1

**PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT**

PRECISIONSM RESOLUTION TARGETS

PIONEERS IN METHYLENE BLUE TESTING SINCE 1974

1800 COUNTRY ROAD 5 BURNINGWELL, WY 82407 USA
TEL 812 426 7887 FAX 812 426 7887 TLA 940009400

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ON FRACTALS, CHAOS AND THE HAILSTONE NUMBERS

by

ROBERT D. BLACK, B.MATH

A thesis submitted to
the faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of

Master of Computer Science

**School of Computer Science
Carleton University
Ottawa, Ontario**

16 April 1992

© copyright 1992, Robert D. Black

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-75982-8

Canada

The undersigned recommend to the Faculty of Graduate Studies
and Research acceptance of the thesis

On Fractals, Chaos and the Hailstone Numbers

submitted by Robert D. Black, B.Math
in partial fulfilment of the requirements for
the degree of Master of Computer Science

A handwritten signature in cursive script, appearing to read "Nicola Pirk", written over a horizontal line.

Thesis Supervisor

A handwritten signature in cursive script, appearing to read "J. Falk", written over a horizontal line.

Chair, School of Computer Science

Carleton University

April 16, 1992

Abstract

In the last twenty years it has been widely recognized that many deterministic mathematical systems can exhibit *chaotic* behaviour, wherein small uncertainties in the system state grow at a roughly exponential rate, making accurate long-term predictions impossible. Chaotic systems often possess a geometry that is *fractal* in nature, having complex structure across all scales. Several definitions of chaos exist, all of which are defined for continuous-space systems. In this thesis we review the current theory of chaos and fractals, and report on claims regarding the existence of chaos in a certain discrete-space dynamical system over the natural numbers. We conclude by considering whether fractals and chaos can exist within discrete-space systems, and give recommendations for further research in this area.

For my Parents:
Robert and Christina Black

Acknowledgements

I would like to thank my supervisor Nicola Santoro for his guidance and friendship these last few years, and for many hours spent in enlightening discussion. From him I have learned that good research demands an attention to detail, an ability to write well, and an adventurous spirit.

My thanks to Angelo Mingarelli for his encouragement and guidance into the deeper aspects of this work. His careful review of the thesis and suggestions for improving the mathematics are very much appreciated. I would also like to thank the other members of the Examining Committee: Sivarama Dandamundi, Frank Fiala, and Jeff Sidney, for useful feedback during the defense.

Thanks go to Kevin McGuire, for many enjoyable discussions and especially for help unlooked-for during the final days. He helped to improve many parts of the thesis and gave me energy when I needed it most. Thanks also to Vincenzo Acciaro for his help with questions of number theory, and for suggesting useful books on the area. I am also grateful to Ralph Nevins and Patrick Hayden for sharing the results of numerous experiments on the computer.

Finally, my love and thanks to Kira for providing me with "sanity checks" along the way, help with the final draft, and most of all for love and friendship throughout.

Table of Contents

Abstract	iii
Acknowledgements	v
Table of Contents	vi
List of Figures	viii
1. INTRODUCTION	1
1.1 Motivation	1
1.2 Research results	3
1.3 Organization	5
PART I: OVERVIEW OF FRACTALS AND CHAOS	8
2. Fractals and dimension	9
2.1 Motivation	9
2.2 Historical Background	10
2.3 On measuring the size of fractal sets	12
2.4 Euclidean dimension	32
2.5 Topological dimension	34
2.6 Hausdorff dimension	40
2.7 Similarity dimension	44
2.8 Box-counting dimension	49
2.9 Correlation dimension	57
2.10 Information dimension	59
2.11 Summary	60
3. Chaos	62
3.1 Motivation	62
3.2 Historical Background	62
3.3 Dynamical systems	65
3.4 Sensitive dependence	71
3.5 Topological chaos	79
3.6 Symbolic dynamics	84
3.7 Formal chaos	93
3.8 Sarkovsky's theorem	95
3.9 Li/Yorke Chaos	96
3.10 Strange attractors	99
3.11 Summary	109

PART II: THE HAILSTONE NUMBERS	113
4. The Hailstone Function	114
4.1 Motivation	114
4.2 A dynamical system on the natural numbers	114
4.3 The $3x+1$ Conjecture	116
4.4 Is the $3x+1$ Conjecture Undecidable?	124
4.5 A Relationship with the Fibonacci Numbers	128
4.6 On a definition of chaos	135
4.7 Hailstone sequences and the Sarkovsky ordering	137
4.8 Summary	144
5. The Hailstone Map	145
5.1 Motivation	145
5.2 A dynamical system on the unit interval	145
5.3 The arrowhead fractal	155
5.4 Summary	166
PART III: SEARCHING FOR DISCRETE CHAOS	167
6. Conclusions	168
6.1 On the theory of fractals and chaos	168
6.2 On the hailstone numbers	170
6.3 On fractals and chaos in fully discrete systems	173
6.4 Recommendations for further research	176
APPENDIX A: Concepts from Metric Topology	178
A.1 Metric spaces	178
A.2 Compact metric spaces	181
A.3 Lebesgue measure	183
A.4 Continuous functions	186
A.5 Notes on dimension	187
REFERENCES	191

List of Figures

Figure 2.1	The Triadic Cantor Set	14
Figure 2.2	Peano curves for levels 1,2,3 and 4	21
Figure 2.3	von Koch snowflake curves for levels 0,1,2 and 3	23
Figure 2.4	The limiting von Koch snowflake curve	24
Figure 2.5	The Sierpinski triangle	29
Figure 3.1	Lorenz attractor projected onto the X-Y and Y-Z planes	106
Figure 3.2	The Lorenz attractor projected into the X-Z plane	107
Figure 3.3	Reconstruction of the Lorenz attractor	108
Figure 4.1	Hailstone sequence plot for $X_0=25$	118
Figure 4.2	Hailstone sequence plot for $X_0=26$	118
Figure 4.3	Hailstone sequence plot for $X_0=27$	119
Figure 4.4	Total stopping times for $X \leq 1000$	123
Figure 4.5	First iterate of the hailstone function	129
Figure 4.6	Second iterate of the hailstone function	129
Figure 4.7	Third iterate of the hailstone function	130
Figure 4.8	Effect of the hailstone function on the Sarkovsky sequence	140
Figure 5.1	Hailstone map, first iterate	151
Figure 5.2	The arrowhead fractal	157
Figure 5.3	Invariant set of the Arrowhead IFS	159

1. INTRODUCTION

In the last twenty years it has been widely recognized that simple mathematical models can exhibit a particular kind of complicated, random-like behaviour now known as *deterministic chaos*. More surprisingly, it has been discovered that the same chaotic behaviour observed in mathematical systems could also be observed in the physical systems they represent. Closely related to the theory of chaos is *fractal geometry*, which is the study of recursive geometric figures that exhibit a non-trivial structure at all scales. It is very often the case that chaotic systems have geometric properties that are fractal in nature.

In this thesis we examine the theory of fractal geometry and chaotic dynamical systems, and investigate claims regarding the existence of chaos in a certain number theoretic system known as the *hailstone numbers*.

1.1 Motivation

The study of chaos represents a special part of the wider area of *dynamical systems theory*, which is concerned with the long-term behaviour of mathematical models of a certain kind. Within dynamical systems theory there are two broad categories:

continuous dynamical systems which model time as a continuous flow, and *discrete dynamical systems* which model time as a discrete sequence of steps. Both types of systems have been successfully used to model a wide range of physical phenomena, and both are capable of exhibiting chaotic behaviour.

Although continuous and discrete dynamical systems differ in how they model time, they model space in the same way: as a continuum. This is an interesting fact, since the physical systems that they model are inherently discrete, being composed of large numbers of isolated particles. For example, population dynamics are often modelled using continuous-space maps, even though there can only be an integer number of individuals in a population. Similarly, the motion of a fluid is modelled using systems of differential equations, even though the fluid itself is composed of isolated particles.

In recent years there has been significant interest in dynamical systems that are discrete in both time and space; we refer to these as *fully discrete dynamical systems*. Two familiar examples of fully discrete dynamical systems are Turing machines [12, 30] and cellular automata [44, 167].

Although chaos is currently defined only for continuous-space systems [7, 34, 94, 159], it has been reported in Turing machines [20, 106], cellular automata [25, 57, 83, 90, 140, 146, 166], and pseudo-random number generators [76]. It is the fact that chaos is defined only for continuous-space systems, together with the existence of claims that it

exists in fully discrete systems which motivates the following

OBJECTIVES:

1. *To present a synthesis of the literature on the existing theory of fractals and chaos, and;*
2. *To investigate claims regarding the existence of chaos in a particular fully discrete system, known as the "hailstone numbers".*

1.2 Research results

We have found that, although the original definition of *fractal* due to Mandelbrot [100] applied only to uncountable sets, its usage has broadened to include discrete (countable) sets as well. In fact, even certain finite sets can possess "fractal" properties across a finite range of scales.

The many definitions of *fractal dimension* can be reduced to the following. The fractal dimension of a set is a real number that indicates how the "size" of the set changes with the scale at which it is measured. Since there are many ways to measure the size of a set, there are correspondingly many ways to measure its dimension. For example, Hausdorff dimension is the most useful in theoretical settings, whereas box-counting and correlation dimensions are most useful in practice.

Sensitive dependence on initial conditions is a central concept in chaos theory, wherein any uncertainty in the initial state of a system grows at a roughly exponential rate as the system evolves, making accurate long-term predictions impossible. Because sensitive dependence can be directly observed in physical systems, many authors consider this to be a sufficient definition of chaos. Some authors impose additional restrictions on the definition, usually with regard to the periodic structure of the system. Still others require that a chaotic system be dynamically equivalent to a "shift map", whose properties are well-known. Unfortunately, there is still no general agreement on the definition of chaos.

It is often the case that the evolution of a chaotic system traces out a fractal pattern commonly known as a *strange attractor*, which results from the repeated "stretching and folding" action of the system. Although the existence of a strange attractor usually indicates the presence of chaos, this is not always the case.

Our investigation into the hailstone numbers revealed that the claim by Agnes and Rasetti [1] that the hailstone numbers are chaotic, is in fact false. In particular, we show that the hailstone numbers do not act as a shift map upon Sarkovsky's ordering of the natural numbers. Agnes and Rasetti also defined a functionally equivalent one-dimensional map on the unit interval, and claimed that this map was chaotic and had fractal orbits. We show these claims to be false.

During our investigations we also examined the recently made claim that the $3x+1$ problem, which is an open problem related to the hailstone numbers, is undecidable [1]. We showed that the arguments used to support the claim are incorrect.

In an original result, we establish a relationship between the hailstone numbers and the Fibonacci numbers, determining the number of ways in which integers can be transformed by higher iterates of the hailstone function. Another original result establishes a lower bound on the "total stopping time" function, which determines the minimum number of steps required in a sequence of hailstone numbers until the value 1 is reached. Although these results are not directly related to the question of chaos in the hailstone numbers, they may prove useful in future research.

1.3 Organization

Our review of the literature represents a synthesis of the major concepts in the theory of fractals and chaos. Whenever possible, we provide historical background to the ideas being discussed, and include derivations of the relevant mathematical formulas. We reproduce a number of existing results from fractal geometry, often providing our own proofs in order to reinforce the techniques involved. Additionally, we present the various definitions of fractal dimension and chaos and put them into the proper context. This in itself is a significant contribution, since most sources present only one definition of each.

In Chapter 2: *Fractals and Dimension* we present the most important definitions of fractal dimension, and give several examples to illustrate the properties of fractal sets. Where possible, we derive these definitions from first principles to provide the appropriate motivation. In order to reinforce the techniques being discussed, we often provide our own proofs to results that are standard in the literature (or can be attributed to particular authors). We use the dagger symbol ($\dagger$) to identify results for which we have provided our own proof.

In Chapter 3: *Chaos* we present several definitions of chaos, and show how they are related. In addition, we discuss *sensitive dependence on initial conditions* as the cause of unpredictability in a dynamical system, *topological conjugacy* as the means by which different chaotic systems can be shown to be equivalent, and *symbolic dynamics* as a qualitative technique for establishing the dynamical properties of a chaotic system. We also discuss the meaning of *strange attractors* in the context of chaotic systems.

Chapters 4 and 5 present our case study of the fully discrete dynamical system commonly known as the hailstone numbers [72, 89]. This system was selected for examination because of the existence of specific claims that the system is chaotic [1]. In Chapter 4: *The Hailstone Function*, we present the basic known properties of the system in order to provide a context for the discussion to follow. In this regard we contribute to the existing knowledge by establishing a relationship between the hailstone numbers and the Fibonacci numbers. We also establish a lower bound on a certain

function related to the transient length of hailstone sequences.

In Chapter 5: *The Hailstone Map*, we discuss a one-dimensional map that is functionally equivalent to the hailstone function defined in the previous chapter. We examine the claims regarding the existence of chaos in the hailstone map, and show them to be false.

The thesis concludes with a summary of the important aspects of the theory of fractals and chaos. Following this we discuss the elements of the theory that can be related to fully discrete systems, and present our conclusions regarding the existence of chaos in the hailstone numbers. Based on what was learned from the literature and from the case study, we also present a number of promising avenues for future research.

Finally, the Appendix provides a brief summary of the various concepts from metric topology that appear in the thesis.

PART I: OVERVIEW OF FRACTALS AND CHAOS

2. Fractals and dimension

2.1 Motivation

In this chapter we present the central concepts behind the notions of *fractal sets* and *fractal dimension*. We attempt to provide historical context, to emphasize that fractal sets are not a recent discovery (although they have only recently been called by that name), and that the theory of fractional dimensions has existed for several decades.

The definitions of set dimension discussed here fall into four general categories, according to whether they are primarily based on

- the number of coordinates required to specify points in the set;
- the extent to which a set is connected;
- the type of symmetry exhibited by the set, or;
- a measure of the size of a set.

Most definitions are measure-based, and as many ways as there are to measure the *size* of a set, there are ways to measure its dimension. We illustrate these ideas with several well-known examples of fractal sets.

2.2 Historical Background

The term *fractal* was invented by Benoit Mandelbrot in the 1970's to describe certain mathematical sets of a highly irregular nature [99, 100]. Mandelbrot derived the word *fractal* from the latin *fractus*, meaning "broken"; the examples appearing later will show this to be an appropriate choice of terms.

Mandelbrot defined a fractal formally as "a set for which the Hausdorff Besicovitch dimension strictly exceeds the topological dimension". Hausdorff dimension (as it is usually called) was developed at the turn of the century in order to measure sets which cannot be adequately described using the usual notions of length, area, and volume. The Hausdorff dimension of a set can have non-integer values, whereas the topological dimension is always an integer. Hence, according to Mandelbrot's definition, a fractal is any set for which the Hausdorff dimension is not an integer.

Fractals are most commonly defined in the literature as sets that have fractional dimension. There are many ways to define the dimension of a set; Hausdorff dimension, various box-counting dimensions, packing dimensions, topological dimension and Euclidean dimension to name a few. Because these set dimensions are generally not equivalent, the definition of *fractal* has become somewhat muddled. In [43], Falconer writes:

Inconsistent usage has sometimes led to considerable confusion. In particular, warning lights flash in my mind (as in the minds of other mathematicians) whenever the term 'fractal dimension' is seen. Though some authors attach a precise meaning to this, I have known others to interpret it inconsistently in a single piece of work.

In fact, Mandelbrot's original definition has met with some dissatisfaction; it has been suggested in [43, 100] that the definition is too restrictive, since there exist highly irregular sets of *integer* Hausdorff dimension that strongly deserve to be called fractals. Falconer [43] suggests that instead of using a precise definition, fractals be described in more qualitative terms:

(2.2.1) DEFINITION. [Falconer, 1990]. A *fractal* is any set that has most or all of the following properties:

- fine structure at all scales;
- self-similarity;
- an irregular structure not adequately described by traditional geometry, and;
- a simple, recursive definition.

This definition is more intuitive than Mandelbrot's original one, as it is not based on the concept of set dimension.

Mathematical sets that are now known as fractals were first studied near the turn of the century by Cantor, Hausdorff, Peano and others [84]. These sets were described variously as sets of fractional dimension, sets of Hausdorff measure, sets with a fine

structure or irregular sets [42]. Most mathematicians considered them to be pathological, as they challenged the geometric intuition of the time. Poincaré went so far as to call them "a gallery of monsters" [144]. A few of these sets are described in the following sections.

2.3 On measuring the size of fractal sets

In this section we establish the properties of several well-known fractals, in order to illustrate how fractals have no well-defined size in the traditional sense. As we provide a fair amount of detail in these examples, the reader may choose to browse the proofs in favour of gaining a general understanding of the concepts involved.

The triadic Cantor set

Probably the earliest and best known example of a fractal is the *triadic Cantor set*. The set is named after Georg Cantor¹, who made use of the set in 1883, although it was actually defined by Henry Smith in 1875 [100, 145]. This set is also known as *Cantor's middle-third set*, the *triadic Cantor dust*, *Cantor's ternary set*, and occasionally as 'the' Cantor set. The triadic Cantor set can be defined by a process which, starting with a line

¹ Georg Cantor (1845-1919) is best known for his work on the foundations of set theory, including his "diagonalization" proof that the real numbers are uncountable. An interesting account of Cantor's work can be found in [88].

segment of unit length, removes an infinite number of smaller and smaller line segments, leaving only a 'dust' of points. The construction proceeds as follows.

Let C_0 be the unit interval $[0,1]$. Define the *approximant* C_{i+1} as the set of points that remain after removing the open middle-third of every interval in C_i , where $i \geq 0$. The sequence of sets C_i are thus defined as

$$\begin{aligned} C_0 &= [0, 1] \\ C_1 &= [0, \frac{1}{3}] \cup [\frac{2}{3}, 1] \\ C_2 &= [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1] \\ &\vdots \\ C &= \bigcap_{i \geq 0} C_i \end{aligned}$$

This process is illustrated in Figure 2.1. The triadic Cantor set C is the set of points that appear in *every* iterate C_i ; that is, C is the limiting set of the sequence $\{C_i\}$ as i goes to infinity. In what follows, we show that C is in fact non-empty.

Figure 2.1 The Triadic Cantor Set

The triadic Cantor set is the limit of a process that starts with a unit line segment and recursively removes the middle third from each interval remaining in the set. It has an uncountable number of points but has a total "length" of zero.

We now measure the size of the triadic Cantor set, in terms of the number of points in the set, as well as its overall length.

(2.3.1) DEFINITION. [Falconer, 1990]. A *basic interval* of the approximant C_i is an interval $[a,b] \in C_i$ of length 3^{-i} . Each approximant C_i is made up of 2^i basic intervals.

(2.3.2) PROPOSITION.* [Falconer, 1990]. The boundary points of each approximant C_i are contained in the intersection C .

PROOF.

The boundary points of C_i are simply the endpoints of the basic intervals contained in the set. We show that every approximant C_i contains the boundary points of all other approximants, and hence that all boundary points are contained in the intersection C .

Because $C_i \supset C_{i+1}$ for all $i \geq 0$, it is clear that each C_i contains the boundary points of all $C_{i,j}$ for $j > 0$. Thus, it is required to show that every C_i contains the boundary of all previous $C_{i,j}$ for $j < i$. Intuitively, we are trying to show that no boundary points will ever be removed by the construction process.

This is proved by induction, using C_0 and C_1 as the basis. Since $C_0 = [0,1]$ is the first approximant, it trivially contains the boundary points of all "previous" approximants. This is also true for $C_1 = [0,1/3] \cup [2/3,1]$, since it contains the boundary points $\{0,1\}$ of C_0 .

Assume that C_k contains the boundary points of all previous approximants C_j , where $j < k$. The set C_{k+1} is formed by removing the open middle-third of each basic

interval in C_k . Consider the basic interval $[a,b] \in C_k$. Removing the middle-third of $[a,b]$ will not remove either of its endpoints $\{a,b\}$, nor will it remove any of the other boundary points of C_k (since $[a,b]$ contains no other boundary points). Thus the boundary points of C_k are all contained in C_{k+1} , completing the induction.

