certain quadratic maps. It turns out that they can prove this equality for an even larger class of systems under the assumption that the systems in question have the *shadowing property*.

Definition 5.8. A point $z \in X$ is said to *ε-shadow* a sequence $\langle x_0, x_1, x_2, \dots \rangle$ for some $\varepsilon > 0$ if $d(x_i, f^i(z)) < \varepsilon$ for each $i \in \mathbb{N}_0$.

Definition 5.9 (Shadowing)**.** A dynamical system $f : X \to X$ is said to have *shadowing* provided that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that every δ-pseudo-orbit is ε-shadowed by some point in X.

Barwell et al. posed a conjecture that under shadowing it is always $\omega_f = ICT(f)$. Not long ago, Meddaugh and Raines in [MR13] have proved that the set $ICT(f)$ must be closed in the space 2^X furnished with the Hausdorff distance. Assuming the shadowing condition they further show that closure of ω_f in 2^X is equal to $ICT(f)$:

$$\overline{\omega_f} = ICT(f).$$

This is a particularly nice result, because it has been known for some time that for any continuous map of the interval $f : [0, 1] \to [0, 1]$ the set ω_f is closed, i.e. $\overline{\omega_f} = \omega_f$, and so assuming that f has shadowing $\omega_f = ICT(f)$.

That ω_f is closed for interval maps was proved by Blokh, Bruckner, Humke, and Smítal in [BBHS96]. Pokluda then showed this for circle maps in [Pok02], and finally the general case of maps on any finite graph was covered by Mai and Shao in [MS07].

This meant that Meddaugh and Raines's result gave a positive resolution to the conjecture in all of these cases. It remained open if the shadowing property itself implies closedness of ω_f. Note that in the light of new results this is equivalent to Barwell's question.

We shall now show that the answer is no by constructing a system with shadowing in which $\omega_f \neq ICT(f)$. To do this we once again use shift spaces and methods of symbolic dynamics.

As before, let $\mathcal{A}$ be an alphabet — a finite discrete set of symbols. Recall that a *word* over $\mathcal{A}$ is a finite sequence of elements in $\mathcal{A}$. If one can find a finite collection of words W such that a shift space X is precisely the set of sequences in which none of the words from W appears, then that shift space is said to be of *finite type*. Walters in [Wal78] showed that shifts of finite type are precisely those shift spaces with the shadowing property.

Proposition 5.10 (Walters [Wal78])**.** *A shift space over a finite alphabet is of finite type if and only if it has the shadowing property.*

We can now proceed with the construction. We fix the alphabet $\mathcal{A} = \{0, 1\}$ and for each $k \in \mathbb{N}_0$ we set

$$X_k = \{\xi \in \Sigma_2 \mid \text{any two 1s in } \xi \text{ are separated with at least } (k+1) \text{ 0s}\},$$

and

$$X_\infty = \{\xi \in \Sigma_2 \mid \xi \text{ has at most one symbol 1}\}.$$

Note that each X_k is in fact a shift space of finite type where the set of forbidden words is exactly

$$\{1\underbrace{0\ldots0}_{l \text{ zeros}}1 \mid 0 \leq l \leq k\} = \{11, 101, 1001, \ldots, 1\underbrace{0\ldots0}_{k \text{ zeros}}1\}.$$

We also set $N = \{1/2^k \mid k \in \mathbb{N} \cup \{0, \infty\}\}$ where $1/2^\infty = 0$ by the convention. The topology on N is taken to be inherited from the real line, and the observant reader might realise that N and X_∞ are in fact homeomorphic. The space $N \times \Sigma_2$ is a compact metric space equipped with the max-distance:

$$d((a_1, \xi_1), (a_2, \xi_2)) = \max\{|a_1 - a_2|, d_{\Sigma_2}(\xi_1, \xi_2)\},$$

where d_{Σ_2} is the standard metric on the full shift Σ_2 defined in the previous section.