Therefore the intersection C contains the boundary points of all approximants C_i , completing the proof. ■

Next, we prove that there are an uncountable number of points in the triadic Cantor set. This is done by establishing a one-to-one correspondence between these points and the real numbers in the unit interval. To this end, we identify the real numbers between 0 and 1 with their ternary (base-3) representations. We note, however, that some of these values have two ternary representations; these are the rational numbers of the form $p/3^k$ for some integer $0 \leq p \leq 3^k$, where the ternary representation of p ends in a 1. In particular, if the ternary expansion of p has the form $0.p_{k-1} \dots p_3 p_2 p_1 1$ then the number $p/3^k$ can be written in ternary either as $0.p_{k-1} \dots p_3 p_2 p_1 1$, or as $0.p_{k-1} \dots p_3 p_2 p_1 02222 \dots$. We shall ignore the first form, and use the second.

(2.3.3). LEMMA. Each basic interval $[a,b]$ of the approximant C_i can be written in the form $3^{-i}[p,p+1]$, where $[p,p+1]$ is a real interval for some integer p .

PROOF.

By induction. As a basis we observe that the single basic interval of C_0 is of the required form, since $[0,1] = 3^0[0,1]$. Assume that the basic intervals of C_k are of the required form;

that is, if $[a,b]$ is a basic interval in C_k then $[a,b]=3^{-k}[p,p+1]$ for some integer p . The set C_{k+1} is formed by removing the open middle-third of each basic interval in C_k . Thus C_{k+1} contains the intervals $L=[a, a+(b-a)/3]$ and $R=[a+2(b-a)/3, b]$, corresponding to $L=3^{-k}[p, p+1/3]=3^{-(k+1)}[3p, 3p+1]$ and $R=3^{-k}[p+2/3, p+1]=3^{-(k+1)}[3p+2, 3p+3]$. Observing that the endpoints of these intervals are of the required form completes the induction. ■

(2.3.4). LEMMA. Let C_i be an approximant to the triadic Cantor set, with $i \geq 0$. Then none of the points in C_i have ternary representations having a 1 in the first k positions.

PROOF.

By induction. We use $C_0=[0,1]$ as the basis, for it is trivial that none of the ternary representations of the points in C_0 have a 1 in the first 0 positions. Assume this condition to hold for all points in C_k , and let $[a,b]$ be a basic interval in this set. By Lemma 2.3.3 we know that $[a,b]$ is of the form $3^{-k}[p,p+1]$ for some p . Now the points that are 'removed' from this interval are given by $3^{-(k+1)}(3p+1, 3p+2)$, which are exactly those points having a 1 in the $(k+1)$ -st position. Therefore none of the points in C_{k+1} can have a 1 in the $(k+1)$ -st position, and the induction is complete. ■

(2.3.5) PROPOSITION.[†] The points in the triadic Cantor set correspond to those real numbers between 0 and 1 whose ternary representations do not contain the digit 1.

PROOF.

By Lemma 2.3.4 we know that each approximant C_i corresponds to those real numbers between 0 and 1 whose ternary representations do not contain a 1 in the first i positions.

Therefore none of the points in the intersection C can have ternary representations containing the digit 1 in any position. ■

(2.3.6) PROPOSITION.[†] There is an uncountable number of points in the triadic Cantor set C .

PROOF.

It is clear that C contains a countably infinite subset, being the endpoints of the basic intervals in each C_i (which are enumerable). We must show that there are an uncountable infinity of other points in C .

By Proposition 2.3.5 we know that the triadic Cantor set corresponds to the set of infinite ternary strings containing only the digits 0 and 2. This set can be put into one-to-one correspondence with the set of infinite binary strings, with the ternary digit 2 corresponding to the binary digit 1. Since the set of infinite binary strings is uncountable the triadic Cantor set is itself uncountable. ■

One of the interesting properties of the triadic Cantor set is that although it contains an uncountably infinite number of points, it nevertheless has a "total length" of zero. We show this with respect to Lebesgue measure (defined in the Appendix) in the following.

(2.3.7) PROPOSITION.[†] The triadic Cantor set has (linear) Lebesgue measure zero.

PROOF.

We observe that because $C \subset C_i$ for all i , we have $L^1(C) \leq L^1(C_i)$, where L^1 denotes the linear Lebesgue measure. Now the set C_i contains 2^i intervals of size 3^{-i} , hence $L^1(C_i) = (2/3)^i$. This implies that $L^1(C) \leq (2/3)^i$ for all $i \geq 0$. Letting $i \rightarrow \infty$ gives $L^1(C) = 0$ to complete the proof. ■

Because the triadic Cantor set has total length zero it can contain no intervals, and there must be unfilled space between any two points in the set [84, 100, 118]. This result also implies that the set of all points whose ternary representations are missing the digit 1 has Lebesgue measure zero. It is interesting to note that Hardy and Wright give a more general result in [70], which we reproduce here without proof:

(2.3.8) THEOREM. [Hardy and Wright]. The set of points whose decimals, in any scale, miss any digit is null: almost all decimals contain all possible digits.² ■

The Peano curve

Euclidean geometry dictates that points are zero dimensional objects, curves one-dimensional, surfaces two-dimensional and solids three-dimensional. In 1890 Giuseppe

² In this context the term 'decimal' refers to the expansion of a real number in some base, or 'scale'. A 'null' set is a set of Lebesgue measure zero.

Peano³ defied this view by inventing a continuous curve that visits every point in the unit square [88]. The *Peano curve* was defined by a mapping between the unit line segment and the filled-in unit square.⁴ In this mapping, each point (x,y) in the square is specified by two continuous functions, $f(t)$ and $g(t)$ such that for every $(x,y) \in [0,1] \times [0,1]$ there exists $0 \leq t \leq 1$ for which $(f(t), g(t)) = (x,y)$.

The Peano curve represents the limit of a constructive process which repeatedly maps the unit line segment to a curve consisting of smaller and smaller line segments, as shown in Figure 2.2. The first step in the construction maps the line segment to the curve shown in Figure 2.2(a), which consists of nine line segments, each of size $1/3$. The process continues by replacing each of these line segments by nine smaller line segments, and so on *ad infinitum*. Figures 2.2(b) through 2.2(d) shows the next few approximations to the curve, which do indeed give the impression of filling in the square.

In establishing the correspondence of a line segment to a square, Peano showed that set dimension cannot depend on the multiplicity of points in a set. He also showed that the dimension of a set is not simply a matter of the number of coordinates required to identify each point, since each point (x,y) in the square can be specified by a single value $0 \leq t \leq 1$ [88, 144].

³ Giuseppe Peano (1858-1932) is also known for his work on the fundamental axioms of arithmetic, a major goal in mathematics in the late nineteenth century.

⁴ Actually, the term "curve" is inappropriate, since equivalence between the Peano curve and a square means that the curve is in fact two dimensional. However, we shall continue to use the term, since its use is standard in the literature.

a)

b)

c)

d)

Figure 2.2 Peano curves for levels 1,2,3 and 4

Part a) shows the first level Peano curve, where the diagonal line has been replaced by nine line segments at one-third scale. Parts b)-d) shows how higher level curves fill in the square. The limiting Peano curve is in fact two-dimensional.

The von Koch snowflake

In 1906 Helge von Koch (1870-1924) defined a set with the interesting ability to enclose a finite area with a continuous curve of infinite length. Like Peano's space-filling curve, von Koch's "snowflake" curve challenged the geometrical intuition of the time [84, 88].

The snowflake curve can be constructed by the following recursive process. Let S_0 be the curve formed by the perimeter of an equilateral triangle. Define the *approximant* S_{i+1} as the curve that results from replacing each line segment in S_i with four non-overlapping line segments of one third the size, as shown in Figure 2.3. The snowflake curve S is defined as the limit of this sequence of approximants:

$$S = \lim_{i \rightarrow \infty} S_i$$

This limiting curve is depicted in Figure 2.4.

a)

b)

c)

d)

Figure 2.3 von Koch snowflake curves for levels 0,1,2 and 3

Construction of the snowflake begins with a triangle, as shown in a). Each side of the triangle is replaced by four line segments at one-third scale, as shown in b). The process continues for smaller and smaller line segments as shown in c) and d).

4

Figure 2.4 The limiting von Koch snowflake curve

Next, we measure the length of the snowflake curve, and beginning with the following.

(2.3.9). LEMMA. Let E_0 be the length of the edges of the equilateral triangle in S_0 . Then the approximant S_i consists of $3 \cdot 4^i$ line segments each of length $3^{-i}E_0$. In addition S_i has a total length of $3(4/3)^i E_0$.

PROOF.

By induction. As a basis, we observe that S_0 consists of $3(4^0)=3$ line segments of length $3^0E_0=E_0$. Assume that S_k consists of $3(4^k)$ line segments of length $3^{-k}E_0$. The approximant S_{k+1} is formed by replacing each of these line segments by four line segments that are one third the size. Hence S_{k+1} consists of $3(4^k)4=3(4^{k+1})$ line segments of length $(3^{-k}/3)E_0=3^{-(k+1)}E_0$, completing the induction. Thus S_i has a length of $3(4/3)^i E_0$. ■

(2.3.10) PROPOSITION.[†] The snowflake curve S has infinite length.

PROOF.

This follows directly from the fact that the length of S_i is $3(4/3)^i E_0$, where E_0 is the size of the edges in S_0 . Since this value grows without bound as i goes to infinity, the length of the limiting curve S is infinite. ■

Next, we consider the area enclosed by the snowflake curve.

(2.3.11). LEMMA. Let A_0 be the area enclosed by S_0 . Then each of the small triangles 'added' to S_i in the construction of S_{i+1} encloses an additional area of $(1/9)^{i+1}A_0$.

PROOF.

Let E_0 be the size of the edges in S_0 , as before. By the construction of S , the edge size decreases by a factor of three at each stage; hence $E_{i+1}=E_i/3$, and $E_i=(1/3)^iE_0$. Now S_{i+1} is constructed from S_i by 'adding' small triangles of edge size E_{i+1} . Thus

$$\left(\frac{E_{i+1}}{E_0}\right)^2 A_0 = \left(\frac{(\frac{1}{3})^{i+1}E_0}{E_0}\right)^2 A_0 = \left(\frac{1}{3}\right)^{2i+2} A_0 = \left(\frac{1}{9}\right)^{i+1} A_0$$

completing the proof. ■

(2.3.12). LEMMA. Let A_i be the area enclosed by S_i , for $i \geq 0$. Then A_{i+1} exceeds A_i by an amount equal to

$$\frac{1}{3}A_0\left(\frac{4}{9}\right)^i$$

PROOF.

In the construction of S_{i+1} , a small triangle of area $(1/9)^{i+1}A_0$ is added to each edge in S_i .

Since there are $3(4^i)$ edges in S_i , the total area added is given by

$$A_{i+1}-A_i = 3(4^i)\left(\frac{1}{9}\right)^{i+1}A_0 = \frac{1}{3}A_0\left(\frac{4}{9}\right)^i$$

which completes the proof. ■

(2.3.13) PROPOSITION.* Let A_0 be the area enclosed by the equilateral triangle S_0 . Then the snowflake curve S encloses an area of $(8/5)A_0$.

PROOF.

Let A_i be the area enclosed by the approximant S_i , and let A be the area enclosed by the snowflake curve S . Then A can be expressed as the sum of the differences between the A_i :

$$A = A_0 + (A_1 - A_0) + (A_2 - A_1) + \dots + (A_{i+1} - A_i) + \dots$$

The difference in values between A_{i+1} and A_i is equal to $(1/3)(4/9)^i A_0$, by Lemma 2.1.12.

Thus, the total area A is given by

$$\begin{aligned} A &= A_0 + \left(\frac{1}{3}\right) A_0 + \left(\frac{1}{3}\right) \left(\frac{4}{9}\right) A_0 + \dots + \left(\frac{1}{3}\right) \left(\frac{4}{9}\right)^i A_0 + \dots \\ A &= A_0 + \left(\frac{1}{3}\right) A_0 \left\{ 1 + \left(\frac{4}{9}\right) + \dots + \left(\frac{4}{9}\right)^i + \dots \right\} \end{aligned}$$

Except for the first term, this is a geometric progression of common ratio $4/9$. Thus

$$\begin{aligned} A &= A_0 + \left(\frac{1}{3}\right) A_0 \lim_{i \rightarrow \infty} \left(\frac{1 - \left(\frac{4}{9}\right)^{i+1}}{1 - \left(\frac{4}{9}\right)} \right) \\ A &= A_0 + \left(\frac{1}{3}\right) A_0 \left(\frac{9}{5}\right) \\ A &= \frac{8}{5} A_0 \end{aligned}$$

and the proof is complete. ■

The fact that the snowflake curve bounds a finite area with a curve of infinite length represented a divergence from traditional geometric intuition. Next, we discuss an

example of a curve that has infinite length but bounds a region of zero area.

The Sierpinski triangle

By the early part of this century many fractal sets had been invented, with each representing a counterexample to some accepted law of geometry. The Sierpinski triangle, invented by the Polish mathematician Wacław Sierpinski (b 1882), is a well-known example of a fractal set. Its construction is similar to that of the triadic Cantor set, except that the Sierpinski triangle begins as a two dimensional object.

Let T_0 be a filled equilateral triangle with edge size E_0 and area A_0 with the sequence of approximants T_i defined as follows. Construct T_1 by extracting the middle half-scale triangle formed by joining the midpoints of the three sides. T_1 is thus composed of three half-scale triangles in a pyramid shape, with the centre triangular area empty. In general, T_{i+1} is constructed by removing the inverted centre triangles of each of the triangles contained in T_i , where the removals are done in such a way that the boundary of the triangle is left behind. The Sierpinski triangle T is the limiting set of this process, and is depicted in Figure 2.5.

Figure 2.5 The Sierpinski triangle

The triangle is constructed by a recursive process that repeatedly removes the centre of each triangle at the previous level.

(2.3.14). LEMMA. Let A_0 be the area of the initial triangle T_0 . Then T_i is composed of 3^i filled triangles, each having an area of $(1/4)^i A_0$, with the total area of T_i equal to $(3/4)^i A_0$.

PROOF.

By induction. We use T_0 as a basis, since it contains 3^0 triangles of size $(1/4)^0 A_0$. Assume this property to hold for T_k . To construct T_{k+1} , each of the 3^k triangles in T_k are broken into four sub-triangles at half-scale, with one being removed. Thus the number of triangles contained in T_{k+1} is equal to $3(3^k) = 3^{k+1}$, and each has area $(1/4)(1/4)^k A_0 = (1/4)^{k+1} A_0$, completing the induction. Thus, T_i has a total area of $A_i = (3/4)^i A_0$. ■

(2.3.15) PROPOSITION.[†] The Sierpinski triangle has area zero.

PROOF.

Let A_i be the area of T_i , where $A_i = (3/4)^i A_0$. Observing that $A_i \rightarrow 0$ as $i \rightarrow \infty$ is enough to prove the result. ■

Define the *perimeter* of the Sierpinski triangle as the curve formed by the boundaries of the smaller triangles removed during the construction process.

(2.3.16) PROPOSITION.[†] The Sierpinski triangle has an infinite perimeter.

PROOF.

Let P_i be the perimeter of the approximant T_i for $i \geq 0$. Then P_i is equal to the number of edges in T_i times the edge size of 2^{-i} :

$$P_i = 3 (3^i) (2^{-i}) E_0 = 3 E_0 \left(\frac{3}{2}\right)^i$$

where E_0 is the edge size of the original triangle. Taking the limit as $i \rightarrow \infty$ implies that the Sierpinski triangle has an infinite perimeter. ■

This result shows that, while the Koch snowflake had an infinite perimeter enclosing a finite area, the Sierpinski triangle has an infinite perimeter enclosing zero area.

The limitations on measuring set sizes

These examples have shown that the traditional notions of set size (namely cardinality, length, area, or volume) are unsatisfactory when applied to fractal sets. For example, saying that the triadic Cantor set contains an uncountably infinite number of points but has an overall length of zero seems to be an inadequate description of the set. For fractal sets, then, a different measure of their character is required.

In the following sections we discuss the notion of *set dimension*, a concept which proves to be more useful in describing fractal sets. As will be seen, there are many definitions of set dimension; we begin with the earliest, due to Euclid.

2.4 Euclidean dimension

Euclid (circa 300 B.C.) set forth the foundations of what we now know as Euclidean geometry in his treatise the *Elements* [87].⁵ Book I begins with the following definitions and axioms of planar geometry.

1. A point is that which has no part.
2. A line is breadthless length.
3. The extremities of a line are points.
4. A straight line is a line which lies evenly with the points on itself.
5. A surface is that which has length and breadth only.
6. The extremities of a surface are lines.

Kline [87] remarks that Euclid likely presented his ideas in a way that would appeal to the reader's physical intuition, though at the expense of rigor. Certainly the above definitions agree with our intuitive notions of standard planar objects. The following definitions of spatial objects appear in Book XI.

1. A solid is that which has length, breadth, and depth.
2. An extremity of a solid is a surface.

⁵ The *Elements* contains thirteen books in which Euclid summarized the mathematical knowledge of the time, ranging from the geometry of planar objects, the theory of numbers, the study of proportions and incommensurables (rational and irrational numbers) and spatial geometry. It includes many famous theorems, including the Pythagorean theorem and Euclid's algorithm for finding the greatest common divisor of two numbers.

The standard figures of Euclidean geometry can be classified according to their *Euclidean dimension*, which denotes whether the object is a point, curve, surface, or solid. In current usage, the Euclidean dimension of a set is determined by the number of coordinates required to specify the points in the set; in other words, Euclidean dimension is just the "n" in $\mathbb{R}^n$. For example, we see that the triadic Cantor set has Euclidean dimension 1 since it "lives" in the real line $\mathbb{R}^1$, and the Peano curve, Koch snowflake, and Sierpinski triangle each have Euclidean dimension 2 since they each "live" in the plane $\mathbb{R}^2$.

Unfortunately, Euclidean dimension has limited utility, due to its dependence on a particular coordinate system. In particular, the Euclidean dimension of a set can change with its orientation in $\mathbb{R}^n$. For example, let L be the unit line segment $L=[0,1] \subset \mathbb{R}^1$; clearly, $\dim_E L=1$. Let M be the rotation of L through 45 degrees such that the endpoints of M lie at $(0,0)$ and $(1/\sqrt{2}, 1/\sqrt{2})$. Now since $M \subset \mathbb{R}^2$ we have $\dim_E M=2$; intuitively, however, M should have dimension one because it is only a line segment.

In general, Euclidean dimension is not preserved under homeomorphisms and is therefore not a topological property [88]. This limitation leads to a more general definition of dimension, described in the next section.

2.5 Topological dimension

Euclidean dimension suffers from a dependence on the co-ordinate system in which a set is represented. In this section we discuss the notion of *topological dimension*, which successfully avoids this problem. We illustrate the use of this definition by applying it to some of the fractal sets introduced earlier.

In 1912 Henri Poincaré⁶ proposed an inductive definition of set dimension, wherein a continuum (a closed, connected set) is said to have dimension N if it can be separated into two parts whose boundaries are continua of dimension $N-1$ [88, 100, 122]. Poincaré's definition was put on a firm mathematical basis by Menger and Urysohn in 1926, and is now known as *topological dimension*. There are now many definitions of topological dimension, although it is typically known by the name *small inductive dimension*. We shall use [39] as the main source for material in this section.

Intuitively, the definition begins with zero dimensional sets, containing only isolated points. One dimensional sets are defined as sets whose boundaries are zero dimensional. For example, the unit line segment $[0,1]$ is one dimensional because its boundary set $\{0,1\}$ has dimension 0. Similarly, the unit square is two dimensional since its boundary consists of four line segments which have dimension 1, and a cube is three

⁶ Henri Poincaré (1854-1912), considered the greatest mathematician of his age, has often been called "the last universalist", as he roamed freely through all major areas of mathematics [88, 145]. Poincaré gives an interesting philosophical discussion of dimension in his essay "Why space has three dimensions" [122].

dimensional since its boundary consists of six squares which have dimension 2. In general, a set can be thought of as being N -dimensional if its boundary has dimension $N-1$. We lead up to the formal definition of topological dimension by giving a few preliminary definitions on metric spaces.

(2.5.1) DEFINITION. [Edgar, 1990]. A *clopen* subset of a metric space is a subset that is both closed and open.

Let A be a clopen subset of a metric space S . By definition, A contains all of its accumulation points (since it is closed), and has the property that all points in A are at the centre of an open ball contained entirely within A (since it is open). Another way to think of a clopen set is as a set with no boundary. A set of this type appears in Example 2.5.4 below.

(2.5.2) DEFINITION. [Edgar, 1990]. A family B of open subsets of a metric space S is called a *base for the open sets* of S if and only if, for every open set $A \subset S$, and every $x \in A$, there exists a $U \in B$ such that $x \in U \subset A$.

(2.5.3) DEFINITION. [Edgar, 1990]. A metric space S is said to have small inductive dimension zero, written $\dim_7 S = 0$, if there exists a base B for the open sets in S such that B consists only of clopen sets.

(2.5.4) EXAMPLE. [Edgar, 1990]. Any finite set $S = \{x_1, x_2, \dots, x_n\}$ has $\dim_T S = 0$.

PROOF.

Let δ be a small value such that $|x_i - x_j| > \delta$ for all $x_i, x_j \in S$, and let $B_\delta(x_i)$ be an open ball centred on x_i . Because δ is smaller than the minimum distance between any pair of points in S we have $B_\delta(x_i) \cap B_\delta(x_j) = \emptyset$. Now the closure of $B_\delta(x_i)$ is simply $\{x_i\}$ and the closure of $S - B_\delta(x_i)$ is simply $S - \{x_i\}$, which are disjoint sets. Hence every ball $B_\delta(x_i)$ is without a boundary and thus clopen. Since the B_δ 's form a base for the open sets in S we have it that S has topological dimension zero. ■

Earlier, we showed that the triadic Cantor set C has Lebesgue measure zero. We are now able to see that C also has topological dimension zero.

(2.5.5) PROPOSITION. [Edgar, 1990]. The triadic Cantor set C has $\dim_T C = 0$.

PROOF.

We are required to find a base for the open sets of C that consists only of clopen (boundaryless) sets. Let C be the triadic Cantor set, defined as

$$C = \bigcap_{k \geq 0} C_k$$

where the approximants C_k contained 2^k disjoint closed intervals, each of length 3^{-k} .

Writing I_{kj} for such an interval we have

$$C_k = \bigcup_{j=1}^{2^k} I_{kj}$$

Let $M_{kj} = C \cap I_{kj}$, namely that part of C covered by the interval I_{kj} . Now M_{kj} is closed in

C since both C and I_{k_j} are closed. On the other hand, M_{k_j} is open in C because every point $x \in M_{k_j}$ is the centre of an open ball $B_\delta(x) \subset C$ ($\delta < 3^{-k}$) containing only points from M_{k_j} .

It remains to show that the set of all M_{k_j} for $k \geq 0$, $1 \leq j \leq 2^k$ forms a base for the open sets in C . Given any point $x \in C$ and $\delta > 0$, choose k such that $3^{-k} < \delta$ and j such that $x \in I_{k_j}$. Then $C \cap B_\delta(x)$ contains M_{k_j} , implying that the set of all M_{k_j} forms the required base. ■

Sets of higher topological dimension are defined in terms of zero-dimensional sets, as follows.

(2.5.6) DEFINITION. Let N be a positive integer. Then a set S is said to have *small inductive dimension* $\dim_T S \leq N$ if and only if there is a base for the open sets of S consisting of sets U such that $\dim_T \partial U \leq N-1$, where ∂U is the boundary of U . In particular, $\dim_T S = N$ if $\dim_T S \leq N$ but $\dim_T S \not\leq N-1$.

(2.5.7) PROPOSITION. The Peano curve P has topological dimension 2.

PROOF.

According to its construction, the Peano curve P visits every point in the unit square, and is hence equivalent to the square. Let S be the unit square minus its four boundary lines. A base for the open sets in S is the family of open balls $B_\delta(x) \subset S$ where $x \in S$; hence $\dim_T S \leq 2$. Since the boundary of each $B_\delta(x)$ is a one-dimensional circle we have $\dim_T S = 2$. ■

Before discussing the topological dimension of the von Koch snowflake curve, it is necessary to establish its continuity.

(2.5.8) DEFINITION. A *continuous curve* in a metric space T is a continuous function $f:[0,1] \rightarrow T$.

(2.5.9) PROPOSITION. The von Koch snowflake curve S is continuous.

PROOF.

Recall that the snowflake curve represents the limit of a sequence of polygonal approximations (or approximants), as shown in Figure 2.3. Let S_k be the k -th approximant, for $k \geq 0$, and let S be the final snowflake curve. Certainly each S_k is continuous, as it consists only of connected line segments. Thus each S_k can be represented by a continuous function $f_k:[0,1] \rightarrow S_k$.

It is a standard result that if a sequence of continuous functions f_k converges "uniformly" to a function f , then f is continuous. We claim that the f_k do indeed converge uniformly.⁷ This implies that the limit function, and hence the snowflake curve, are continuous. ■

(2.5.10) PROPOSITION. The von Koch snowflake curve S has topological dimension 1.

PROOF.

Consider the family of open sets

⁷ Although we have not discussed uniform convergence here, the claim can nonetheless be justified (see, for example, [39]).

$$\{ f((x-\delta, x+\delta)) : x \in [0, 1], \delta > 0 \}$$

where f is the continuous function generating the snowflake curve. Let V be an open set in S . Then for every $y \in V$ there exist $x \in [0, 1]$ and $\delta > 0$ such that $y = f(x)$ and $f((x-\delta, x+\delta)) \subseteq V$. Thus the family defined above is a base for the open sets of S .

Let U be one of these open sets; that is, $U = f((x-\delta, x+\delta))$ for some x and δ . Now, the interval $(x-\delta, x+\delta)$ is a one-dimensional set with a zero-dimensional boundary. Due to the continuity of f , U must also have this property. This implies that the snowflake curve S must itself have topological dimension 1. ■

By similar arguments, it can be shown that the Sierpinski triangle also has topological dimension 1.

Finally, we remark that the topological dimension of a set is invariant under homeomorphisms. That is, if S and T are homeomorphic sets, then $\dim_T S = \dim_T T$. The proof can be found in [39]. In general, we have $\dim_T \leq \dim_E$. It will turn out that all definitions of fractal dimension will fall between the topological and Euclidean dimensions.

2.6 Hausdorff dimension

In this section we discuss the first definition of set dimension that could be applied to fractal sets. This definition differs from the Euclidean and topological definitions, as it is based on a *measure* of the size of a set, which is an important concept in all definitions of fractal dimension.

In 1919 Felix Hausdorff⁸ used an extension of Lebesgue measure to define the notion of fractional set dimension [39]. The notion of *Hausdorff dimension*⁹ has proved to be the most useful mathematically, although it is usually the hardest to work with [42]. We lead up to Hausdorff dimension by defining the *Hausdorff measure* of a set.

(2.6.1) DEFINITION. [Falconer, 1985]. Let F be a set, and let $\{U_i\}$ be a countable collection of sets of diameter at most δ that cover F . That is

$$F \subset \bigcup_{i=0}^{\infty} U_i$$

where the diameter $|U_i|$ is such that $0 < |U_i| \leq \delta$ for $i \geq 0$. Then $\{U_i\}$ is said to be a δ -cover of F .

⁸ Felix Hausdorff (1868-1942) worked in the theory of metric spaces. In 1914 he developed the notion of a topological space, which can be thought of as a metric space in which the elements themselves are sets of points S [88].

⁹ The name *Hausdorff-Besicovitch dimension* often appears in the literature. Although Hausdorff provided the mathematical foundation, it was his student A.S. Besicovitch who later developed the theory.

Let F be a set and s be a non-negative real number. For $\delta > 0$, define the following function:

$$H_{\delta}^s(F) = \inf \left\{ \sum_{i=1}^{\infty} |U_i|^s : \{U_i\} \text{ is a } \delta\text{-cover of } F \right\}$$

This function considers the sets U_i in all δ -covers of F , and minimizes the sum of the s -th powers of the diameters of these sets. This function has a limit as δ goes to zero, which leads to the following

(2.6.2) DEFINITION. The s -dimensional Hausdorff measure of the set F is defined as

$$H^s(F) = \lim_{\delta \rightarrow 0} H_{\delta}^s(F)$$

It is interesting to note that the s -dimensional Hausdorff measure is equal to the s -dimensional Lebesgue measure, to within a constant factor.

The following proposition establishes the relationship between Hausdorff measures for different values of the exponent. This relationship is central to the definition of Hausdorff dimension.

(2.6.3) PROPOSITION. Let F be a set. Then for $t > s$ and $\delta < 1$, $H_{\delta}^t(F) \leq \delta^{t-s} H_{\delta}^s(F)$.

PROOF.

Consider some δ -cover $\{U_i\}$ of the set F . Now $|U_i| \leq \delta$ implies $|U_i|^t = |U_i|^{t-s} |U_i|^s$, hence $|U_i|^t \leq \delta^{t-s} |U_i|^s$, giving

$$\begin{aligned}
& \sum_i |U_i|^t \leq \delta^{t-s} \sum_i |U_i|^s \\
& \Rightarrow \inf \sum_i |U_i|^t \leq \delta^{t-s} \inf \sum_i |U_i|^s \\
& \therefore H_\delta^t(F) \leq \delta^{t-s} H_\delta^s(F)
\end{aligned}$$

which completes the proof. ■

Thus, depending on s , the value of H^s may be zero, finite, or infinite. In the limit as δ goes to zero, if $H^s(F)$ is finite, then $H^t(F)$ must be zero for $t > s$. On the other hand, if $H^s(F)$ is infinite, then $H^t(F)$ must be finite for $s < t$ [39, 43]. This leads us to the following

(2.6.4) DEFINITION. The *Hausdorff dimension* of a set F is the value of s for which the Hausdorff measure $H^s(F)$ 'jumps' from zero to infinity. That is,

$$\begin{aligned}
\dim_H F &= \inf \{s : H^s(F) = 0\} = \sup \{s : H^s(F) = \infty\} \\
\Rightarrow H^s(F) &= \begin{cases} \infty & \text{if } s < \dim_H F \\ 0 & \text{if } s > \dim_H F \end{cases}
\end{aligned}$$

In other words, $\dim_H$ can be described as the scaling factor s for which the s -dimensional Hausdorff measure of the set takes on a reasonable (non-zero, finite) value.¹⁰

It is interesting to note that the Hausdorff dimension of any discrete (countable) subset of $\mathbb{R}$ is zero [42]. We sketch a proof of this as follows. For $s=0$, all values in the

¹⁰ Actually, the definition also applies to sets for which the value of H^s is always either zero or infinity even for $s = \dim_H$, provided $H^s = \infty$ for $s < \dim_H$ and $H^s = 0$ for $s > \dim_H$.

summation

$$\sum_{j=0}^{\infty} |U_j|^s$$

are raised to the power 0 resulting in an infinite sum. Thus the 0-dimensional Hausdorff measure of any set is infinite. We claim that the s-dimensional Hausdorff measure of a countable set is zero for $s > 0$. The proof is similar to that for the proposition given in the Appendix that all countable sets have Lebesgue measure zero. This implies that the measure jumps from infinity to zero for $s = 0$, and that the Hausdorff dimension of a discrete set is zero. ■

(2.6.5) PROPOSITION. [Falconer, 1985]. The Hausdorff dimension of the triadic Cantor set is $\log 2 / \log 3$.

PROOF. See Appendix A. ■

In general, the direct calculation of Hausdorff dimensions can be rather difficult [42, 43]. Because of this, we do not calculate the Hausdorff dimensions of the Peano curve, Koch snowflake curve, or Sierpinski triangle here. Instead, their Hausdorff dimensions will be established using a convenient result given in the next section.

2.7 Similarity dimension

Many fractals, including the ones discussed thus far, have the interesting property of being *self-similar* (or recursively symmetric), wherein the entire set is similar to some proper subset of itself. For example, the triadic Cantor set can be broken into left and right halves, each of which is similar to the entire set. Self-similar sets often have agreeable properties which make the calculation of Hausdorff dimension much easier. Self-similarity is discussed in [32, 80, 100], as well as in the popular book by Michael Barnsley [8]. The material in this section is based on the books by Edgar and Falconer [39, 42, 43]. We begin with several definitions.

(2.7.1) DEFINITION. A function $f:S \rightarrow T$ is a *similarity* if and only if there is a positive number r such that

$$|f(y)-f(x)| = r|y-x|$$

for all $x,y \in S$. The number r is called the *ratio* of f . Two metric spaces are said to be *similar* if there is a similarity of one onto the other.

(2.7.2) DEFINITION. An *iterated function system (IFS)* over a set S is a list $(f_1, \dots, f_n)$ of functions, where $f_i: S \rightarrow S$ is a similarity with ratio r_i . The *ratio list* of an IFS is written $(r_1, r_2, \dots, r_n)$, where r_i is the ratio of the similarity function f_i .

(2.7.3) DEFINITION. A ratio list $R=(r_1, r_2, \dots, r_n)$ is said to be *contracting* if $r_i < 1$ for $1 \leq i \leq n$.

An iterated function system is said to be *contracting* if it has a contracting ratio list.

(2.7.4) DEFINITION. A non-empty set $K \subseteq S$ is said to be an *invariant* set of the iterated function system $(f_1, \dots, f_n)$, if and only if there is a finite decomposition of K

$$K = \bigcup_{i=1}^n f_i(K)$$

(2.7.5) DEFINITION. A set is said to be *self-similar* if it is the invariant set of an iterated function system.

(2.7.6) DEFINITION. The *similarity dimension* of a self-similar set K is the number s such that

$$\sum_{i=1}^n r_i^s = 1$$

where $(r_1, r_2, \dots, r_n)$ is the ratio list of the corresponding iterated function system. The similarity dimension of K is written $\dim_s K$.

A theorem due to Edgar [39] guarantees the existence and uniqueness of the compact invariant set K , and of its dimension s . As a corollary, Edgar shows that an arbitrary initial set K_0 will converge under iteration of the similarity functions $\{f_i\}$ to the invariant set K . We reproduce this result (without proof) as

(2.7.7) THEOREM. [Edgar, 1990]. If K_0 is any non-empty compact set in S , and if

$$K_{j+1} = \bigcup_{j=1}^n f_j(K_j)$$

for $i \geq 0$, then the sequence $\langle K_i; i \geq 0 \rangle$ converges in the Hausdorff metric to the invariant set of the iterated function system. ■

This result suggests a method for depicting the invariant set of a contractive iterated function system. First, let K_0 denote the entire set S , and then compute K_1 as $K_1 = f_0(K_0) \cup f_1(K_0) \cup \dots \cup f_n(K_0)$. By the theorem, the Hausdorff distance between K_1 and the invariant set K will be less than the Hausdorff distance between K_0 and K . Continuing this process, there will eventually be some iterate K_i that is sufficiently "close" to K that the two sets would be indistinguishable on a computer display. This condition could be determined by noticing when two successive iterates are indistinguishable on the display.

Although this method for displaying the invariant set would work in theory, it would be quite inefficient in practice. A better method makes use of the fact that the invariant set K is equal to the set of cluster points of a sequence $\langle x_i \rangle$ generated by "random" applications of the similarity functions [39]. That is, if $x_0 \in S$, and $k_1, k_2, \dots$ is an (algorithmically) random sequence in $\{1, 2, \dots, n\}$ then the sequence

$$x_i = f_{k_i}(x_{i-1}), \quad i \geq 1$$

has as its set of cluster points the invariant set K . In practice, this method involves choosing an arbitrary point on the screen, then computing successive iterates of that point

by applying the similarity functions at random. A classic example of a self-similar set generated by an algorithm of this type is the Sierpinski triangle, shown in Figure 2.5.

The Hausdorff dimension of a self-similar set is related to its similarity dimension in the following way. Suppose F is a self-similar set, such that

$$F = \bigcup_{i=1}^n F_i$$

where the F_i are disjoint subsets of F that are similar to F . Then we have

$$\begin{aligned} H^s(F) &= \sum_{i=1}^n H^s(F_i) \\ H^s(F) &= \sum_{i=1}^n r_i^s H^s(F) \end{aligned}$$

Assuming that the value of $H^s(F)$ is non-zero and finite for $s = \dim_H F$, we can divide by $H^s(F)$ to obtain

$$1 = \sum_{i=1}^n r_i^s$$

which is the definition of the similarity dimension s . It is a standard result that the Hausdorff dimension of the set F is equivalent to the similarity dimension of F when the sets F_i do not overlap "too much". In particular, if the F_i are disjoint or overlap at a countable set of points, then the Hausdorff and similarity dimensions are equivalent. This criterion is made precise in [39, 43].

(2.7.8) PROPOSITION. The similarity dimension of the triadic Cantor set is $\log 2 / \log 3$.

PROOF.

By its construction (see Figure 2.1), the triadic Cantor set C is similar to each of the one-third scale copies of itself at the left and right sides of the set. The similarity dimension of C is given by the number s satisfying

$$\left(\frac{1}{3}\right)^s + \left(\frac{1}{3}\right)^s = 1$$

giving $s = \log 2 / \log 3$. ■

(2.7.9) **PROPOSITION.** The similarity dimension of the Peano curve is 2.

PROOF.

At each stage in the construction of the Peano curve (see Figure 2.2), every line segment is replaced by nine smaller line segments at one-third scale. Hence the curve has similarity dimension given by the number s satisfying

$$9 \left(\frac{1}{3}\right)^s = 1$$

giving $s = \log 9 / \log 3 = 2$. ■

(2.7.10) **PROPOSITION.** The similarity dimension of the boundary of the von Koch snowflake is $\log 4 / \log 3$.

PROOF.

Note that the complete snowflake, consisting of the boundary and enclosed area, is not self-similar since it does not contain any smaller copies of the entire set. However, the boundary is self-similar, as can be seen by its construction in Figure 2.3. Each side of

the snowflake curve is similar to four one-third scale copies of itself. Therefore the dimension of the boundary is equal to $\log 4 / \log 3$. ■

(2.7.11) PROPOSITION. The similarity dimension of the Sierpinski triangle is $\log 3 / \log 2$.

PROOF.

The result follows directly from the fact that the Sierpinski triangle is similar to three half-scale copies of itself, as can be seen in Figure 2.5. ■

We remark that the Hausdorff dimensions of the above sets are equal to the corresponding similarity dimensions, since in each case the sets can be partitioned into similar subsets that overlap in at most a countable number of points.

2.8 Box-counting dimension

Until now we have discussed the concept of dimension as applied to sets of purely mathematical origin. The triadic Cantor set, for example, is the physically unattainable result of a recursive process that is repeated an infinite number of times. The same is true for the Peano curve, the von Koch snowflake and the Sierpinski triangle - none of these sets could accurately describe any object in the physical world.

In spite of this fact, however, the notion of dimension can be useful in describing

many natural phenomena [46, 100]. Essentially, the dimension of a natural object is related to how a measure of its size behaves at different scales. In this context, the size of an object refers to its n -dimensional volume, which for $n=1$ corresponds to length, for $n=2$ corresponds to area, and so forth.

In this section we discuss the notion of box-counting dimension which, like Hausdorff dimension, is based on the measurement of a set's size. Unlike Hausdorff dimension, however, box-counting dimension is useful in many practical situations. It is interesting to note that box-counting dimension has been known by many names, including *Kolmogorov entropy*, *entropy dimension*, *capacity dimension*, *metric dimension*, *logarithmic density*, and *information dimension* [43].

An estimate of the n -dimensional volume of a set can be obtained by counting the number of small volume elements required to cover the set.¹¹ For example, no finite number of line elements can be used to cover a solid square, since the line elements only have Euclidean dimension one, whereas the square has dimension two. In most cases, two dimensional volume elements are used, resulting in the name 'box-counting'.

For simplicity, it can be assumed that volume elements take the form of cells in an n -dimensional grid, where n is the Euclidean dimension of the set being measured. In two dimensions, for example, the cells would correspond to the 'boxes' found on a

¹¹ The Euclidean dimension of the volume elements must be at least as large as that of the set; otherwise they could not form a cover.

sheet of graph paper. We shall refer to n -dimensional cells of side length δ as δ -cells.¹²

A collection of δ -cells that form a cover of a set will be called a δ -cell cover.