On $N \times \Sigma_2$ we define a continuous map f as the product of the identity on N and the shift map σ on Σ_2:

$$f(a, \xi) = (a, \sigma(\xi)).$$

This is easily seen continuous. Finally, we take

$$X = \{(a, \xi) \in N \times \Sigma_2 \mid a = \frac{1}{2^k} \text{ and } \xi \in X_k, \text{ for some } k \in \mathbb{N} \cup \{0, \infty\}\},$$

or equivalently

$$X = \{0\} \times X_\infty \cup \bigcup_{k=0}^{\infty} \{1/2^k\} \times X_k.$$

This is clearly an f-invariant subset of $N \times \Sigma_2$. We below show that X is also closed and that the map f restricted to X provides the counter-example we have been looking for.

Let us briefly describe the idea behind the construction. The map f on each space $\{1/2^k\} \times X_k$ is conjugated to a shift of finite type and hence, by Proposition 5.10, f has shadowing on those subspaces. The space $\{0\} \times X_\infty$ on the other hand is not of finite type and does not have shadowing. In the construction we exploit the fact that the sequence of spaces $(\{1/2^k\} \times X_k)_{k \in \mathbb{N}_0}$ converge to $\{0\} \times X_\infty$ in the hyperspace 2^X as $k \to \infty$. This allows us to shadow pseudo-orbits in the subspace $\{0\} \times X_\infty$ using real orbits in the space $\{1/2^k\} \times X_k$ for k large enough. In this way we succeeded (Lemma 5.12) to impose shadowing on f in the whole space by having shadowing on a family of proper subspaces approximating X in the limit.

It remains to be seen that for this system $\omega_f \neq ICT(f)$. The counter-example is the set $\{0\} \times X_\infty$ which is not an ω-limit set of any of the points in X (Theorem 5.13). Yet, it is the limit of the sequence of subspaces $\{1/2^k\} \times X_k$ as $k \to \infty$, each of which is an ω-limit set of a point in X. This, when combined with the result of Meddaugh and Raines, implies that $\{0\} \times X_\infty$ is an ICT set but not an ω-limit set. We shall now proceed with proving these claims.

Lemma 5.11. *X is a closed and hence a compact subset of $\mathcal{N} \times \Sigma_2$.*

Proof. Let (a, ξ) be a point in $\mathcal{N} \times \Sigma_2 \setminus X$. If $a = 1/2^k > 0$, this means that $\xi \in \Sigma_2 \setminus X_k$. Since X_k is closed in Σ_2 there is an open set V around ξ that does not intersect X_k. Since $U = \{a\}$ is an open (and closed) set in $\mathcal{N}$, $U \times V$ is an open neighbourhood containing (a, ξ) that does not intersect X.

If $a = 0$ and $\xi \notin X_\infty$ then there exists a $k \in \mathbb{N}_0$ so that word $1\underbrace{0 \ldots 0}_{k \text{ zeros}}1$ occurs somewhere in ξ. Take V to be the set of all 0-1 sequences in Σ_2 which have this word at the same position as ξ does. This set is easily seen to be clopen. It is indeed what is called a cylinder set in Σ_2 (see e.g. [LM95]). Setting $U = [0, 1/2^k) \cap \mathcal{N}$ one readily checks that $U \times V$ is an open neighbourhood containing (a, ξ) but not intersecting X. □

Lemma 5.12. *(X, f) has shadowing.*

Proof. Let $\varepsilon > 0$ and additionally assume $\varepsilon < 1$. Choose a $k \in \mathbb{N}$ so that $\varepsilon/2 \leq 1/2^k < \varepsilon$. Set $\delta = \min\{\varepsilon/4, \delta_1(\varepsilon), \ldots, \delta_k(\varepsilon)\} > 0$, where each $\delta_j(\varepsilon)$ for $1 \leq j \leq k$ is a positive number chosen so that every $\delta_j(\varepsilon)$-pseudo-orbit in X_j is ε-shadowed. This can be done by Proposition 5.10. We claim that for this δ, every δ-pseudo-orbit in X is ε-shadowed by a real orbit.