We now give some preliminary notation. Let F be a set and let N_δ be the number of δ -cells required to cover F . That is,

$$N_\delta(F) = \min \{ \# \{U_i\} : \{U_i\} \text{ is a } \delta\text{-cell cover of } F \}$$

Next, we define the *approximate s -dimensional volume* of F at scale δ by

$$B_\delta^s(F) = N_\delta(F) \delta^s$$

where s is allowed to take on non-integer values. This leads us to the following.

(2.8.1) DEFINITION. The *s -dimensional box-counting volume* of F is the limit of the approximate s -dimensional volume as $\delta \rightarrow 0$, and is given by

$$B^s(F) = \lim_{\delta \rightarrow 0} B_\delta^s(F)$$

assuming that the limit on the right side exists.

In the case where the exponent s is too small, B_δ^s will grow unbounded as δ goes to zero, resulting in an infinite value for B^s . This would occur, for example, in an attempt to measure the length of a surface, or the area of a solid. On the other hand, if s is too large, then B_δ^s will vanish as δ goes to zero, resulting in a zero value for B^s .

¹² The use of δ -cells as volume elements is convenient but not necessary; a δ -cell covering can be replaced by any other covering in which the volume elements have diameter at most δ [43].

However, there may exist a value of s for which B_δ^s tends to a finite non-zero limit. In what follows, we shall assume that such an s exists and is unique.¹³

(2.8.2) DEFINITION. Let F be a set. If there exists a unique value s for which $B_\delta^s(F)$ tends to a finite non-zero limit, then s is called the *box-counting dimension* of F , denoted by $\dim_b F$.

Henceforth, we shall assume that a well-defined box-counting dimension always exists.

(2.8.3) PROPOSITION.¹ Let F be a set and let δ and $N_\delta(F)$ be defined as above. Then,

$$\dim_b(F) = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta}$$

PROOF.

Let $s = \dim_b(F)$. Then by the definition of dimension, B_δ^s has a constant value, which we denote by c :

$$B_\delta^s = \lim_{\delta \rightarrow 0} N_\delta(F) \delta^s = c$$

Let c_δ denote the approximate s -dimensional box-counting dimension of F at scale δ ; that is, $c_\delta = B_\delta^s(F) = N_\delta(F) \delta^s$. Thus

¹³ In the case where there are more than one such s , the notions of *upper* and *lower* box-counting dimensions can be used [43]. We elect not to discuss this situation further, however.

¹ Recall that this symbol indicates an original proof to an existing result.

$$s = \frac{\log N_\delta(F) - \log c_\delta}{-\log \delta} \quad \forall \delta$$

$$s = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F) - \log c_\delta}{-\log \delta}$$

Observing that $c_\delta \rightarrow c$ yields

$$s = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta}$$

which is the desired result. ■

Intuitively, the box-counting dimension represents the logarithmic rate at which the number of boxes increases as $\delta \rightarrow 0$. Next, we derive an expression for computing the box-counting dimension.

Determining the box-counting dimension empirically

Whereas it is generally difficult to determine the Hausdorff dimension of a set, it is often easy to find its box-counting dimension if the construction of the set is known. This is shown in the following.

(2.8.4) PROPOSITION. The box-counting dimension of the triadic Cantor set is $\log 2 / \log 3$.

PROOF.

Recall that the triadic Cantor set represents the limit of a sequence of *approximant* sets

C_k , each containing 2^k intervals of length 3^{-k} . Thus

$$B_{3^{-k}}^s(F) = 2^k (3^{-k})^s$$

Setting $s = \log 2 / \log 3$ gives $B^s(F) = 1$. Observing that this gives a positive finite value for $B^s(F)$ is enough to complete the proof. ■

In practice, this method of determining the box-counting dimension cannot be used, since knowledge of how a set is constructed will not generally be available. Nevertheless, the box-counting dimension can be estimated from a collection of measurements of the set volume $B_\delta^n(F)$ taken at different scales, as shown in the following.

(2.8.5) PROPOSITION.[†] Let $B_\delta^n(F)$ be related to δ by a power law of the form

$$B_\delta^n(F) = c\delta^r$$

Then the box-counting dimension is given by $\dim_B(F) = n - m$.

PROOF.

Let $s = \dim_B(F)$ as usual. For s to act as a dimension, the value of the box-counting volume must be independent of the scale. That is,

$$\begin{aligned} B_\delta^s(F) &= B_{\frac{1}{2}}(F) \quad \forall \delta \\ \Rightarrow N_\delta(F) \delta^s &= N_{\frac{1}{2}}(F) 1^s \end{aligned}$$

By the power law relation above, we have

$$N_1(F) \delta^{n-s} = c \delta^n$$

$$\Rightarrow N_1(F) = c, \quad \dim_B(F) = s = n - m$$

which completes the proof. ■

In practice, it is an easy matter to determine to what extent B_δ^n obeys a power law.

For if $B_\delta^n(F) = c\delta^m$, then taking logs on both sides gives

$$\log B_\delta^n(F) = \log c + m \log \delta$$

implying that a log-log plot of $B_\delta^n(F)$ against δ would form a line of slope m .

Note that for objects in the physical world, B_δ^n obeys a power law only across a limited number of scales - as $\delta \rightarrow 0$ the atomic level is eventually reached, and the scaling properties of the set no longer hold. However, for many natural sets the relation between $B_\delta^n(F)$ and δ can be approximated by a power law across a wide range of scales, thereby giving the set a characteristic box-counting dimension. For example, certain cloud structures have been found to have a box-counting dimension of about $4/3$ across seven orders of magnitude [100].

Finally, we note that discrete sets can have non-zero box-counting dimension, whereas the Hausdorff dimension of a discrete set is always zero. In particular, Falconer [43] shows that for any bounded set $F \subset \mathbb{R}^n$, the box-counting dimension of F is equal to

the box-counting dimension of its closure.¹⁴ Thus, at least in theoretical settings, box-counting dimension must be used with some care since the introduction of a measure-zero set can drastically affect the dimension, in the sense that there exist sets of Lebesgue measure zero whose box-counting dimension is non-zero.

Intuitively, this difference between box-counting and Hausdorff dimensions is due to the nature of the set coverings involved. Box-counting dimension is defined in terms of finite δ -cell coverings $\{U_i\}$, where $|U_i|=\delta$ for all i . Thus,

$$B_\delta(F) = N_\delta(F) \delta^s = \min \left\{ \sum_{j=1}^{N_\delta(F)} |U_j|^s : \{U_j\} \text{ is a } \delta\text{-cell cover of } F \right\}$$

will be bounded below by $\delta^s > 0$. However, Hausdorff dimension is based on countable δ -covers $\{U_i\}$ where $|U_i| \leq \delta$ for all i . Hence

$$H_\delta^s(F) = \inf \left\{ \sum_{j=1}^{\infty} |U_j|^s : \{U_j\} \text{ is a } \delta\text{-cover of } F \right\}$$

can be arbitrarily small for each δ .

¹⁴ This implies, for example, that a discrete dense set in $\mathbb{R}^n$ would have n as its box-counting dimension, even though its Hausdorff dimension would be zero.

2.9 Correlation dimension

Another type of dimension that is useful in practical situations is called the *correlation dimension*, proposed by Grassberger and Procaccia [7, 58]. It is similar to the box-counting dimension, except that it also takes the local *density* of points in a set into account.¹⁵

The correlation dimension is based on a measure of what fraction of points in a set are within some radius R of each other. Let x_i, x_j be members of a bounded set F , and let N be a certain number of points from the set, chosen at random. Then the fraction of these points that are within a radius R of x_i is given by

$$\frac{1}{N} \sum_{j=1}^N H(R - |x_i - x_j|)$$

where the Heaviside step function $H(x)$ simply counts points x_j for which $|x_i - x_j| \leq R$ and excludes the rest.¹⁶ Define the *approximate correlation function* as

$$\begin{aligned} C_N(R) &= \frac{1}{N} \sum_{i=1}^N \frac{1}{N} \sum_{j=1}^N H(R - |x_i - x_j|) \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N H(R - |x_i - x_j|) \end{aligned}$$

Intuitively, $C_N(R)$ is the average fraction of the N points that are within R of each other.

¹⁵ For this reason the box-counting and Hausdorff dimensions are referred to as *metric* dimensions, whereas the correlation dimension is referred to as a *frequency* dimension.

¹⁶ The Heaviside step function $H: \mathbb{R} \rightarrow [0, 1]$ is defined as $H(x) = 0$ for $x < 0$, and $H(x) = 1$ for $x \geq 0$.

The *correlation function*

$$C(R) = \lim_{N \rightarrow \infty} C_N(R)$$

is the limiting average as increasing numbers of points are included from the set. The *correlation dimension*, written $\dim_c(F)$, is defined by the variation of $C(R)$ with R :

$$C(R) \approx R^{\dim_c(F)} \text{ as } R \rightarrow 0$$

Thus $\dim_c(F)$ can be interpreted as the logarithmic rate at which the correlation function increases with R . In a manner similar to that of the box-counting dimension, the existence of a well-defined correlation function determines a straight line in the plot of $\log C(R)$ versus $\log R$. Because $\dim_c$ takes into account the local density of points, which $\dim_b$ ignores, they are related as

$$\dim_b \leq \dim_c$$

The correlation dimension has the advantage of being easier to compute in practical situations, where it is often used to detect the presence of a *strange attractor* in a physical system from a time-series of observations [7, 49, 58, 112].

2.10 Information dimension

The last dimension that we will discuss is the *information dimension* of a set, which is closely related to the notion of *entropy* in thermodynamics.¹⁷ Like $\dim_c$, the information dimension takes into account the distribution of points in a set by considering the probability of finding a point in a particular region.

Let F be a bounded subset of $\mathbb{R}^n$, covered by $N_\delta(F)$ n -dimensional "boxes" of side-length δ , and let p_i be the probability that a point is contained in the i -th box. Then the *approximate metric entropy* of F at scale δ is given by

$$I_\delta(F) = - \sum_{i=1}^{N_\delta(F)} p_i \log p_i$$

Intuitively, $I_\delta(F)$ measures that amount of information necessary to specify an arbitrary point in the set. If all of the boxes are equally probable, then $I_\delta(F)$ reaches its maximum value of $\log N$. The information dimension of a set represents the scaling exponent in the variation of I_δ with δ :

$$\dim_i(F) = \lim_{\delta \rightarrow 0} \left[- \frac{I_\delta(F)}{\log \delta} \right] = \lim_{\delta \rightarrow 0} \left[\frac{\sum_{i=1}^{N_\delta(F)} p_i \log p_i}{\log \delta} \right]$$

¹⁷ We include this section in the interest of completeness, and to emphasize the fact that fractal dimensions are defined with respect to some measure of the "size" of a set.

In practical situations the information dimension can be estimated by partitioning a set into N bins and then assigning randomly selected points from the set to the bins in which they reside. The relative sizes of the bins determines approximate values from the probabilities p_i . Grassberger and Procaccia [61] report that the information dimension is related to box-counting and correlation dimensions by $\dim_B \leq \dim_I \leq \dim_C$.

2.11 Summary

A wide variety of definitions exist for the notion of set dimension, with each one designed to capture some different property of sets. However, the term *fractal dimension* is ambiguous, and careful writers should avoid using it (or at least they should give it a precise definition; for example, see [154]). More generally, the term *dimension* must itself be interpreted in one of the following contexts.

- *With respect to a co-ordinate system.* $\dim_E$ indicates the number of co-ordinates required to specify a point in a Cartesian space. Other co-ordinate systems are also possible.
- *With respect to connectedness.* $\dim_T$ indicates the extent to which a set corresponds to a continuum, and is defined recursively in terms of set boundaries.
- *With respect to self-similarity.* $\dim_S$ indicates the extent to which a set is

similar to proper subsets of itself. It is akin to the box-counting dimension, except that the "boxes" covering the set are in fact scaled-down copies of the whole set.

- *With respect to a measure.* The dimension represents the scaling exponent for which the "measure" (be it Hausdorff, box-counting, correlation, or entropy) is scale-invariant. It can be thought of as the logarithmic rate at which the measure varies with the scale.

Hausdorff dimension is the most useful in theoretical settings, but it is the most difficult to work with practically. Self-similarity dimension is often easy to determine, provided the construction of the set is known. Box-counting dimension is used most often in practice, especially in classifying the structure of physical objects [46]. It has also been used in the study of fractal-like patterns observed in the evolution of cellular automata [9, 26, 57, 149, 155, 160, 161, 162, 163, 166]. Finally, the correlation dimension has been used in the study of fractal patterns observed in time-series measurements from physical systems [58].

3. Chaos

A very small cause, which escapes us, determines a considerable effect which we cannot ignore, and we then say that this effect is due to chance.

Henri Poincaré, 1908

3.1 Motivation

In this chapter we present most of the important concepts in the theory of chaos and dynamical systems. The chapter begins by providing some historical background to the area, emphasizing the fact that the elements of chaos theory were in existence at the turn of the century. Following this, we provide several definitions of chaos and discuss how they are related. In addition we discuss the notion of a *strange attractor*, wherein lies the relationship between chaos and fractal geometry.

3.2 Historical Background

It has long been believed that nature obeys mathematical laws, and that given knowledge of these laws and the state of the world as it is, mathematical methods can be used to predict its future. This belief can be traced back to Descartes, who in 1637

advocated a scientific method that would enable humans "to become the lords and possessors of nature". The elements of Descartes' method are to (a) begin with self-evident truths, (b) proceed from the simple to the complex by a sequence of clear and logical steps, (c) break larger problems into smaller ones, and (d) check when done.¹⁸

Descartes' dream was that all of areas of human knowledge (including science, art, law, and politics) could be enlightened by reason alone [31]. This idea was later encouraged by the tremendous success of Newtonian mechanics in describing and predicting the behaviour of natural systems (especially in celestial mechanics).¹⁹ Descartes' dream was also widely promoted by another French mathematician, Laplace, who conceived of

An intelligence which, at a given instant, would know all the forces by which Nature is animated, and the respective situation of all the elements of which it is composed, if furthermore it were vast enough to submit all these data to analysis, would in the same formula encompass the motions of the largest bodies of the universe, and those of the most minute atom: nothing for it would be uncertain, and the future as well as the past would be present to its eyes.²⁰

¹⁸ René Descartes 1637, *Discourse in the Method of Properly Guiding the Reason in the Search of Truth in the Sciences*.

¹⁹ Newton laid out his "Rules of Reasoning in Philosophy", in the same spirit as Descartes' method, in book three of the *Principia* (published in 1686) [15].

²⁰ Pierre Simon Marquis de Laplace, *Essai philosophique sur les probabilités*, Paris, 1814. For an English translation, see [91]. Laplace (1749-1827) is best known for his book *Mécanique céleste*, in which he describes in detail the application of the law of gravitation to astronomical phenomena [109].

This view, now referred to as *classical determinism*, has been a part of our scientific tradition for over three hundred years [24, 131]. It was not until the early twentieth century that its validity came into question.

The first challenge to classical determinism came with the development of quantum mechanics, which relied on a probabilistic view of events at the atomic scale. In particular, Heisenberg's uncertainty principle put strict limits on the precision with which physical measurements can be made:

Nature puts up with our probings into its mysteries only on conditions. The more we clarify the secret of position, the more deeply hidden becomes the secret of velocity... We can distribute the uncertainty as we wish, but we can never get away from it.²¹

Thus the uncertainty principle effectively ruled out the accurate prediction of events at the atomic scale, making phenomena such as radioactive decay inherently unpredictable [24, 73].

Classical determinism has been further eroded by the discovery that deterministic mathematical systems can behave in a way that is indistinguishable from a random process. In recent years, such *deterministic randomness* has come to be known as *chaos*. Chaos has been found in mathematical models of many natural phenomena, including

²¹ Werner Heisenberg, quoted in [74]. Heisenberg (1902-1976) won the Nobel prize in Physics in 1932 for his contributions to quantum theory and the development of the uncertainty principle.

population growth [33, 101, 102], heart arrhythmias [54], epileptic seizures [49], weather patterns [96, 97], turbulent fluid flow [133, 150], chemical reactions [147, 157], quantum mechanics [64], laser dynamics [45, 48], economic trends [4, 105], and the motion of bodies in the solar system [164].

In the next section, we introduce a number of concepts from dynamical systems theory, in order to provide an appropriate context for the discussion to follow.

3.3 Dynamical systems

Toward the end of the nineteenth century it was realized that not all problems in mathematics can be solved analytically, and that qualitative methods would have to be used instead.²² This problem was addressed by Henri Poincaré (1854-1912) who laid the foundation for a qualitative theory now known as *dynamical systems theory*.²³

In what follows we shall use the term *evolution* to mean the sequence of state changes experienced by a system over time, and the term *evolution law* to mean the rules

²² A solution is said to be *analytic* if it consists only of an indefinite integral involving elementary functions and uniformly convergent power series [81]. A well-known example is the N-body problem in celestial mechanics, for which no analytic solutions exist when $N \geq 3$.

²³ Poincaré actually developed dynamical system theory in his attempt to determine the stability of the solar system [145]. The term "dynamical system" is due to the American topologist Stephen Smale, who used it instead of the traditional "system of differential equations" in order to emphasize the topological approach [141].

that govern these changes. Evolution laws with no dependence on time are said to be *autonomous*; the greater part of dynamical systems theory deals with systems in this category. This leads to the following informal

(3.3.1) DEFINITION. A *dynamical system* is any *autonomous system* that obeys a deterministic evolution law.²⁴

It is important to note that the evolution of a dynamical system is by definition *deterministic*, wherein the future behaviour of the system is completely determined by its current state. This contrasts the situation in which random chance plays a part in the evolution law.

Types of dynamical systems

Traditional models of physical systems have been based on systems of differential equations. These systems, being continuous both in time and space, are called *continuous dynamical systems*, or *flows*. Although analytical tools exist for solving certain kinds of differential equations, most systems resist analysis and require numerical techniques [107]. Here, the usual approach is to approximate the continuous system by a model in which time is represented as a sequence of discrete steps.

The time discretization of a continuous dynamical system leads to a *discrete*

²⁴ More formal definitions can be found in, for example, [5, 81, 159].

dynamical system, or *map*, which is more amenable to solution by computer. Care must be taken, however, as the process of discretization can sometimes lead to a model having significantly different behaviour from that of the original system [10, 35, 98]. In addition to being used in the numerical approximation of flows, maps are often used to model certain phenomena directly. They appear, for example, in the study of (seasonal) population growth, where the number of individuals changes abruptly from one generation to the next. The most familiar example is the logistic map, which models the dynamics of a single population in a bounded space [34, 101].

Although discrete dynamical systems lend themselves to implementation on a computer, their continuous spatial component cannot be truly represented. Many calculations will inevitably require precision greater than that offered by the computer. For some systems this may not be a problem, but if the system is chaotic, small errors can quickly accumulate to corrupt the overall computation. Chaotic systems effectively defy computation, and make accurate long-term predictions impossible. Lorenz's chaotic model of fluid convection is the classic example of this phenomenon [96, 143].

Recent use has been made of models that are discrete in both time and in space. These models represent *fully discrete dynamical systems*, and unlike systems having continuous components, they can be implemented exactly on a computer (provided enough memory is available). Examples of fully discrete systems include finite state machines, Turing machines, cellular automata, and recurrence relations on integers. Fully

discrete models have been used in applications ranging from chemical systems [53] to population dynamics [33].

Phase space

Central to the study of dynamical systems is the notion of *phase space*, wherein the state of a system is represented as a point in a multidimensional space. For example, consider a general system of three coupled differential equations:

$$\frac{dx}{dt} = F(x,y,z)$$

$$\frac{dy}{dt} = G(x,y,z)$$

$$\frac{dz}{dt} = H(x,y,z)$$

The phase space of this system is contained in $\mathbb{R}^3$, since the system state is uniquely specified by the triple (x,y,z) . Similarly, the phase space for a general differential system in n variables is contained in $\mathbb{R}^n$.

In most cases the phase space of a system can be represented as some bounded subset of $\mathbb{R}^n$, where the system variables can only take on a finite range of values. A dynamical system is said to be *bounded* if it has a bounded phase space. As will be seen later, this boundedness is a requirement for chaotic behaviour [82].

The evolution of a dynamical system is represented by a *trajectory* in phase space. Trajectories appear as smooth curves in continuous time systems, and as sequences of

points in discrete time sequences. A computed solution to a dynamical problem is also represented by a trajectory in phase space, although it may not correspond to a *theoretical* trajectory, since there are usually errors in the computation at each step. However, it may be that such a computed trajectory "shadows" an existing theoretical trajectory. A significant amount of work has gone into studying the extent to which computed solutions reflect actual system behaviour [23, 104, 115, 116, 117].

Asymptotic behaviour

The limiting, or asymptotic behaviour of a dynamical system can be described by the structure of its trajectories for large time values. We base the following definition on Wiggins [159]:

(3.3.2) DEFINITION. A point $x_0 \in \mathbb{R}^n$ is a *limit point* of a trajectory $\{x_t\}$ of $x \in \mathbb{R}^n$ if there exists a sequence of time values $\{t_i\}$, $t_i \rightarrow \infty$ such that $x_{t_i} \rightarrow x_0$.

In other words, x_0 is a limit point for x if the trajectory $x(t)$ passes closer and closer to x_0 over time.

The set of all limit points is called the *limit set* of a dynamical system. Various features can be found in the limit set of a dynamical system. These include

- *sinks*, toward which all nearby trajectories converge, representing a stable steady

- state;
- *sources*, from which all nearby trajectories diverge, representing an unstable steady state, and;
- *saddle points*, being steady states that are stable in some directions and unstable in others.

Each case describes a different kind of *fixed point* which, loosely put, represents a state of equilibrium within a system. Periodic behaviour is represented by a

- *limit cycle*, being a closed loop in phase space and which may be either stable or unstable

In the terminology of dynamical systems, stable structures are said to be *attracting* and unstable structures *repelling*.

Structural stability

An important notion is that of a dynamical system that is *structurally stable*, wherein small perturbations to the evolution rule result in systems having limit sets that are qualitatively the same [34, 81, 159]. The features appearing in structurally stable systems are said to be *typical*.

A standard result due to Poincaré and Bendixson states that the only typical features in two-dimensional continuous dynamical systems are sources, sinks, saddle points, and limit cycles. For three dimensional systems it was known that typical features included these four, plus a

- *torus*, being the superposition of two periodic motions with incommensurate periods.

In 1967 Stephen Smale [141] showed that the features typical in three-dimensional dynamical systems included the expected ones, plus a new type of feature now called a

- *strange attractor*, being a structurally stable attracting set that is neither a fixed point, limit cycle, or a torus

Attractors of this kind were called *strange* as they possessed unexpected, often fractal properties. The name *strange attractor* was first used by Floris Takens and David Ruelle in [133]. Strange attractors will be discussed later on in this chapter.

3.4 Sensitive dependence

We now discuss the notion of *sensitive dependence*, which is arguably the most important concept in the study of chaos. Sensitive dependence is a feature common to all chaotic systems, wherein small changes in the initial state of a system can have a large impact on its behaviour. Such systems effectively defy prediction by numerical methods, since any uncertainty in the state of the system is amplified throughout the computation, until it attains a size comparable to that of the phase space itself.

(3.4.1) DEFINITION. A dynamical system S is said to have *sensitive dependence* at a point x in S provided trajectories passing near to x diverge at an exponential rate (on

average) for a significant period of time.

The phrase *on average* allows trajectories to approach each other on occasion provided they otherwise diverge at an exponential rate. This leads us to our first definition of chaos.

(3.4.2) DEFINITION. [Sensitive chaos]. A dynamical system is said to be *chaotic* if it has sensitive dependence at almost every point in its phase space.

The phrase *almost every point* allows the existence of a set of points over which the system does not have sensitive dependence, provided this set has measure zero.²⁵ This is the most common usage of the term *chaotic*, and can also be expressed by saying that the system has *sensitive dependence on initial conditions* [130].

Exponential growth in errors (both due to measurement and computation) puts strict limits on the extent to which accurate predictions are possible. In the absence of exact measurement and computation, all possible evolutions of a chaotic system will be unpredictable.

Sensitive dependence was first discovered in 1898 by Jacques Hadamard in his

²⁵ An example would be the system $f(x)=x^2-2$, which is chaotic even though it has a dense set of points (the pre-images of $x=0$) over which the system is stable. (Courtesy of R.L. Devaney, personal communication).

study of *geodesic flows on surfaces of negative curvature*.²⁶ Hadamard found that the motion of point particles on such surfaces was unstable, and stated explicitly that any uncertainty in the initial condition would grow exponentially with time, making long-term predictions useless. We remark that Hadamard's work is sometimes mentioned in the literature as part of introductory discussions of chaos. Unfortunately, some authors give the impression that an understanding of chaos requires knowledge of these 'geodesic flows', which is not the case.

The dynamical system studied by Hadamard is similar in spirit to the game of billiards [65, 66, 159]. Here, any uncertainty in the trajectory of a ball is amplified with each collision until eventually the computed trajectory 'hits' an obstacle that the theoretical trajectory does not (or vice versa). Although convincing arguments can be made geometrically, it was not until 1970 that the dynamics of billiards (in general, hard sphere gases) was actually proved to be chaotic [139].

The importance of Hadamard's discovery was recognized by Poincaré [121], who gave a crisp description of sensitive dependence, wherein "a very small cause, which escapes us, determines a considerable effect which we cannot ignore, and we then say that this effect is due to chance".²⁷ Poincaré considered a system to be *predictable* if its

²⁶ Jacques Hadamard (1865-1963), a contemporary of Poincaré, is primarily known for his proof of the prime number theorem in 1896 (and for living to a very old age!).

²⁷ The original reference, in French, is Henri Poincaré, *Le hasard*, chapter 4 in: *Science et Méthode* (Paris: Ernest Flammarion, 1908). *Science et Méthode* is a classic text on the philosophy of science.

long-term behaviour could be approximated with no more uncertainty than that found in the initial condition.

Due to the subjective nature of the phrase *long-term*, we shall say that a dynamical system is *effectively predictable* if the uncertainty in the system state remains sufficiently small for an amount of time that is large enough for predictions to remain useful. Otherwise, we say that the system is *effectively unpredictable*.²⁸

Calculations have actually been performed to estimate the degree of sensitive dependence in physical systems. According to Atlee Jackson [80],

Borel pointed out in 1914 that the displacement of 1 gram by 1 centimeter on the star Sirius would result in a variation of the Earth's gravitational field whose perturbation on the motion of gas molecules would lead to our ignorance about the molecular collisions in roughly 10^6 s.²⁹

The existence of chaos in many (if not most) physical systems, together with the impossibility of absolute precision in measurement, implied that the dream of classical determinism could never be realized. Said Poincaré bluntly, "Determinism is a fantasy due to Laplace" [47].

²⁸ This will, of course, depend on the application. However, it is generally true that a system in which errors grow exponentially will be effectively unpredictable.

²⁹ E. Atlee Jackson, page 343 in [82]. See also Ruelle in [131].

Lyapunov exponents

Points that are initially close in the phase space of a chaotic dynamical system have trajectories that tend to diverge exponentially with time. Of course, exponential divergence of nearby trajectories cannot continue indefinitely in a bounded system; eventually the rapidly expanding region of phase space is forced to fold back on itself. For example, this corresponds to the situation in billiards where the perturbed trajectory of a ball causes it to hit an object that it would otherwise have not. The degree of sensitive dependence in a dynamical system is measured in terms of its *Lyapunov exponents*, which we define as follows.³⁰

Let $F:S \rightarrow S$ be a continuous map, and let x be a point in S .³¹ Then the trajectory or "orbit" of x consists of the sequence of points $\langle x, F(x), F^2(x), F^3(x), \dots \rangle$, where as usual $F^n(x)$ denotes the n -th iterate of F evaluated at x . This trajectory is said to be *unstable* if S has sensitive dependence at every point in the trajectory. The degree of this instability is measured by the Lyapunov exponent of the trajectory. Now it may happen that the distance between the trajectories of x and some nearby point $x+\epsilon$ can be modelled

³⁰ Also called the *characteristic exponent*, this term is named after the Russian mathematician Alexander Mikhailovich Lyapunov (1857-1918) who worked on the stability of dynamical systems at the turn of the century. His doctoral thesis is considered a classic in the area - an English translation is available as *Stability of Motion*, Academic Press, New York 1966. It is a sad fact that Lyapunov himself died by violence in the aftermath of the Russian revolution [13, 138].

³¹ Lyapunov exponents can also be defined for differential systems, by defining a map that represents the action of the dynamical system over time steps of fixed length [132].

by an exponential:

$$|F^n(x+\epsilon) - F^n(x)| \approx \epsilon e^{\lambda n}$$

for some value λ . Solving,

$$\lambda \approx \frac{1}{n} \log \left| \frac{F^n(x+\epsilon) - F^n(x)}{\epsilon} \right|$$

For small ϵ , the expression in the absolute value sign can be replaced by the derivative of F^n assuming, as we do, that the derivative exists:

$$\lambda \approx \frac{1}{n} \log \left| \frac{dF^n(x)}{dx} \right|$$

The Lyapunov exponent is defined to be the limiting value of this expression as n goes to infinity assuming, as we do, that the limit exists:

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \log \left| \frac{dF^n(x)}{dx} \right|$$

Thus λ represents the average rate of (exponential) growth in the derivative of the n -th iterate of F , evaluated at the point x . Notice that because λ is averaged over the entire forward trajectory, it takes on the same value for all points along the trajectory.

Unfortunately, it is not practical to compute Lyapunov exponents in this way. Instead, the expression above can be simplified by applying the chain rule to the derivative of $F^n(x)$:

$$\begin{aligned}
\frac{dF^n(x)}{dx} &= F'(F^{n-1}(x))F'(F^{n-2}(x))\dots F'(F(x))F'(x) \\
&= \prod_{i=0}^{n-1} F'(F^i(x)) \\
&= \prod_{i=0}^{n-1} F'(x_i)
\end{aligned}$$

where $x_i = F^i(x)$ is the i -th point in the trajectory of x . Inserting this into the definition of λ gives

$$\begin{aligned}
\lambda &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left| \prod_{i=0}^{n-1} F'(x_i) \right| \\
\lambda &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log |F'(x_i)|
\end{aligned}$$

Thus λ is computed as the average rate of divergence measured over all points in the forward trajectory.³² In a similar fashion, the Lyapunov exponent for the entire system can be found by averaging the logarithmic rate of divergence at each point in the phase space:

$$\lambda = \int p(x) \log |F'(x)| dx$$

where $p(x)$ is the probability of finding the system at x . In the case where we have an unstable *dense* trajectory, then the exponent of the trajectory will represent that of the system itself, since the trajectory comes arbitrarily close to each point in the phase space. In many cases, however, the system-wide Lyapunov exponent may be difficult to find,

³² In [165], Wolf suggests that exponential growth be measured with respect to base 2 instead of base e , thereby interpreting the exponent as the rate of information loss measured in bits per iteration. Although this is an appealing interpretation, most authors still use base e .

and is in general not equal to the exponents for individual trajectories [82].

Positive values for the Lyapunov exponent indicate the presence of sensitive dependence, and set the time scale over which predictions can be made. Negative values indicate that neighbouring trajectories tend to converge rapidly, and set the time scale over which initial transients die out. Finally, zero values indicate that neighbouring trajectories change their relative distance only slowly. Thus we have the following

(3.4.3) DEFINITION. A dynamical system S is *chaotic* if it possesses at least one positive Lyapunov exponent.

In dynamical systems having many degrees of freedom, Lyapunov exponents are defined for each independent variable, giving the next

(3.4.4) DEFINITION. A dynamical system is *hyperchaotic* if it possesses more than one positive Lyapunov exponent.

Hyperchaos corresponds to stretching of the phase space in more than one direction, and can be observed by sampling more than one variable of a multidimensional system.³³

As can be seen from the derivation above, Lyapunov exponents can be determined

³³ *Hyperchaos* is a recent term in the literature, and is due to Rössler [124, 125, 126]. The invention of a new term seems unnecessary in this case given the modest generalization that it represents.

from the system equations, but in practice one studies systems where the evolution laws are unknown. Accordingly, techniques have been developed for estimating the Lyapunov exponents from time series taken from observations of a system [7, 14, 36, 38, 41, 49, 92, 102, 119, 165]. The basic idea is to find elements of the time series that represent nearby points in the phase space, and measure the distance between the corresponding values at the next time step.

3.5 Topological chaos

Chaos has been described as the middle ground between order and randomness; having order through a deterministic evolution law, and randomness in the rapid growth of initial uncertainty in the system state. In this section we discuss an alternative definition due to Robert L. Devaney, which requires that a chaotic system possess the attributes of uncertainty, underlying regularity, and indecomposability [34, 86]. The definition is given in terms of a one-dimensional continuous map over the reals, but can be extended to flows in a natural way.

As will be seen, Devaney's definition requires knowledge of the evolution law of a system, and hence cannot be used in many practical situations. Instead, it is concerned with the qualitative nature of the evolution rule itself, and for this reason (in part) we refer to it as *topological chaos*.

Unpredictability

As before, unpredictability is expressed in terms of the sensitive dependence of the evolution law, although this definition differs from that given in the previous section.

(3.5.1) DEFINITION. [Devaney, 1989]. A continuous map $F:S \rightarrow S$ has sensitive dependence on initial conditions if there exists $\delta > 0$ such that, for any $x \in S$ and any neighbourhood N of x , there exists $y \in N$ and $n \geq 0$ such that $|F^n(x) - F^n(y)| > \delta$.

We shall call this property *topological sensitivity*. Essentially, the definition says that nearby trajectories eventually diverge by an amount at least equal to δ , everywhere in the phase space S . We note some differences from the definition of sensitive chaos:

1. The rate of trajectory divergence need not be *exponential*; polynomial (even linear) divergence would suffice, provided the distance between the diverging trajectories eventually exceeds δ . However, the time required will depend on the initial distance between the trajectories, and may be very large.³⁴
2. It is not required that *all* trajectories near a point diverge; it is only required that *there exists* a divergent trajectory in every neighbourhood of a point. Thus it may even be the case that most nearby trajectories stay close to each other, with a

³⁴ Poincaré noted that exponential divergence is not required for a system to become unpredictable, although it may take a very long time for this to happen otherwise [121, 131]

measure-zero set of trajectories actually being divergent.

(3.5.2) DEFINITION. $F:S \rightarrow S$ is *expansive* if there exists $\nu > 0$ such that, for any two distinct points x and y in S , there exists an n such that $|F^n(x) - F^n(y)| > \nu$.

Expansiveness is a stronger form of topological sensitivity in that *all* nearby points eventually separate by at least ν . This is closer to the traditional notion of sensitive dependence (where all nearby points diverge exponentially).

Inherent regularity

In addition to having determinism in the evolution law, topological chaos obtains a degree of regularity from the *density of periodic points*. By definition, if periodic points are dense in the phase space of a system then there will be periodic points in every neighbourhood of every point in the space. This represents a degree of regularity in a system, since the perturbation of a non-periodic trajectory, for example, might cause it to hit one of the periodic points, thereby causing the trajectory itself to become periodic. Of course, if non-periodic points are also dense then the reverse is also true.

We remark, however, that if most periodic points are repelling, then this form of regularity will be hidden. Additionally, the periodic points may form a set of measure zero even though they would be infinite in number. In either case, the probability of

observing periodic behaviour in such a system would effectively be zero. Thus it seems that the value of dense periodic points as representing an element of regularity is somewhat limited.

Indecomposability

It is part of the scientific method that one should attempt to understand a system by breaking it down into smaller parts. In the context of dynamical systems theory, this corresponds to studying the properties of isolated regions of the phase space. However, if the dynamics of a system causes its phase space to be well *mixed*, the phase space becomes *indecomposable* and must be studied as a whole.

Devaney captures this notion of indecomposability by requiring that chaotic systems be *topologically transitive*, as given in the following.

(3.5.3) DEFINITION. $F:S \rightarrow S$ is said to be *topologically transitive* if for any pair of open sets U, V in S there exists a $k > 0$ such that $F^k(U) \cap V \neq \emptyset$.

This states that any region of the phase space will be deformed by the system until it overlaps with any other region. Intuitively, this corresponds to the system being well mixed.

Summary

Devaney's definition of chaos can be stated in the following

(3.5.4) DEFINITION. Let the map $F:S \rightarrow S$ be continuous. Then F is said to exhibit *topological chaos* if

- i) F is topologically sensitive;
- ii) the periodic points of F are dense in S , and;
- iii) F is topologically transitive over S .

This definition of chaos is appealing in that it tries to capture the intuitive notions of unpredictability, regularity, and mixing. Unfortunately, it is not useful in practice, since it is not possible to verify the requirements in a physical system. Topological sensitivity is more general than the traditional notion of sensitive dependence, as systems with only polynomial growth in errors (and thus a zero Lyapunov exponent) can be chaotic. In addition, the definition is somewhat redundant, since topological sensitivity is implied by the second two conditions (existence of a dense set of periodic points together with topological transitivity).³⁵

Finally, we note that the three prerequisites of topological chaos are each invariant

³⁵ R.L. Devaney, in a personal communication. Devaney also mentioned that this definition of chaos has the agreeable property of being easier to teach!

under homeomorphisms, and are thus said to be *topological properties* of the system. The equivalence of two systems under a homeomorphism is referred to as *topological conjugacy*. This invariance of topological chaos under homeomorphisms can be used to generate new chaotic systems; once an archetype chaotic system is found, other systems can be proved chaotic just by demonstrating topological conjugacy to the first.

3.6 Symbolic dynamics

Topological conjugacy plays an important role in dynamical systems theory as it allows the properties of one dynamical system to be determined from another system whose properties are already known. This fact is used in the study of *symbolic dynamics*, wherein the behaviour of a system is expressed in terms of an equivalent, simpler system defined over a space of *symbol sequences*. The power of symbolic dynamics lies in the fact that much can be learned about a dynamical system by studying the properties of the corresponding symbolic system.

In the symbolic system, each symbol sequence represents the *itinerary* of a trajectory in the original system. The basic technique is to prove that the two systems are topologically conjugate, and then establish the dynamical properties of the symbolic system. In so doing, the corresponding properties in the original system are also established. The technique of symbolic dynamics has been successfully used in the study

of several systems, including the logistic map [5, 34], the Lorenz equations [142], Smale's horse-shoe map [141], and even in number theory [136]. Symbolic dynamics is an important part of the study of chaotic systems, and has received much attention in the literature [2, 11, 55, 63, 69, 108, 119, 135, 159].

The first step in studying the symbolic dynamics of a system is to create the symbol space. To this end, let A be an alphabet of N symbols, and let Σ be the set of all infinite sequences of symbols from A :

$$A = \{\alpha_1, \alpha_2, \dots, \alpha_N\}$$

$$\Sigma = \{s = \langle s_0, s_1, s_2, \dots \rangle : s_i \in A \ \forall i \geq 0\}$$

Next, define the distance between any two elements x and y of Σ :

$$d(x, y) = \sum_{i=0}^{\infty} \frac{\delta_i}{N^i}$$

$$\delta_i = \begin{cases} 0 & \text{if } x_i = y_i \\ 1 & \text{if } x_i \neq y_i \end{cases}$$

It is straightforward to show that d is a metric on Σ [34]. Endowed with this metric, Σ is then called the *sequence space on N symbols*.

The full shift on N symbols

The next step is to find an agreeable mapping of Σ to itself that can be shown to

be equivalent to the evolution law of the original dynamical system. The map of choice is called the *full shift on N symbols*:

$$\begin{aligned}\sigma: \Sigma &\rightarrow \Sigma \\ s = \langle s_0 s_1 s_2 \dots \rangle &\in \Sigma \\ \sigma(s) &= \langle s_1 s_2 s_3 \dots \rangle\end{aligned}$$

The action of σ is to simply shift every element of the sequence s left one position, discarding the first element.

We now describe the dynamics of σ in terms of the different types of elements in Σ . To be consistent with earlier usage, we refer to an element x of Σ as a *point*, and the sequence of elements $\langle x, \sigma(x), \sigma(\sigma(x)), \dots \rangle$ as its *trajectory*. A point x of Σ is said to be *periodic* if $\sigma^k(x) = x$ for some $k > 0$. The period of a point is the smallest value of k for which this is true.