Let $\langle (a_n, \xi_n) \rangle_{n \in \mathbb{N}_0}$ be a δ-pseudo-orbit in X. We shall distinguish two cases.

Case 1. We first suppose that $a_0 > \varepsilon/2$. If $a_1 > a_0$ then $a_1 \geq 2a_0$ and hence $|a_1 - a_0| > \varepsilon/2 > \delta$. On the other hand, if $a_1 < a_0$ then $2a_1 \leq a_0$ and hence $|a_0 - a_1| \geq a_0/2 > \varepsilon/4 \geq \delta$. Therefore, it must be that $a_1 = a_0$ and inductively $a_n = a_0$ for all $n \in \mathbb{N}$. Which means that in this case the whole pseudo orbit is actually contained in the same subspace $\{1/2^m\} \times X_m$ where $a_0 = 1/2^m$.

Clearly $m \leq k$. Since $\delta \leq \delta_m(\varepsilon)$, we have that $\langle \xi_n \rangle_{n \in \mathbb{N}_0}$ is a $\delta_m(\varepsilon)$-pseudo-orbit in X_m, hence we can choose a point ξ^* that ε-shadows it. But then the point (a_0, ξ^*) clearly ε-shadows the initial pseudo-orbit.

Case 2. We now suppose $a_0 \leq \varepsilon/2$. A similar argument to the one above shows that $a_n \leq \varepsilon/2$, and hence $a_n \leq 1/2^k$ for all $n \in \mathbb{N}$. Since $(X_n)_{n \in \mathbb{N}_0}$ from a decreasing sequence of sets, each ξ_n is contained in the space X_k. The sequence $\langle \xi_n \rangle_{n \in \mathbb{N}_0}$ is a $\delta_k(\varepsilon)$-pseudo-orbit in X_k, hence there exists a point ξ^* that ε-shadows it. Again, it is readily checked that $(1/2^k, \xi^*)$ ε-shadows the initial pseudo-orbit. □

Theorem 5.13. *(X, f) is a dynamical system on a compact metric space which exhibits shadowing but for which $\omega_f \neq ICT(f)$.*

Proof. It suffices to note that $\{0\} \times X_\infty$ is an ICT set that is not an ω-limit set of any of the points in X. If it were an ω-limit set of some point $(a, \xi) \in X$, it would have to be that $a = 0$. But it is not hard to see that the ω-limit set of any point in $\{0\} \times X_\infty$ is the singleton $\{(0, 0^\infty)\}$ as they are all pre-fixed points. Here by 0^∞ we denote the sequence in Σ_2 consisting only of zeros. Therefore the set $\{0\} \times X_\infty$ is not in ω_f.

It remains to be shown that $\{0\} \times X_\infty$ is in $ICT(f)$. To simplify notation we shall instead show that the set X_∞ is ICT under the shift map σ. This is clearly an equivalent statement. Let $\delta > 0$ and let ξ and η be any two points in X_∞. We can always choose $k \in \mathbb{N}$ such that $\sigma^k(\xi) = 0^\infty$. If $\eta = 0^\infty$ we are done as

$$\langle \xi, \sigma(\xi), \ldots, \sigma^k(\xi) = \eta \rangle$$

is a δ-pseudo-orbit between ξ and η.

If otherwise $\eta = 0^m 10^\infty$ for some $m \in \mathbb{N}_0$, choose $n > m$ large enough so that the point $\zeta = 0^n 10^\infty$ is δ-close to 0^∞. Then one can check that

$$\langle \xi, \sigma(\xi), \ldots, \sigma^k(\xi) = 0^\infty, \zeta, \sigma(\zeta), \ldots, \sigma^{n-m}(\zeta) = \eta \rangle$$

is a δ-pseudo-orbit between ξ and η. □

We mentioned already that combining this theorem with Meddaugh and Raines's result we obtain the following corollary.