The periodic points in Σ are those sequences containing a repeating pattern of symbols. For example, the period-1 (fixed) points are those sequences consisting of a single symbol, the period-2 points are those consisting of an alternating pattern of two symbols, and so forth. In general, the periodic points are those consisting of a repeating pattern of symbols. *Eventually periodic* points contain an initial pattern of symbols, followed by some other pattern which repeats for the remainder of the sequence. Together, the periodic and eventually periodic points form a countably infinite subset of Σ .

The *non-periodic* points in Σ are those sequences having a non-repeating pattern of symbols. For example, the decimal expansion of π would form such a sequence, since it is non-repeating. Recalling that the irrational numbers have non-repeating expansions in any base, we can map the non-periodic points in Σ to the irrationals in $[0,1]$ by simply placing a "decimal point" in front of the sequence. Since the irrational numbers in $[0,1]$ are uncountable, we conclude that the non-periodic points form an uncountable subset of Σ .

Finally, we observe that Σ contains points whose trajectories wind densely throughout the set; that is, the trajectory of such a point would come arbitrarily close to every other point. We can construct such a sequence x by concatenating all patterns of k symbols, for increasing k . Then for each element y of Σ , there will be some iterate of x that approximates y to any number of digits desired. With this, we have just proved the following standard result [34, 159].

(3.6.1) THEOREM. [Devaney, 1989]. The shift map σ acting on the space of infinite sequences of symbols from a finite alphabet, Σ , has

- i) a countable infinity of periodic orbits of arbitrarily high period;
- ii) an uncountable infinity of non-periodic orbits;
- iii) a dense orbit.

We are also now able to prove the following results.

2

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

PIONEERS IN METHYLENE BLUE TESTING SINCE 1974

MICROD
CALCULATORS

10000 COUNTRY ROAD 6, BURNINGWELL, WY 82004 USA
TEL: 812 646 7887 FAX: 812 625 7887 TLX: 618000000

(3.6.2) PROPOSITION.[†] [Devaney, 1989]. $\sigma: \Sigma \rightarrow \Sigma$ is topologically chaotic.

PROOF.

Topological sensitivity follows from the fact that the trajectories of any two distinct points must eventually separate, under the metric d , by a distance of one half. Secondly, the periodic points are clearly dense in Σ , since every element of Σ can be represented by a convergent sequence of periodic points. Finally, the existence of a dense trajectory guarantees that σ is topologically transitive, completing the proof. ■

(3.6.3) PROPOSITION.[†] [Devaney, 1989]. $\sigma: \Sigma \rightarrow \Sigma$ has sensitive dependence on initial conditions.

PROOF.

By the nature of σ , the distance between any two nearby points grows by a factor of N with each iteration, resulting in the exponential divergence of nearby trajectories and a positive Lyapunov exponent of $\log N$. Thus σ is expansive, and in particular, has sensitive dependence on initial conditions. ■

The full shift is appealing from a theoretical point of view because it allows one to determine the number and identity of all of the periodic orbits, and is easily seen to be chaotic. Although it might appear to be an artificially abstract system, the dynamics of the shift have been found in many other systems, including the Lorenz system [96, 142, 143], the Smale horseshoe map [34, 141], the Hénon map [29], and the logistic map [86]. Together with *stretching and folding*, the notion of *shift dynamics* has become a

metaphor for chaotic behaviour.

The tent map

To illustrate the use of topological conjugacy in studying the symbolic dynamics of a system, we consider a system known as the *tent map*. Letting I be the unit interval $[0,1]$, we define the tent map by:

$$F:I \rightarrow \mathbb{R}$$

$$F(x) = \begin{cases} mx & \text{if } x \leq \frac{1}{2} \\ m-mx & \text{if } x > \frac{1}{2} \end{cases}$$

with $m > 2$. Label the intervals $[0, 1/m]$ and $[1-1/m, 1]$ as I_0 and I_1 respectively. Each of these intervals is mapped onto the entire set I . However, the points lying between I_0 and I_1 escape from I in one iteration of the map. Similarly, I_0 and I_1 contain central subsegments that escape from I in two iterations. We consider the *invariant set* of F , namely those points in I whose trajectories do not escape the set.

The invariant set, denoted by Λ , may be defined as the infinite intersection of the sub-intervals that do not escape I under higher and higher iterations of F . Given I_0 and I_1 above, we define

$$\begin{aligned}
I_{00} &= \{x \in I_0 : F(x) \in I_0\} \\
I_{01} &= \{x \in I_0 : F(x) \in I_1\} \\
I_{10} &= \{x \in I_1 : F(x) \in I_0\} \\
I_{11} &= \{x \in I_1 : F(x) \in I_1\}
\end{aligned}$$

which generalizes to

$$\begin{aligned}
I_{s_0 s_1 \dots s_n} &= \{x \in I_{s_0} : F(x) \in I_{s_1}, F^2(x) \in I_{s_2}, \dots, F^n(x) \in I_{s_n}\} \\
&= I_{s_0} \cap F^{-1}(I_{s_1}) \cap F^{-2}(I_{s_2}) \cap \dots \cap F^{-n}(I_{s_n}) \\
&= I_{s_0} \cap F^{-1}(I_{s_1 s_2 \dots s_n})
\end{aligned}$$

Each point in Λ is uniquely represented by the infinite intersection

$$I_{s_0 s_1 \dots} = \bigcap_{n \geq 0} I_{s_0 s_1 \dots s_n}$$

for a particular infinite sequence $s_0 s_1 \dots s_n \dots$. The set itself can thus be defined as

$$\Lambda = \bigcup_{s_0 s_1 \dots} I_{s_0 s_1 \dots}$$

where the union is taken over all combinations of $s_0 s_1 \dots s_n \dots$. As the construction of Λ parallels the construction of the triadic Cantor set, we see that Λ is itself a Cantor set having measure zero and an uncountable number of points.

Next, we construct a map from Λ to the sequence space on two symbols, denoted by Σ . Observe that $\Lambda \subset I_0 \cup I_1$, thus for every x in Λ all points in the trajectory of x also lie in Λ and thus on either I_0 or I_1 . This leads to the following

(3.6.4) DEFINITION. The *itinerary* of a point x in I is the sequence $S(x) = \langle s_0, s_1, \dots \rangle$, where

$$s_i = \begin{cases} 0 & \text{if } F^{-i}(x) \in I_0 \\ 1 & \text{if } F^{-i}(x) \in I_1 \end{cases}$$

The itinerary of a point is an infinite binary sequence, making S a map from Λ to Σ . In fact, S can be shown to be a *homeomorphism* implying that Λ and Σ are structurally equivalent sets. Furthermore, letting σ denote the full shift on two symbols, we have the following.

(3.6.5) THEOREM. [Devaney, 1989]. $S \circ F = \sigma \circ S$

PROOF.

Any point x in Λ can be uniquely defined by the nested sequence of intervals

$$\bigcap_{n \geq 0} I_{s_0 s_1 \dots s_n}$$

where $s_0 s_1 \dots s_n$ is itinerary $S(x)$. Now

$$\begin{aligned} I_{s_1 s_2 \dots} &= I_{s_0} \cap F^{-1}(I_{s_1 s_2 \dots}) \\ \Rightarrow F(I_{s_1 s_2 \dots}) &= I_{s_0 s_1 s_2 \dots} \end{aligned}$$

Thus

$$\begin{aligned} SF(x) &= SF\left(\bigcap_{n \geq 0} I_{s_0 s_1 \dots s_n}\right) \\ &= S\left(\bigcap_{n \geq 0} I_{s_1 s_2 \dots}\right) \\ &= s_1 s_2 \dots = \sigma S(x) \end{aligned}$$

completing the proof. ■

The functions $F:\Lambda\rightarrow\Lambda$ and $\sigma:\Sigma\rightarrow\Sigma$ are thus topologically conjugate via the itinerary function $S(x)$. This implies that F on Λ and σ on Σ have completely equivalent dynamics. In particular, F on Λ has

- i) a countable infinity of periodic orbits of arbitrarily high period;
- ii) an uncountable infinity of non-periodic orbits;
- iii) a dense orbit.

Additionally, F on Λ is topologically chaotic and has sensitive dependence on initial conditions.

It is important to note that this theorem applies only to $F:\Lambda\rightarrow\Lambda$, and not to $F:I\rightarrow\mathbb{R}$. That is, the theorem states that the action of F over its invariant set is equivalent to a shift map; it does not state that the action of F over the entire interval is equivalent to a shift map. We remark that the trajectory of any point not on Λ will eventually leave the interval I . Since Λ is a totally disconnected set (as are all Cantor sets), any discrepancy in the initial state would make the chaotic nature of F effectively unobservable.

The typical application of symbolic dynamics is to prove that a given dynamical system is topologically conjugate to a shift map. This concept is used in the definition of formal chaos, which we discuss next.

3.7 Formal chaos

In this section we present two definitions of chaos that involve the invariant set of a dynamical system, and the property that the chaotic behaviour may be effectively unobservable. The term *formal chaos* (as opposed to observable chaos) was proposed by Oono and Osikawa [111] to capture this notion.

The first definition is due to Wiggins [159]:

(3.7.1) DEFINITION. [Wiggins]. Suppose F is the evolution rule of a dynamical system with Λ as its closed invariant set. Then Λ is said to be *chaotic* if F has topological sensitivity over Λ and is topologically transitive on Λ .³⁶

This is similar to the definition of topological chaos, except that

- i) chaos is defined for a set of points, rather than for an evolution rule;
- ii) topological sensitivity and transitivity need only hold over a closed invariant set, rather than the entire space, and;
- iii) periodic points need not be dense.

Point i) seems counter-intuitive since one normally thinks of a *process* as being chaotic,

³⁶ Wiggins originally used the term "sensitive dependence on initial conditions" in the sense of Devaney, which we have termed "topological sensitivity".

rather than a *set* being chaotic.³⁷ Point ii) allows for chaotic behaviour to occur on a set that may have measure zero (as with the tent map), and thus be effectively unobservable. Point iii) dispenses with the attempt at preserving the notion of underlying regularity in a chaotic system.

The following definition is based on that given by Oono and Osikawa [111]. We give the definition for a continuous map F defined over some closed real interval I .³⁸

(3.7.2) DEFINITION. [Formal chaos]. F is formally chaotic if there exist some positive integer M and F^M -invariant set $\Lambda(\subset I)$ and one-to-one map $\psi: \Sigma \rightarrow \Lambda$ such that $F^M \cdot \psi = \psi \sigma$, where Σ is the sequence space on two symbols, and σ be the corresponding shift.

Observations:

- i) This is similar to requiring that F over Λ be topologically conjugate to σ , except that the one-to-one correspondence need not be continuous. It is not clear what implications this lack of continuity would have.
- ii) F itself need not have an invariant set, provided some higher iterate of F does have an invariant set. This appears to be a reasonable generalization.

³⁷ It would seem that the author chose this definition in order to define "strange attractors" as attractors that are "chaotic".

³⁸ The original definition gives an added condition to cover the case when F is not continuous. We choose to avoid this complication here.

3.8 Sarkovsky's theorem

A famous theorem due to the Russian mathematician A.N. Sarkovsky describes the relationships between the sizes of periodic orbits in a one-dimensional real map [134]. In presenting the theorem, it is useful to define the *Sarkovsky ordering* of the natural numbers:

$$3 \triangleright 5 \triangleright 7 \triangleright \dots \triangleright 2(3) \triangleright 2(5) \triangleright \dots \triangleright 2^2(3) \triangleright 2^2(5) \triangleright \dots \triangleright 2^n(3) \triangleright 2^n(5) \triangleright \dots \triangleright 2^m \triangleright 2^{m-1} \triangleright \dots \triangleright 2^2 \triangleright 2 \triangleright 1$$

This ordering lists the odd numbers first (excluding 1), followed by two times the odds, followed by higher powers of two times the odds, followed finally by all the powers of two in descending order. The notation $a \triangleright b$ indicates that a appears to the left of b in the ordering.

(3.8.1) THEOREM. [Sarkovsky]. Suppose F is a continuous map over the reals. If F has an orbit of primary period p , then F also has orbits of primary period q , for all q such that $p \triangleright q$ in the Sarkovsky ordering.

PROOF. See [34, 111]. ■

The theorem applies only to one-dimensional real maps, although some authors have claimed to have found a two-dimensional analogue [71].

Observe that Sarkovsky's result is topological in nature, since it establishes the existence of orbits having specific periods, but does not identify the orbits themselves. A consequence of the theorem is that if the function f has only finitely many orbits, then they must all have periods that are powers of two. Conversely, if f has an orbit whose period is not a power of two, then it must have infinitely many periodic points. In particular, if f has a period-3 orbit, then it has orbits of all other periods. This is related to the Li/Yorke definition of chaos, which we present next.

3.9 Li/Yorke Chaos

In 1975 Tien-Yien Li and James A. Yorke published a paper entitled "Period Three Implies Chaos" which has become a standard reference in the literature [94]. In fact, the first use of the term "chaos" has been attributed to this paper [101, 131].³⁹

The theorem of Li and Yorke has two main parts. The first states that if a continuous map of an interval has a point of period three, then there exist periodic points of all other periods. This part of their result is actually a consequence of Sarkovsky's theorem, although it was discovered independently [34, 134]. In 1978 Li and Yorke's result was extended by Oono, who showed that any period other than a power of two implied chaos [110]. Of course, Oono's result is also a consequence of Sarkovsky's

³⁹ It would appear from the paper that the authors used the term "chaos" loosely, as it did not appear in any theorem. They probably did not expect it to become as popular as it has!

theorem.

The second part of the theorem states that, given the existence of a period three orbit, the map must possess an uncountable infinity of non-periodic orbits. This is the behaviour they call *chaos*. It is interesting to note that this is the same phenomenon discovered earlier by Lorenz, albeit in a system of differential equations and not in a one-dimensional map [96, 97, 142]. We present the theorem below, without proof.

(3.9.1) THEOREM. [Li & Yorke, 1975]. Let J be an interval and let $F: J \rightarrow J$ be continuous. Assume there is a point $a \in J$ for which the points $b=F(a)$, $c=F^2(a)$ and $d=F^3(a)$, satisfy $d \leq a < b < c$ (or $d \geq a > b > c$). Then

T1: for every $k=1,2,\dots$ there is a periodic point in J having period k .

Furthermore

T2: there is an uncountable set $S \subset J$ (containing no periodic points), which satisfies the following conditions:

(A) For every $p, q \in S$ with $p \neq q$,

$$\limsup_{n \rightarrow \infty} |F^n(p) - F^n(q)| > 0$$

and

$$\liminf_{n \rightarrow \infty} |F^n(p) - F^n(q)| = 0$$

(B) For every $p \in S$ and periodic point $q \in J$,

$$\limsup_{n \rightarrow \infty} |F^n(p) - F^n(q)| > 0$$

■

Note that although the theorem does not actually require a point of period three, this is the simplest condition that will satisfy the hypothesis.

Condition (A) implies that the orbits of any two non-periodic points can come arbitrarily close to each other (via the infimum), but that they do not stay close (via the supremum). Condition (B) guarantees that none of the aperiodic points are even asymptotically periodic.⁴⁰ Note that even though the set S of non-periodic orbits is uncountable, it may have measure zero.⁴¹ Thus a system may be chaotic in the sense of Li and Yorke, even if almost all points lie on periodic orbits [98].

It is important to note that the theorem of Li and Yorke, as with Sarkovsky's theorem, only applies to continuous maps of an interval. In [34], Devaney gives an interesting discussion of the periodic points in a map containing an attracting period three orbit, using the technique of symbolic dynamics.

⁴⁰ An asymptotically periodic point is one whose orbit converges to an existing periodic orbit, without actually becoming periodic itself.

⁴¹ Recall that the triadic Cantor set is an uncountable measure-zero set.

3.10 Strange attractors

The term *strange attractor* has been in use for almost twenty years, and yet there is no generally accepted definition of the term. It is sometimes used to describe the geometry of the attractor in a system, and sometimes to describe the sensitivity of trajectories on the attractor, and sometimes it is used to denote both. We begin the discussion by defining what is meant by the term *attractor*. The following definitions, for a continuous map $F:\mathbb{R}^n \rightarrow \mathbb{R}^n$, are based on those given by Wiggins [159].

(3.10.1) DEFINITION. Let $S \subset \mathbb{R}^n$ be a set. Then S is an *invariant set* if, $\forall x \in S, F^n(x) \in S \forall n \geq 0$.

(3.10.2) DEFINITION. A closed invariant set S is *topologically transitive* if, for any two open sets $U, V \subset S$, $\exists n \geq 0$ such that $F^n(U) \cap V \neq \emptyset$.

(3.10.3) DEFINITION. A closed invariant set $A \subset \mathbb{R}^n$ is called an *attracting set* if there is some neighbourhood U of A such that $\forall x \in U \exists n \geq 0$, such that $F^n(x) \in U$ and $F^n(x) \rightarrow A$ as $n \rightarrow \infty$.

(3.10.4) DEFINITION. An *attractor* is a topologically transitive attracting set.

Intuitively, an attractor is an indecomposable set representing the long-term behaviour of

the dynamical system. Devaney [34] gives a similar definition, although he omits the transitivity requirement.

Geometry versus behaviour

It is a standard result in dynamical systems theory that in one-dimensional flows the only structurally stable⁴² attractors are sinks (stable fixed points), and that in two-dimensions the only structurally stable attractors are sinks and stable limit cycles [5, 145, 159]. Until the 1960's it was known that in three dimensions the structurally stable attractors included sinks, stable limit cycles and torii, but it was not known if there were any others [145, 159].

In 1967 Stephen Smale [141] showed that in three-dimensional flows there can also exist structurally stable attractors that are not one of these types. In other words, he had found that fractal attractors could exist in continuous dynamical systems. Smale also demonstrated the existence of a fractal invariant set in his two-dimensional *horseshoe map* of the plane, and although this set was not attracting, it did illustrate how such features can come about. In fact, an attractor of this type was observed in 1963 by Edward Lorenz [96, 142]. More recent observations can be found in [127, 137, 147, 148, 150].

The term *strange attractor* was coined in 1971 by David Ruelle and Floris Takens

⁴² Recall that a dynamical feature is *structurally stable* if it persists under small perturbations to the evolution rule.

[132, 133], in reference to the features discovered by Smale. Since then, many authors have come to use the term *strange attractor* in this geometric sense [7, 34, 62, 101, 143]. However, various authors also define an attractor to be *strange* if the underlying dynamical system is chaotic on the attractor [37, 38, 52, 78, 79, 113, 127, 159]. A representative definition is thus given by Wiggins [159]:

These two views of strange attractors are complimentary; the fact that an attractor is fractal does not prevent it from being chaotic, and vice versa. Indeed, most chaotic systems that were studied early on had attractors that were both fractal and chaotic, leading many authors to use the term *strange* to mean both at once [24, 58, 59, 128, 129]. This view is not unreasonable, as both do occur in dissipative dynamical systems as we now describe.

Consider a chaotic n -dimensional dissipative dynamical system. Because the system is dissipative, there will be a net contraction in the volume of any n -dimensional phase space element; thus the long-term behaviour of the system will lie on a limit set of zero volume. Because the system is chaotic, however, there will be a stretching of phase space elements in at least one direction, as indicated by the presence of at least one positive Lyapunov exponent. We remark that, because the system is dissipative, the sum of the n Lyapunov exponents must be negative.

Now, in an unbounded system, the stretching and contraction could continue in

their respective directions forever, resulting in a region of phase space that is at once infinitely 'long' and infinitesimally 'thin'. However, in a *bounded* dynamical system, a region of phase space can be stretched only so far in one direction, and must eventually fold back on itself. The long-term behaviour of the system is thus represented by a region of phase space (the limit set) that has undergone an infinite amount of stretching and folding. In effect, the phase space has undergone the same kind of process used to construct the fractal sets described in Chapter 2. Thus, provided the limit set is attracting, it will be a strange attractor that is at once fractal and chaotic.