Corollary 5.14. *(X, f) is a dynamical system on a compact metric space which exhibits shadowing but for which ω_f is not closed in the hyperspace of compact subsets 2^X.*

We remark that Barwell's results in particular show that (X, f) cannot be conjugated to any shift system.

Related to these results, Good and Meddaugh in [GM16] very recently obtained a characterisation of systems in which $\omega_f = ICT(f)$. They show that this is equivalent to another technical property named *orbital limit shadowing*, a definition of which can be found in their paper. Our example thus shows that a system can exhibit shadowing without having orbital limit shadowing.

LIST OF REFERENCES

[AEG15] Massoud Amini, George A. Elliott, and Nasser Golestani, *The category of Bratteli diagrams*, Canad. J. Math. **67** (2015), no. 5, 990–1023. MR3391730

[AGW08] Ethan Akin, Eli Glasner, and Benjamin Weiss, *Generically there is but one self homeomorphism of the Cantor set*, Trans. Amer. Math. Soc. **360** (2008), no. 7, 3613–3630. MR2386239

[Ban05] John Banks, *Chaos for induced hyperspace maps*, Chaos Solitons Fractals **25** (2005), no. 3, 681–685. MR2132366 (2006a:37013)

[Bar10] Andrew D. Barwell, *A characterization of ω-limit sets of piecewise monotone maps of the interval*, Fund. Math. **207** (2010), no. 2, 161–174. MR2586009 (2011k:37059)

[Bar11] Andrew Barwell, *Omega-limit sets of discrete dynamical systems*, Ph.D. Thesis, University of Birmingham, 2011.

[BBHS96] Alexander Blokh, A. M. Bruckner, P. D. Humke, and J. Smítal, *The space of ω-limit sets of a continuous map of the interval*, Trans. Amer. Math. Soc. **348** (1996), no. 4, 1357–1372. MR1348857 (96j:58089)

[BC86] Louis Block and Ethan M. Coven, *Maps of the interval with every point chain recurrent*, Proc. Amer. Math. Soc. **98** (1986), no. 3, 513–515. MR857952 (87j:54059)

[BC92] L. S. Block and W. A. Coppel, *Dynamics in one dimension*, Lecture Notes in Mathematics, vol. 1513, Springer-Verlag, Berlin, 1992. MR1176513 (93g:58091)

[BD12] Nilson C. Bernardes Jr. and Udayan B. Darji, *Graph theoretic structure of maps of the Cantor space*, Adv. Math. **231** (2012), no. 3-4, 1655–1680. MR2964619

[BDG12] Andrew D. Barwell, Gareth Davies, and Chris Good, *On the ω-limit sets of tent maps*, Fund. Math. **217** (2012), no. 1, 35–54. MR2914921

[BGKR10] Andrew Barwell, Chris Good, Robin Knight, and Brian E. Raines, *A characterization of ω-limit sets in shift spaces*, Ergodic Theory Dynam. Systems **30** (2010), no. 1, 21–31. MR2586343 (2011e:37026)

[BGL13] Francisco Balibrea, Juan L. G. Guirao, and Marek Lampart, *A note on the definition of α-limit set*, Appl. Math. Inf. Sci. **7** (2013), no. 5, 1929–1932. MR3072393

[BGO12] Andrew D. Barwell, Chris Good, and Piotr Oprocha, *Shadowing and expansivity in subspaces*, Fund. Math. **219** (2012), no. 3, 223–243. MR3001240

[BGOR13] Andrew D. Barwell, Chris Good, Piotr Oprocha, and Brian E. Raines, *Characterizations of ω-limit sets in topologically hyperbolic systems*, Discrete Contin. Dyn. Syst. **33** (2013), no. 5, 1819–1833. MR3002729

[BLT03] F. Thomas Bruss, Guy Louchard, and John W. Turner, *On the N-tower problem and related problems*, Adv. in Appl. Probab. **35** (2003), no. 1, 278–294. In honor of Joseph Mecke. MR1975514 (2004d:60181)

[Bow75] Rufus Bowen, *ω-limit sets for axiom* A *diffeomorphisms*, J. Differential Equations **18** (1975), no. 2, 333–339. MR0413181

[BR15] Andrew D. Barwell and Brian E. Raines, *The ω-limit sets of quadratic Julia sets*, Ergodic Theory Dynam. Systems **35** (2015), no. 2, 337–358. MR3316915

[BS75] Walter Bauer and Karl Sigmund, *Topological dynamics of transformations induced on the space of probability measures*, Monatsh. Math. **79** (1975), 81–92. MR0370540 (51 #6767)

[Bur07] Keki Burjorjee, *Sufficient conditions for coarse-graining evolutionary dynamics*, Foundations of genetic algorithms, 2007, pp. 35–53.

[CE80] Pierre Collet and Jean-Pierre Eckmann, *Iterated maps on the interval as dynamical systems*, Progress in Physics, vol. 1, Birkhäuser, Boston, Mass., 1980. MR613981

[CGH+14] Chris Cannings, Chris Good, John Haslegrave, David Parker, Mate Puljiz, and Jonathan E. Rowe, *WP2: Invariance properties of sub-systems — D2.1: Theory of topological and dynamical invariants* (2014), available at http://www.hieratic.eu/deliverables.php. Accessed: 11/08/2016.

[CGH+15] ———, *WP2: Invariance properties of sub-systems — D2.2: Combination of complex sub-systems* (2015), available at http://www.hieratic.eu/deliverables.php. Accessed: 11/08/2016.

[Dan01] Alexandre I. Danilenko, *Strong orbit equivalence of locally compact Cantor minimal systems*, Internat. J. Math. **12** (2001), no. 1, 113–123. MR1812067 (2002j:37016)

[DDF07] Peter Dittrich and Pietro Speroni Di Fenizio, *Chemical organisation theory*, Bulletin of mathematical biology **69** (2007), no. 4, 1199–1231.

[DHH+15] Peter Dittrich, Jan Huwald, Richard Henze, Bashar Ibrahim, and Mate Puljiz, *WP4: Multi-scale simulation — D4.2: Integration of coarse-graining algorithms* (2015), available at http://www.hieratic.eu/deliverables.php. Accessed: 11/08/2016.

[DHS03] S. Derisavi, H. Hermanns, and W. Sanders, *Optimal state-space lumping in Markov chains*, Information Processing Letters **87** (September 2003), no. 6, 309–315.

[Doo03] Anthony H. Dooley, *Markov odometers*, Topics in dynamics and ergodic theory, 2003, pp. 60–80. MR2052275 (2005c:37007)

[Eng89] Ryszard Engelking, *General topology*, Second, Sigma Series in Pure Mathematics, vol. 6, Heldermann Verlag, Berlin, 1989. Translated from the Polish by the author. MR1039321

[Eng93] Arthur Engel, *The computer solves the three tower problem*, The American Mathematical Monthly **100** (1993), no. 1, pp. 62–64.

[ES09] M. A. El-Shehawey, *On the gambler's ruin problem for a finite Markov chain*, Statist. Probab. Lett. **79** (2009), no. 14, 1590–1595. MR2547416 (2010i:60133)

[FB94] Walter Fontana and Leo W Buss, *The arrival of the fittest: Toward a theory of biological organization*, Bulletin of Mathematical Biology **56** (1994), no. 1, 1–64.

[Fel68] William Feller, *An introduction to probability theory and its applications. Vol. I*, Third edition, John Wiley & Sons, Inc., New York-London-Sydney, 1968. MR0228020 (37 #3604)

[FG16] Leobardo Fernández and Chris Good, *Shadowing for induced maps of hyperspaces*, Fund. Math. (2016), to appear.

[FGH+16] Leobardo Fernández, Chris Good, John Haslegrave, Mate Puljiz, and Ártico Ramírez, *WP2: Invariance properties of sub-systems — D2.3: Invariance properties of sub-systems* (2016), available at http://www.hieratic.eu/deliverables.php. Accessed: 11/08/2016.