We note, with interest, that there exist dynamical systems having invariant sets that are either fractal or chaotic, but not both.

i) The logistic map

$$x_{n+1} = cx_n(1-x_n), \quad x_n \in [0,1]$$

has a non-chaotic fractal invariant set⁴³ at the parameter value $c=3.569944\dots$, for which the Hausdorff dimension is approximately 0.538 [59, 62]. However, the invariant set is not structurally stable, and disappears under small perturbations to the parameter c .

ii) The logistic map also has a chaotic non-fractal invariant set at $c=4$, namely the entire unit interval [95, 168]. However, it is debatable whether this set can be considered to be an attractor. In fact, since the invariant set

⁴³ The invariant set occurs at the accumulation point of the sequence of bifurcations leading up to the parameter value $c=3.56944\dots$. It consists of the periodic orbits having periods that are all the powers of two, with the "largest" of these orbits being attracting.

comprises the entire phase space, there are no points "outside" it for the set to attract. This invariant set is also structurally unstable.

- iii) Grebogi *et al* [62] have devised a two-dimensional map having a non-chaotic fractal attractor that is structurally stable.

Detecting strange attractors

Strange attractors (of either variety) can be detected in physical systems by the analysis of time series data. The observance of a trajectory having a positive Lyapunov exponent is evidence of chaotic attractor [14, 38, 41, 49, 92, 165]. On the other hand, the determination that a trajectory traces out a set of points having fractional dimension suggests the existence of a fractal attractor [6, 40, 60, 61, 75, 77, 103, 153]. The correlation and information dimensions are typically used, as they are empirical in nature and are easier to compute from time series data than the more familiar box-counting dimension [7].

Reconstructing strange attractors

Fractal attractors can be detected through dimensional estimates, but they can also be reconstructed pictorially with the *method of delays* devised in 1980 by Packard and Takens [28, 114, 123]. The Packard/Takens method (as it is also called) takes a time series of a single observable and reconstructs the corresponding trajectory, and hence the

underlying attractor, if there is one. If the system has only one degree of freedom, then the observed trajectory is embedded in one-dimensional space and can be plotted directly. If the number of degrees of freedom is not known, as is generally the case, then there is the problem of finding the proper dimension in which to "embed" the trajectory.⁴⁴ Various methods exist to calculate the embedding dimension [3, 40, 49, 150].

The Packard/Takens method reconstructs the trajectory in n -dimensional space by creating $n-1$ 'fake' observables, from shifted values in the time series. Given a time series of the form $\langle x_1, x_2, \dots, x_n, \dots \rangle$, the method reconstructs points in n -dimensions as

$$\begin{aligned} &\langle x_1, x_2, \dots, x_n \rangle \\ &\langle x_2, x_3, \dots, x_{n+1} \rangle \\ &\vdots \\ &\langle x_k, x_{k+1}, \dots, x_{k+n-1} \rangle \\ &\vdots \end{aligned}$$

Packard and Takens showed that, under some reasonable conditions, the resulting set has the same topological features as that defined by the original trajectory. Thus, if the reconstructed set has a fractal appearance, then there is strong evidence that the original attractor also has a fractal structure.

As an example, we briefly discuss the strange attractor discovered in 1963 by Edward Lorenz [96]. Lorenz developed an idealized model of a hydrodynamic system

⁴⁴ This explains the use of the term *embedding dimension*, often found in the literature. By definition, the value of the embedded dimension is always an integer (like the Euclidean dimension), as it represents the number of co-ordinates required to uniquely specify the system state.

in which a thin layer of fluid was held between two heated metal plates.⁴⁵ His model comprised a set of three ordinary differential equations:

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y-x) \\ \frac{dy}{dt} &= rx-y-xz \\ \frac{dz}{dt} &= xy-bz\end{aligned}$$

In general terms, x represents the rate of convection, whereas y and z represent the horizontal and vertical temperature variations respectively (the remaining terms are constants) [143].

The strange attractor contained within this system is now known as the *Lorenz attractor*, which is shown in Figures 3.1 and 3.2. The reconstructed attractor is shown in Figure 3.3, and is seen to have the same general features as the original attractor.

⁴⁵ In this context 'thin' implies an idealized two-dimensional fluid having zero depth.

Figure 3.1 Lorenz attractor projected onto the X-Y and Y-Z planes

Figure 3.2 The Lorenz attractor projected into the X-Z plane

Figure 3.3 Reconstruction of the Lorenz attractor

3.11 Summary

There is no consensus regarding the definition of chaos. Generally, the term is associated with the random-like behaviour appearing in certain dynamical systems. What is certain is that chaos occurs in *dynamical systems*, which are systems that follow deterministic time-independent evolution laws. Chaos is not associated with stochastic systems. In this chapter we have discussed the most common definitions of chaos, the first being

- *sensitive dependence on initial conditions*, wherein nearby trajectories diverge at an exponential rate on average, as indicated by the existence of positive Lyapunov exponents.

In practice, the existence of sensitive dependence in a physical system makes accurate long-term predictions impossible. Chaotic behaviour is expressed in a more theoretical setting as

- *topological chaos*, being the combination of unpredictability (through topological sensitivity), inherent regularity (through the density of periodic points), and indecomposability (through topological transitivity).

Demonstrating the existence of topological chaos in a dynamical system requires knowledge of its evolution rule. However, once this rule is known it is often easier to

prove topological chaos than it is to prove sensitive dependence.

Sensitive dependence and topological chaos are both invariant under homeomorphisms, making them *topological properties* of a system. This allows one to prove the chaotic nature of a system (in either sense) by showing that it is *topologically conjugate* to another system that is known to be chaotic. This is the basis for the study of

- *symbolic dynamics*, wherein the phase space of the system is shown to be homeomorphic to a space of infinite sequences, over which the corresponding evolution rule acts in a simple way.

The usual strategy in studying the symbolic dynamics of a system is to prove topological conjugacy to the

- *shift map*, which simply shifts the symbols in an infinite sequence left by one position, discarding the first symbol in the sequence.

The dynamics of the shift map include: an uncountable number of aperiodic orbits, an infinite number of periodic orbits of all periods, and a dense orbit. The shift map has sensitive dependence on initial conditions, and is also topologically chaotic. Thus any system that is topologically conjugate to the shift will also have these properties. A more general definition captures the notion of

- *chaos over the invariant set*, wherein topological conjugacy to an existing

chaotic system (usually the shift map) can only be demonstrated when the action of the system is restricted to its invariant set.

In the case where the invariant set is attracting, it is often referred to as a

- *strange attractor*, usually occupying a zero-volume fractal subset of the phase space of the system.⁴⁶

Even if the invariant set is repelling, thereby making the chaotic behaviour effectively unobservable, the system is still said to exhibit

- *formal chaos*, wherein the system acting over its invariant set is topologically conjugate to the shift map.

An early definition, restricted to continuous one-dimensional maps, is commonly referred to as

- *Li/Yorke chaos*, where a map having an orbit of period three is proved to have an uncountable number of aperiodic orbits as well as periodic orbits of all periods.⁴⁷

Finally, we remark on the existence of randomness in the behaviour of chaotic

⁴⁶ Note that the term *strange attractor* is also used to denote attracting invariant sets that are fractal, but not chaotic.

⁴⁷ There is a sense in which the chaos of Li and Yorke is equivalent to the chaos found in the shift map, except that asymptotically periodic orbits are not present.

dynamical systems. Contrary to traditional belief, random behaviour is not necessarily due to the complexity of the evolution law. The complexity is in fact due to the continuum of real values that can be taken on by the system [16, 17, 18, 19, 50, 93]. Indeed, much of the interest in chaos theory is that complex behaviour can come from very simple rules [101].

This behaviour is best illustrated by the shift map. Since almost all initial conditions correspond to random real values [19], any finite precision representation will contain within it some random uncertainty. The shift map amplifies this uncertainty by an order of magnitude with each iteration, until it grows to the size of the phase space itself. Thus the long-term behaviour of the system will depend only on the initial uncertainty and not on the actual initial measurement. The existence of sensitive dependence on initial conditions, together with a continuum of random initial conditions, means that no finite-precision computation can accurately follow any trajectory in a chaotic system.

PART II: THE HAILSTONE NUMBERS

4. The Hailstone Function

4.1 Motivation

We now discuss a number theoretic system which produces integer sequences commonly known as *hailstone numbers*. The system itself, which we refer to as the *hailstone system*, has a simple definition yet the sequences it generates can behave in a seemingly erratic fashion. An open conjecture associated with this system has resisted analysis for many years.

Our main interest in the hailstone system is due to the claim by C. Agnes and M. Rasetti that the hailstone system is chaotic [1], and that the related conjecture is undecidable. After examining the properties of the hailstone function, we discuss whether it can be said to exhibit chaotic-like behaviour over the natural numbers.

4.2 A dynamical system on the natural numbers

We examine the fully discrete dynamical system defined by the recurrence relation

$$X_{n+1} = H(X_n) = \begin{cases} \frac{X_n}{2} & \text{if } X_n \text{ is even} \\ 3X_n + 1 & \text{if } X_n \text{ is odd} \end{cases} \quad (4.1)$$

where the values X_n are natural numbers. Given an initial value X_0 , the system generates a sequence of positive integers that tend to rise and fall in value in a way that has been compared to the movement of hailstones in a storm cloud [72, 120]. We shall refer to (4.1) as the *hailstone system*, and H as the *hailstone function*.

(4.2.1) DEFINITION. The *hailstone sequence* $Hs(X_0)$ for a natural number X_0 is the sequence of values $\langle X_0, X_1, \dots, X_n, \dots \rangle$ produced by the recurrence relation (4.1).

Table 1 shows the hailstone sequences for the initial values $\{1, 2, \dots, 10\}$. Observe that every odd number X_n in a hailstone sequence is followed by the value $3X_n + 1$, which is always even. On the other hand, the even numbers in a sequence can be followed by values that are either even or odd. Hailstone sequences are thus composed of runs of even numbers, separated by single odd numbers. The sequences shown in Table 1 are all eventually periodic, since the cycle $\langle 1, 4, 2 \rangle$ is endlessly repeated in every case.

X	Hs(X)
1	<1,4,2, 1,4,2, ...>
2	<2,1,4,2, 1,4,2, ...>
3	<3,10,5,16,8,4,2, 1,4,2, ...>
4	<4,2,1,4,2, 1,4,2, ...>
5	<5,16,8,4,2, 1,4,2, ...>
6	<6,3,10,5,16,8,4,2, 1,4,2, ...>
7	<7,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2, 1,4,2, ...>
8	<8,4,2,1,4,2, 1,4,2, ...>
9	<9,28,14,7,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2, 1,4,2, ...>
10	<10,5,16,8,4,2, 1,4,2, ...>

Table 1 Hailstone sequences for the initial values {1,2,...,10}

4.3 The $3x+1$ Conjecture

It is an open conjecture that all hailstone sequences contain the value 1, and thus endlessly repeat the cycle <1,4,2>. We present this as the " $3x+1$ " conjecture, below:

$3X+1$ CONJECTURE:

For every hailstone sequence $Hs(X_0)=\langle X_0, X_1, \dots \rangle$ there exists an index n such that $X_n=1$. Thus every hailstone sequence is eventually periodic, and endlessly repeats the cycle <1,4,2>.

The conjecture appears to have originated with Lothar Collatz of the University of Hamburg, who worked on similar number-theoretic functions during the 1930's. By the mid 1960's the conjecture had appeared in a variety of forms and was known by several names, including the *Collatz problem*, the *Syracuse problem*, *Ulam's problem* and the *$3x+1$ conjecture*. The survey paper by Lagarias [89] provides an excellent background on this problem.

One way of visualizing the dynamics of the hailstone system is through a *hailstone sequence plot*, which is a graph of the function $F(n)=H^n(X_0)$ for some number X_0 , where as usual $H^n(X_0)$ represents the n -th iterate of X_0 . A hailstone sequence plot can be thought of as a "time series" plot, where the indices of sequence elements represent the time steps. Figure 4.1 shows the sequence plot for $X_0=25$; this sequence reaches the cycle $\langle 1,4,2 \rangle$ after 23 steps. Figure 4.2 shows the sequence plot for $X_0=26$; this sequence reaches the cycle $\langle 1,4,2 \rangle$ after only 10 steps. In marked contrast, the sequence plot for $X_0=27$ (shown in Figure 4.3) requires 111 steps to reach the cycle $\langle 1,4,2 \rangle$, and achieves a maximum value of 9232.

Figure 4.1 Hailstone sequence plot for $X_0=25$

Figure 4.2 Hailstone sequence plot for $X_0=26$

Figure 4.3 Hailstone sequence plot for $X_0=27$

We will now consider the structure of hailstone sequences, with respect to the number of steps required to reach the $\langle 1, 4, 2 \rangle$ cycle, and the maximum values contained in the sequence. To assist in the discussion, some terminology will now be introduced.

(4.3.1) DEFINITION. The *stopping time* of the number X is the smallest number i such that $H^i(X) < X$, and is denoted by $\tau(X)$.⁴⁸

The notion of stopping time can be used to reformulate the $3x+1$ conjecture, as follows:

$3X+1$ CONJECTURE (SECOND FORM): Every positive integer has finite stopping time.

This form of the conjecture implies the first form, by induction. In fact, if every natural number X has a finite stopping time, then the hailstone sequence for X must contain some value X' smaller than X . Similarly, the hailstone sequence for X' must contain a number smaller than X' ; by repeating this argument, 1 must be contained in the sequence.

A theorem due to R. Terras indicates that almost all natural numbers have finite stopping times [151, 152]. The theorem follows, without proof.

(4.3.2) THEOREM. [Terras]. Let $F(k)$ be the *stopping time distribution function*, giving

⁴⁸ The definition of stopping time given by Lagarias differs slightly from ours, as it was based on the function $T(n) = (3n+1)/2$ for n odd.

the density of those integers X whose stopping times $\tau(X) \geq k$; that is

$$F(k) = \lim_{m \rightarrow \infty} \frac{1}{m} \left| \{X \leq m : \tau(X) \geq k\} \right|$$

Then for every positive integer k the limit defining $F(k)$ exists, and goes monotonically to zero as k goes to infinity. ■

Although this result does not prove the conjecture it does show that, if there are natural numbers that generate *aperiodic* hailstone sequences, then these numbers form a set of zero density among the natural numbers.

(4.3.3) DEFINITION. The *total stopping time* of the number X is the smallest number k such that $H^k(X)=1$, and is denoted by $\tau_-(X)$.

As seen in Figures 4.1 through 4.3, the total stopping time can vary significantly even for consecutive numbers. In particular, it is not necessarily the case that if $Y > X$ the hailstone sequence for Y will reach the value 1 after the hailstone sequence for X .

The following original result establishes a lower bound on the value of the total stopping time function for a given number X .

(4.3.4) PROPOSITION. The total stopping time of the hailstone sequence $Hs(X)$ is bounded below by $\log_2 X$. That is, $\tau_-(X) \geq \log_2 X$ for all $X \in \mathbb{N}$.

PROOF.

Consider the hailstone sequence $Hs(X)$ where $X \in \mathbb{N}$. Let $p(X)$ be the minimum value of k such that $H^k(X)$ is a power of two. That is

$$p(X) = \min \{k \geq 0: H^k(X) = 2^j, j \geq 0\}$$

Define $P_n \subset \mathbb{N}$ as the set of $X \in \mathbb{N}$ for which $p(X) = n$. The proof proceeds by induction on n . The proposition is true for $n=0$, since all $X \in P_0$ are powers of two, hence $\tau_-(X) = \log_2 X$. Assume that the proposition is true for $n=k$; that is, $\forall X \in P_k, \tau_-(X) \geq \log_2 X$. Letting X be an element of P_{k+1} , we have $H(X) \in P_k$. The assertion holds if X is an even number, since

$$\begin{aligned} \tau_-(X) &= \tau_-(H(X)) + 1 \\ \Rightarrow \tau_-(X) &= \tau_-\left(\frac{X}{2}\right) + 1 \geq \log_2\left(\frac{X}{2}\right) + 1 = \log_2(X) \end{aligned}$$

Similarly, for X odd,

$$\begin{aligned} \tau_-(X) &= \tau_-(H(X)) + 1 \\ \Rightarrow \tau_-(X) &= \tau_-(3X+1) + 1 \geq \log_2(3X+1) + 1 \geq \log_2(X) \end{aligned}$$

Hence for all $X \in P_{n+1}$ we have $\tau_-(X) \geq \log_2 X$, completing the proof. ■

Although no number X can have a total stopping time that is less than $\log_2 X$, there are numbers whose total stopping times are much greater than $\log_2 X$. Consider, for example, the sequence plot for $X=27$ shown in Figure 4.3. In this case, the total stopping time is 111, while the lower bound established in the proposition yields only $\log_2 27 < 5$. Figure 4.4 plots the total stopping times for $X \leq 1000$; no obvious pattern is apparent.

Figure 4.4 Total stopping times for $X \leq 1000$

Status of the conjecture

Lagarias reports that the $3x+1$ Conjecture has been tested by Nabuo Yoneda for all values of X up to 2^{40} , with no counter-example being found [89]. Should the conjecture be false, there will exist a number whose hailstone sequence does not reach the $\langle 1,4,2 \rangle$ cycle. Such a sequence would either terminate in a different cycle, or be aperiodic. A result by Lynn Garner indicates that if there exists a sequence that enters a cycle not containing 1, then the sequence must contain many thousands of terms [51].

4.4 Is the $3x+1$ Conjecture Undecidable?

Like many problems in number theory, the $3x+1$ conjecture is easy to state but very difficult to analyze. One researcher reported that "For about a month everybody at Yale worked on it with no result... A joke was made that this problem was part of a conspiracy to slow down mathematical research in the U.S." [89].

It may be that the $3x+1$ conjecture is in fact undecidable. A result by John Conway is described in [22]. Conway defines the general *Collatz function* as

$$g(N) = a_N N + b_N$$

where a_N and b_N are rational numbers that only depend on the value of N modulo some fixed number D . The general Collatz problem asks whether for every positive integer N there exists a k for which $g^k(N)=1$. Conway proves the general Collatz problem to be undecidable in [21].

In [1] Agnes and Rasetti claim that the $3x+1$ conjecture itself is undecidable. We present their claim and the argument they use to support it; we will then show that the argument is incorrect.

CLAIM A1 (p5):

Proving the undecidability of the $[3x+1]$ conjecture is straightforward, by mapping it onto a problem equivalent to the halting problem of a Turing machine, which is known to be undecidable.

ARGUMENT (pp 5-6, verbatim):

"The following conjecture, which has been *undecidably* attributed to either Stanislaw Ulam or Lothar Collatz, and has bearings in number theory, is the origin of the first sequence considered. We shall refer to the latter as the Ulam sequence. Starting from any natural number N_0 , the elements of the sequence are generated by the recursive rule:

$$S_n : \begin{cases} N_{n+1} = \frac{1}{2}N_n & \text{if } N_n \text{ is even} \\ N_{n+1} = 3N_n + 1 & \text{if } N_n \text{ is odd} \end{cases} \quad (2.1)$$

The conjecture (which has been checked for N_0 up to $7 \cdot 10^{11}$) states that for any N_0 the sequence shall terminate after a finite, possibly arbitrarily large, number K of steps, at $N_K = 1$. More precisely, the sequence hits for $n = K$ a limit cycle of order three given by $\{1, 4, 2, 1\}$.

Proving the undecidability of the conjecture is straightforward, by mapping it onto a problem equivalent to the halting problem of a Turing machine, which is known to be undecidable.

To this end one writes first N_n as a binary number,

$$N_n = \sum_{i=0}^{L_n} a_i^{(n)} 2^i ; a_i^{(n)} \in \mathbb{Z}_2 \quad (2.2)$$

where

$$L_n = \lfloor \log_2 N_n \rfloor, \quad (2.3)$$

$\lfloor x \rfloor$ denoting the largest integer $\leq x$. Thus N_n is represented by a string of length L_n+1 of 0's and 1's. Now the Ulam sequence (2.1), which can be written in compact form as

$$S_n : N_{n+1} = \frac{1}{2} N_n (5N_n + 3) - (5N_n + 2) \left\lfloor \frac{1}{2} N_n \right\rfloor \quad (2.4)$$

gives rise to a recursive relation among strings. Indeed, upon inserting (2.2) into (2.4), one gets the following recurrent generating scheme for the binary string $\{a_0^{(n)}, a_1^{(n)}, \dots, a_{L_n}^{(n)}\}$:

$$\begin{aligned} a_k^{(n+1)} &= a_0^{(n)} a_k^{(n)} + 3 \sum_{j=0}^{k-1} 2^{-(k-j)} a_j^{(n)} + 2^{-k} \\ &- \sum_{j=1}^{k-1} 2^{-(k-j)} a_j^{(n+1)} + (1 - a_0^{(n)}) a_{k+1}^{(n)} ; \text{ mod } 2 \end{aligned} \quad (2.5)$$

Eq.(2.5) holds for $0 \leq k \leq L_{n+1}$, besides $L_{n+1} \leq L_n + 2$.