[FGP16a] Leobardo Fernández, Chris Good, and Mate Puljiz, *Almost minimal systems and periodicity in hyperspaces* (2016), available at http://arxiv.org/abs/1602.03127. Accessed: 16/8/2016.

[FGP16b] ———, *WP1: Theory of hierarchical structure — D1.3: Theory of hierarchical structure* (2016), available at http://www.hieratic.eu/deliverables.php. Accessed: 11/08/2016.

[FGPR15] Leobardo Fernández, Chris Good, Mate Puljiz, and Ártico Ramírez, *Chain transitivity in hyperspaces*, Chaos Solitons Fractals **81** (2015), no. part A, 83–90. MR3426021

[GGK+06] C. Good, S. Greenwood, R. W. Knight, D. W. McIntyre, and S. Watson, *Characterizing continuous functions on compact spaces*, Adv. Math. **206** (2006), no. 2, 695–728. MR2263719 (2007i:54027)

[GJP+14] Chris Good, Martin Nilsson Jacobi, David Parker, Mate Puljiz, and Jonathan E. Rowe, *WP3: Algorithms for non-linear decompositions — D3.1: Algorithms for identifying linear coarse grainings* (2014), available at http://www.hieratic.eu/deliverables.php. Accessed: 11/08/2016.

[GJP+15] ———, *WP3: Algorithms for non-linear decompositions — D3.2: Algorithms for identifying non-linear and alternative coarse-grainings of non-linear systems* (2015), available at http://www.hieratic.eu/deliverables.php. Accessed: 11/08/2016.

[GKK+06] Aleksandr Nikolaevich Gorban, Nikolas K Kazantzis, Ioannis G Kevrekidis, Hans Christian Öttinger, and Constantinos Theodoropoulos, *Model reduction and coarse-graining approaches for multiscale phenomena*, Springer, 2006.

[GKL+09] Juan Luis García Guirao, Dominik Kwietniak, Marek Lampart, Piotr Oprocha, and Alfredo Peris, *Chaos on hyperspaces*, Nonlinear Anal. **71** (2009), no. 1-2, 1–8. MR2518006 (2010c:37032)

[GM06] Jean-Marc Gambaudo and Marco Martens, *Algebraic topology for minimal Cantor sets*, Ann. Henri Poincaré **7** (2006), no. 3, 423–446. MR2226743 (2006m:37007)

[GM16] Chris Good and Jonathan Meddaugh, *Orbital shadowing, internal chain transitivity and ω-limit sets*, Ergodic Theory Dynam. Systems (2016).

[GPPR14] Chris Good, David Parker, Mate Puljiz, and Jonathan E. Rowe, *WP1: Theory of hierarchical structure — D1.1: Algebraic structure of hierarchical decompositions* (2014), available at `http://www.hieratic.eu/deliverables.php`. Accessed: 11/08/2016.

[GPPR15] ———, *WP1: Theory of hierarchical structure — D1.2: Topological characterisation of property convergence* (2015), available at `http://www.hieratic.eu/deliverables.php`. Accessed: 11/08/2016.

[GPPR16] ———, *A characterisation of compatible state space aggregations for discrete dynamical systems* (2016), available at `http://matepuljiz.tk`. Accessed: 16/8/2016.

[GRIM12] José L. Gómez-Rueda, Alejandro Illanes, and Héctor Méndez, *Dynamic properties for the induced maps in the symmetric products*, Chaos Solitons Fractals **45** (2012), no. 9-10, 1180–1187. MR2979228

[Gut13] Allan Gut, *The gambler's ruin problem with delays*, Statist. Probab. Lett. **83** (2013), no. 11, 2549–2552. MR3144040

[HP17] John Haslegrave and Mate Puljiz, *Reaching consensus on a connected graph*, J. Appl. Probab. **54** (2017), no. 1, to appear, available at `http://arxiv.org/abs/1511.05435`. Accessed: 16/8/2016.

[HPS92] Richard H. Herman, Ian F. Putnam, and Christian F. Skau, *Ordered Bratteli diagrams, dimension groups and topological dynamics*, Internat. J. Math. **3** (1992), no. 6, 827–864. MR1194074 (94f:46096)

[IM12] W. T. Ingram and William S. Mahavier, *Inverse limits*, Developments in Mathematics, vol. 25, Springer, New York, 2012. From continua to chaos. MR3014043

[Jac05] Martin N. Jacobi, *Hierarchical organization in smooth dynamical systems*, Artif. Life **11** (December 2005), no. 4, 493–512.

[JG09] Martin Nilsson Jacobi and Olof Görnerup, *A spectral method for aggregating variables in linear dynamical systems with application to cellular automata renormalization*, Advances in Complex Systems **12** (2009), no. 02, 131–155.

[Kar72] Richard M. Karp, *Reducibility among combinatorial problems*, Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972), 1972, pp. 85–103. MR0378476

[Kat13] Guy Katriel, *Gambler's ruin probability—a general formula*, Statist. Probab. Lett. **83** (2013), no. 10, 2205–2210. MR3093803

[Kat14] ———, *Gambler's ruin: the duration of play*, Stoch. Models **30** (2014), no. 3, 251–271. MR3238565

[KNP11] Marta Kwiatkowska, Gethin Norman, and David Parker, *PRISM 4.0: verification of probabilistic real-time systems*, Computer aided verification, 2011, pp. 585–591. MR2870782

[KO07] Dominik Kwietniak and Piotr Oprocha, *Topological entropy and chaos for maps induced on hyperspaces*, Chaos Solitons Fractals **33** (2007), no. 1, 76–86. MR2301847 (2008b:37024)

[KP02] Andrej Kmet and Marko Petkovšek, *Gambler's ruin problem in several dimensions*, Adv. in Appl. Math. **28** (2002), no. 2, 107–118. MR1888839 (2003a:60012)

[KS76] John G. Kemeny and J. Laurie Snell, *Finite Markov chains*, Springer-Verlag, New York-Heidelberg, 1976. Reprinting of the 1960 original, Undergraduate Texts in Mathematics. MR0410929 (53 #14670)

[Len09] Tamás Lengyel, *The conditional gambler's ruin problem with ties allowed*, Appl. Math. Lett. **22** (2009), no. 3, 351–355. MR2483497 (2010b:60125)

[LM95] Douglas Lind and Brian Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, Cambridge, 1995. MR1369092

[LR89] Genyuan Li and Herschel Rabitz, *A general analysis of exact lumping in chemical kinetics*, Chemical engineering science **44** (1989), no. 6, 1413–1430.

[LR90] ———, *A general analysis of approximate lumping in chemical kinetics*, Chemical engineering science **45** (1990), no. 4, 977–1002.

[LRT94] Genyuan Li, Herschel Rabitz, and János Tóth, *A general analysis of exact nonlinear lumping in chemical kinetics*, Chemical Engineering Science **49** (1994), no. 3, 343–361.

[Mac05] Sergio Macías, *Topics on continua*, Chapman & Hall/CRC, Boca Raton, FL, 2005. MR2147759

[Mio63] J. Mioduszewski, *Mappings of inverse limits*, Colloq. Math. **10** (1963), 39–44. MR0166762

[MR13] Jonathan Meddaugh and Brian E. Raines, *Shadowing and internal chain transitivity*, Fund. Math. **222** (2013), no. 3, 279–287. MR3104074

[MS07] Jie-Hua Mai and Song Shao, *Spaces of ω-limit sets of graph maps*, Fund. Math. **196** (2007), no. 1, 91–100. MR2338540 (2008h:37043)

[MT88] John Milnor and William Thurston, *On iterated maps of the interval*, Dynamical systems (College Park, MD, 1986–87), 1988, pp. 465–563. MR970571

[Nad06] Sam B. Nadler Jr., *Hyperspaces of sets. A text with research questions*, Aportaciones Matemáticas: Textos [Mathematical Contributions: Texts], vol. 33, Sociedad Matemática Mexicana, México, 2006. Unabridged edition of the 1978 original. MR2293338 (2007i:54024)

[Nag70] Keiô Nagami, *Dimension theory*, With an appendix by Yukihiro Kodama. Pure and Applied Mathematics, Vol. 37, Academic Press, New York-London, 1970. MR0271918 (42 #6799)

[Par99] Emanuel Parzen, *Stochastic processes*, Classics in Applied Mathematics, vol. 24, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999. Reprint of the 1962 original. MR1699272

[Pok02] David Pokluda, *Characterization of ω-limit sets of continuous maps of the circle*, Comment. Math. Univ. Carolin. **43** (2002), no. 3, 575–581. MR1920533 (2003d:37048)

[PT87] Robert Paige and Robert E. Tarjan, *Three partition refinement algorithms*, SIAM J. Comput. **16** (1987), no. 6, 973–989. MR917035 (89h:68069)

[RVW05a] Jonathan E. Rowe, Michael D. Vose, and Alden H. Wright, *Coarse graining selection and mutation*, Foundations of genetic algorithms, 2005, pp. 176–191. MR2238251

[RVW05b] ———, *State aggregation and population dynamics in linear systems*, Artif. Life **11** (December 2005), no. 4, 473–492.

[RVW06] ———, *Differentiable coarse graining*, Theoret. Comput. Sci. **361** (2006), no. 1, 111–129. MR2254227 (2007h:68171)

[Šar65] O. M. Šarkovs′kiĭ, *On attracting and attracted sets*, Dokl. Akad. Nauk SSSR **160** (1965), 1036–1038. MR0188992 (32 #6419)

[SB06] Yvik C. Swan and F. Thomas Bruss, *A matrix-analytic approach to the N-player ruin problem*, J. Appl. Probab. **43** (2006), no. 3, 755–766. MR2274798 (2007m:60221)

[She12] Casey Sherman, *Orbit structures of homeomorphisms*, Ph.D. Thesis, Baylor University, 2012.

[Shi14] Takashi Shimomura, *Special homeomorphisms and approximation for Cantor systems*, Topology Appl. **161** (2014), 178–195. MR3132360

[Shi16a] ———, *A combinatorial construction of residually scrambled symbolic systems* (2016), available at http://arxiv.org/abs/1602.01568. Accessed: 9/2/2016.

[Shi16b] ———, *The construction of a completely scrambled system by graph covers*, Proc. Amer. Math. Soc. **144** (2016), no. 5, 2109–2120. MR3460171

[Sti06] David Stirzaker, *Three-handed gambler's ruin*, Adv. in Appl. Probab. **38** (2006), no. 1, 284–286. MR2213975 (2006k:60073)

[Sti94] ———, *Tower problems and martingales*, Math. Sci. **19** (1994), no. 1, 52–59. MR1294785 (95h:60068)

[TLRT97] Alison S Tomlin, Genyuan Li, Herschel Rabitz, and János Tóth, *The effect of lumping and expanding on kinetic differential equations*, SIAM Journal on Applied Mathematics **57** (1997), no. 6, 1531–1556.

[TW06] John Talbot and Dominic Welsh, *Complexity and cryptography*, Cambridge University Press, Cambridge, 2006. An introduction. MR2221458

[Vos99] Michael D. Vose, *The simple genetic algorithm*, Complex Adaptive Systems, MIT Press, Cambridge, MA, 1999. Foundations and theory, A Bradford Book. MR1713436 (2000h:65024)

[Wal78] Peter Walters, *On the pseudo-orbit tracing property and its relationship to stability*, The structure of attractors in dynamical systems (Proc. Conf., North Dakota State Univ., Fargo, N.D., 1977), 1978, pp. 231–244. MR518563 (80d:58055)

[Wil91] David Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991. MR1155402 (93d:60002)