Notice that, if $a_0^{(n)} = 0$ (i.e. if N_n is even), the recursion (2.5) is but a Bernoulli shift. Moreover, one needs a global information on the first portion of the new string in order to be able to decide each new digit of it, as well as its length.

Thus no algorithmic procedure of length less than $\sim n^2$ can be devised to decide whether for a given n , N_n equals 1 (or equivalently $a_0^{(n)}=1$, $a_0^{(k)}=0$, $\forall k: 1 \leq k \leq L_n+1$). In other words, the problem is NP-complete.

The input string $\{a_0^{(0)}, \dots, a_{L_0}^{(L_0)}\}$ together with the instructions for the recursive scheme coded in binary form in (2.5), constitute the abstract (Turing) machine mimicking, by the sequence of its internal states, the sequence of operations that take place in the process of computation.

The halting problem, namely the problem of distinguishing by a finite procedure which - if any - among the internal configurations of the machine lead to the program eventual halting, and which cause the program to compute forever, has been proved to be undecidable for any Turing machine by Michael Rabin".

The excerpt above touches on several issues, but we are concerned only with the undecidability argument. Their argument can be summarized as follows. Consider a program P whose input is any natural number. With input X , the action of P is to compute successive elements in the hailstone sequence for X , and halt when the element 1 is encountered. Therefore P will halt on a particular input X if and only if $Hs(X)$ contains the value 1. Since the "Halting Problem" is known to be undecidable, the problem of determining whether an arbitrary hailstone sequence contains the value 1 is undecidable, thereby proving that the $3x+1$ conjecture is undecidable.

DISCLAIM:

The Halting problem is the problem of determining, through a finite procedure, whether an *arbitrary* Turing machine will halt on arbitrary input. This problem is undecidable [12, 27]. The argument used above is incorrect because it did not map the $3x+1$ conjecture into the Halting problem. Rather, it mapped the $3x+1$ conjecture into the problem of deciding whether a *particular* Turing machine P halts on arbitrary input. Therefore the claim is unjustified. ■

The fact that the undecidability argument used by Agnes and Rasetti is incorrect does not imply that their *claim* is incorrect; the $3x+1$ conjecture may indeed be undecidable.

4.5 A Relationship with the Fibonacci Numbers

There exists an interesting (and, to our knowledge, previously unobserved) relationship between the iterates $H^k(X)$ of the hailstone function and the Fibonacci numbers. Let the function $H_0(X)=X/2$ represent the action of $H(X)$ on the even numbers, and the function $H_1(X)=3X+1$ represent the action of $H(X)$ on the odds. Then each point in the graph of H lies on one of the lines defined by $H_0(X)=X/2$ and $H_1(X)=3X+1$; in particular, one half of the points in the graph lie within each line (see Figure 4.5).

Figure 4.5 First iterate of the hailstone function

Figure 4.6 Second iterate of the hailstone function

Figure 4.7 Third iterate of the hailstone function

In a similar way, each point in the graph of the second iterate (shown in Figure 4.6) lies on one of the lines defined by

$$H_0 \circ H_0(X) = H_0\left(\frac{X}{2}\right) = \frac{X}{4}$$

$$H_0 \circ H_1(X) = H_0(3X+1) = \frac{3X+1}{2}$$

$$H_1 \circ H_0(X) = H_1\left(\frac{X}{2}\right) = 3\frac{X}{2}+1$$

The first line, $y=X/4$, contains those points $(X,H(X))$ of the graph for which X is a multiple of four; one quarter the points lie on this line. The second line, $y=(3X+1)/2$, contains those points $(X,H(X))$ for which X is odd; one half of the points lie on this line. Finally, the third line, $y=3(X/2)+1$ contains those points for which the value of X is twice an odd number. Notice that the line defined by $H_1 \circ H_1(X)$ does not appear in the graph; this is because every odd number maps to an even number under the hailstone function. The number of lines increases with higher iterates of H , as suggested in Figure 4.7. Let us generalize these observations.

(4.5.1) DEFINITION. A *hailstone line* is any line defined by one of the following three equations:

- i) $y = H_1(x)$
- ii) $y = H_0 \circ H_*(x)$
- iii) $y = H_1 \circ H_0 \circ H_*(x)$

where H_* can be either the identity function, or any function defining a hailstone line.

Let us define as *legal* the composition of any sequence of the functions H_0 and H_1 where H_1 is never composed directly with itself. We now show that the number of hailstone lines appearing in the graphs of higher iterates of H is related to the Fibonacci numbers.

(4.5.2) PROPOSITION. Let L_n be the number of hailstone lines appearing in the graph of the n -th iterate of the hailstone function. Then $L_n = F_{n+2}$, where F_{n+2} is the $(n+2)$ -nd Fibonacci number.

PROOF.

We first reformulate the problem in terms of binary n -tuples, and then prove it in that context. Consider a function defining a hailstone line in the graph of $H^n(x)$. Such a function corresponds to a legal composition of length n (or *n -composition*) and is of the form

$$H_{a_n} \circ \dots \circ H_{a_2} \circ H_{a_1}, \quad a_i \in \{0,1\}, \quad 1 \leq i \leq n$$

where H_1 is never composed with itself.

Observe that each n -composition is uniquely represented by the binary n -tuple $a_n a_{n-1} \dots a_1$.⁴⁹ That is, there exists a one-to-one correspondence between the set of hailstone lines in the graph of $H^n(x)$ and the set of binary n -tuples in which the pair (1,1) never occurs.

Let P_n be the set of binary n -tuples that do not contain a pair of consecutive ones.

⁴⁹ For convenience, we write binary n -tuples as a sequence of bits, and omit the usual parentheses and commas.

The first four these sets are shown below:

$$P_1 = \{0, 1\}$$

$$P_2 = \{00, 01, 10\}$$

$$P_3 = \{000, 001, 010, 100, 101\}$$

$$P_4 = \{0000, 0001, 0010, 0100, 0101, 1000, 1001, 1010\}$$

Now P_k corresponds to the set of hailstone lines appearing in the graph of $H^k(X)$.

The proof proceeds by induction. The basis of the induction is given by P_1 - P_4 above, since

$$L_1 = |P_1| = 2 \Rightarrow L_1 = F_3$$

$$L_2 = |P_2| = 3 \Rightarrow L_2 = F_4$$

$$L_3 = |P_3| = 5 \Rightarrow L_3 = F_5$$

$$L_4 = |P_4| = 8 \Rightarrow L_4 = F_8$$

Assume that the proposition holds for $n \leq k$; namely, $L_k = |P_k| = F_{k+2}$. Let A_k be the set of k -tuples from P_k that begin with a zero, and let B_k be the set of k -tuples from P_k that begin with a one; hence $P_k = A_k \cup B_k$. Now the elements of A_{k+1} are constructed from the elements of P_k by placing a zero in front of each:

$$\begin{aligned} A_{k+1} &= \{0p : p \in P_k\} \\ \Rightarrow |A_{k+1}| &= |P_k| = F_{k+2} \end{aligned}$$

However, the elements of B_{k+1} , which all start with a one, can only be constructed from elements from P_k that start with a zero. But these are just the elements of A_k , hence

$$B_{k+1} = \{1p : p \in A_k\}$$

$$\Rightarrow |B_{k+1}| = |A_k| = F_{k+1}$$

Finally, the set P_{k+1} is just the union of A_{k+1} and B_{k+1} , therefore

$$P_{k+1} = A_{k+1} \cup B_{k+1}$$

$$|P_{k+1}| = |A_{k+1}| + |B_{k+1}|$$

$$\Rightarrow |P_{k+1}| = F_{k+2} + F_{k+1} = F_{k+3}$$

Therefore $L_{k+1}=F_{k+3}$, completing the proof. ■

This result determines the number of ways in which integers can be transformed by iterates of the hailstone function. This allows us to establish the following

(4.5.3) PROPOSITION. The number of integers X for which the total stopping time $\tau_{\infty}(X)=k$, $k>0$ is bounded above by F_{k+2} , where F_{k+2} is the $(k+2)$ nd Fibonacci number.

PROOF.

Let X be a natural number, and let k be its total stopping time; that is $\tau_{\infty}(X)=k$. By the definition of total stopping time, $H^k(X)=1$. Hence at least one of the hailstone lines in the graph $H^k(X)$ intersects the point $(X,1)$. Since there are F_{k+2} hailstone lines in the graph of $H^k(X)$, at most F_{k+2} of them can intersect the point $(X,1)$. Thus, at most F_{k+2} integers can have the total stopping time $\tau_{\infty}(X)=k$. ■

4.6 On a definition of chaos

In [1] Agnes and Rasetti give a definition of chaos that is based on the relationship of a system to the Smale horse-shoe map and the binary shift map. The purpose of this section is to discuss and clarify this definition, as it relates to the topics discussed in Part I.

CLAIM A2 (p4, emphasis added):

[Chaos] can be related to the existence of *positive Lyapunov* exponents and can be characterized by the existence of *strange attractors* which may be schematically thought of as the product of a Cantor set and an interval. All of this leads to a rigorous definition based on the *homeomorphism* of chaotic dynamical systems with transitive topological mappings whose *symbolic dynamics* is isomorphic to that of *Smale horse-shoe* systems, invariant under Bernoulli shifts.

CLARIFICATION:

(a) Lyapunov exponents indicate the rate at which nearby points diverge. The existence of positive Lyapunov exponents indicates that a system has sensitive dependence on initial conditions. Sensitive dependence on initial conditions represents the working definition of chaos in much of the literature.

(b) Strange attractors are usually associated with chaotic systems, although the existence of a strange attractor does not imply that a system is chaotic [62]. Conversely,

chaotic systems do not have to possess strange attractors [78]; for example, the chaotic logistic map does not have an attractor at all [101]. As well, strange attractors can have many forms, and are not restricted to being the product of a Cantor set with an interval. An example is the Hénon attractor [7, 113].

(c) A *homeomorphism* is a one-to-one transformation between two spaces that is continuous in both directions. Topological transitivity refers to the indecomposable nature of chaotic systems, and is one of the requirements of topological chaos [34].

The definition of *formal chaos* states that a system is chaotic if there exists an invariant subset of its domain over which the system is topologically conjugate to the shift map [111]. It is known that Smale's horse-shoe map contains such an invariant set; hence the horse-shoe map is chaotic [141].

(d) *Symbolic dynamics* refers to the representation of a system by the action of a function over infinite sequences of symbols [5, 159]. Each symbol characterizes the position of a point in some way. The goal of symbolic dynamics is to transform questions about the behaviour of a given system into questions about the behaviour of an "equivalent" system over a space of infinite sequences. Typically these questions are about the number and lengths of periodic orbits, and are often easier to answer in the context of infinite sequences. The "equivalence" of two systems is made precise by requiring that they be topologically conjugate via a homeomorphism [34], as discussed in Chapter 3.

(e) The *Bernoulli shift* is another name for the binary shift map, or the shift map in the sequence space of two symbols.

We remark that the hailstone function is not topologically conjugate to a shift map, as the hailstone function is defined over a countable space ($\mathbb{N}$), whereas the shift map is defined over an uncountable space of infinite sequences.

4.7 Hailstone sequences and the Sarkovsky ordering

In this section we examine an interesting relationship between the hailstone system and Sarkovsky's ordering of the integers. In [1] the authors claim that the hailstone function acts as a shift map over the Sarkovsky sequence. We discuss this claim and then argue to the contrary.

CLAIM A3 (p12):

[The hailstone function] acts on Sarkowskii's ordered integers as a horse-shoe shift in a quite natural way.

CLARIFICATION:

Recall the Sarkovsky ordering of the natural numbers:

$$3 \triangleright 5 \triangleright 7 \triangleright \dots \triangleright 2 \cdot 3 \triangleright 2 \cdot 5 \triangleright \dots \triangleright 2^2 \cdot 3 \triangleright 2^2 \cdot 5 \triangleright \dots \triangleright 2^n \cdot 3 \triangleright 2^n \cdot 5 \triangleright \dots \triangleright 2^m \triangleright 2^{m-1} \triangleright \dots \triangleright 2^2 \triangleright 2 \triangleright 1$$

Define the *Sarkovsky sequence*, S , as the sequence of natural numbers formed according to this ordering, except for the powers of two which we write in reverse order for later

convenience.⁵⁰ That is,

$$S = \langle 3, 5, 7, \dots, 2 \cdot 3, 2 \cdot 5, \dots, 2^2 \cdot 3, 2^2 \cdot 5, \dots, 2^3 \cdot 3, 2^3 \cdot 5, \dots, 1, 2, 2^2, \dots, 2^m, 2^{m+1}, \dots \rangle$$

Consider S as being composed of an infinite number of "blocks" B_i , which are defined as follows:

$$\begin{aligned} B_0 &= \langle 3, 5, 7, \dots \rangle \\ B_1 &= \langle 2 \cdot 3, 2 \cdot 5, \dots \rangle \\ B_2 &= \langle 2^2 \cdot 3, 2^2 \cdot 5, \dots \rangle \\ &\vdots \\ B_k &= \langle 2^k \cdot 3, 2^k \cdot 5, \dots \rangle \\ &\vdots \\ B_\infty &= \langle 1, 2, 2^2, \dots, 2^{m-1}, 2^m, \dots \rangle \end{aligned}$$

The elements of each block can be written as

$$\begin{aligned} B_{i,j} &= 2^i(2j+3), & i,j=0,1,2,\dots \\ B_{\infty,j} &= 2^j, & j=0,1,2,\dots \end{aligned}$$

The action of the hailstone function on these elements is given by

$$\begin{aligned} H(B_{0,j}) &= 3(2j+3)+1 = 2(3j+5), & i,j=0,1,2,\dots \\ H(B_{i,j}) &= B_{i-1,j}, & i=1,2,3,\dots \\ H(B_{\infty,0}) &= B_{\infty,-2} \\ H(B_{\infty,j}) &= B_{\infty,j-1}, & j>0 \end{aligned}$$

Extend the action of the hailstone function to a set of points S as follows:

⁵⁰ The powers of two are written here in reverse order (with respect to the Sarkovsky ordering) for notational convenience only. This notation does not affect the arguments to follow.

$$H(S) = \{ H(x); x \in S \}$$

Then H acts on the blocks B_i as follows

$$\begin{aligned} H(B_0) &= \{ 2(3j+5); j \geq 0 \} \\ H(B_i) &= B_{i-1}, \quad i > 0 \\ H(B_-) &= B_- \end{aligned}$$

Figure 4.8 illustrates the action of H on the blocks. All blocks B_k $0 < k < \infty$ are mapped into B_{k-1} . In addition, the relative ordering of the elements within a block is preserved, since $B_{k,j}$ is mapped to $B_{k-1,j}$. If an element of B_0 is sent to block B_k , then it will "return" to B_0 after k iterations of the function. However, any element mapped to B_- (containing the powers of two) remains there forever. The elements of B_0 are mapped into *all* of the other blocks.

Figure 4.8 Effect of the hailstone function on the Sarkovsky sequence

The elements of block B_0 are scattered among the elements of all other blocks. The elements of block B_k are mapped onto block B_{k+1} , for $k > 0$. Block B_∞ (containing the powers of two) is mapped to itself.

(4.7.1) PROPOSITION. For every $n > 0$ there exists an $X \in B_0$ such that $H(X) \in B_n$.

PROOF.

By induction on n . As a basis, observe that

$$\begin{aligned} 3 \in B_0 &\Rightarrow H(3) = 10 = 2^1(5) \in B_1 \\ 9 \in B_0 &\Rightarrow H(9) = 28 = 2^2(7) \in B_2 \end{aligned}$$

and the proposition holds for $n=1,2$. Assume that it also holds for $n \leq k$; that is

$$\exists X \in B_0, Y \in B_k: Y = 3X + 1$$

The inductive step applies alternately to n odd and n even. Consider the next odd (resp. even) value $n=k+2$:

$$\begin{aligned} Y \in B_k &\Rightarrow 2^2 Y \in B_{k+2} \\ 2^2 Y &= 4(3X+1) = 3(4X+1)+1 \end{aligned}$$

Since $X \in B_0$, X is odd, and we have it that $X^* = 4X+1$ is also odd. Thus $Y^* \in B_{k+2}$ and $X^* \in B_0$, and the induction is complete for n odd (resp. even). ■

We are now ready to address Claim A3, which is repeated below.

CLAIM A3 (p12):

[The hailstone function] acts on Sarkowskii's ordered integers as a horse-shoe shift in a quite natural way.

DISCLAIM:

Smale's horse-shoe map (as it is properly called) is a two-dimensional map of the unit

square onto itself [34, 141]. The horse-shoe map has an invariant set that is in one-to-one correspondence with the set of all "two-sided" infinite binary sequences [159]. Recalling from Part I that Σ denotes the set of all "one-sided" infinite binary sequences:

$$\Sigma = \{ \pi = \langle \pi_0 \pi_1 \pi_2 \dots \rangle \mid \pi_i \in \{0,1\}, i \geq 0 \}$$

we write the set of all two-sided infinite sequences as:

$$\Sigma_2 = \{ \pi = \langle \dots \pi_{-2} \pi_{-1} \pi_0 \pi_1 \pi_2 \dots \rangle \mid \pi_i \in \{0,1\}, i \in \mathbb{Z} \}$$

Every sequence in Σ_2 represents a point in the invariant set of the horse-shoe map [34, 159]. Let p be a point in the unit square, and π be the corresponding sequence in Σ_2 . Then each bit $\pi_i \in \pi$ identifies the region of the unit square containing the image of p under i iterations of the horse-shoe map.⁵¹ The sequence $\dots \pi_{-2} \pi_{-1}$ represents the backward trajectory of the point, whereas $\pi_0 \pi_1 \pi_2 \dots$ represents the forward trajectory of the point.

When restricted to this invariant set, the horse-shoe map is dynamically equivalent to a "two-sided" shift map σ_2 over Σ_2 , where

$$\sigma_2(\langle \dots \pi_{-2} \pi_{-1} \pi_0 \pi_1 \pi_2 \dots \rangle) = \langle \dots \pi_{-1} \pi_0 \pi_1 \pi_2 \pi_3 \dots \rangle$$

If the hailstone function acts as a two-sided shift map over the Sarkovsky sequence (as claimed), there must exist a one-to-one correspondence between S and Σ_2 . This is impossible, since Σ_2 is an uncountable set while S is a countable set. Thus, the claim is incorrect. ■

⁵¹ The details of this correspondence are not important to this discussion.

There remains the possibility that there exists a countable subset of Σ_2 such that the action of σ_2 on this subset is equivalent to the hailstone function over the natural numbers. We now show that such a set cannot exist.

(4.7.2) PROPOSITION. The hailstone function over S is not equivalent to the shift map σ_2 over any subset of Σ_2 .

PROOF.

The shift map σ_2 on Σ_2 is invertible (a right shift is the opposite of a left shift), whereas the hailstone function H is not invertible. Thus there are elements in S that have two pre-images under H . For example, the pre-images of 10 under H are 3 and 20. However, an element of Σ_2 can have only one pre-image under σ_2 . Therefore the action of H on S cannot be equivalent to σ_2 on any subset of Σ_2 . ■

Finally, we argue that the hailstone function over S is qualitatively different from a shift map with respect to its action on a neighbourhood of points in S . With the exception of blocks B_0 and B_∞ , the hailstone function maps block B_k onto B_{k-1} in such a way that the relative distances between the elements is preserved. The elements of B_∞ are shifted one position to the right within the block such that their relative distances are also preserved (except for the element 1 which is mapped to 4). In addition, the scattering of the elements of block B_0 throughout the other blocks does not constitute a shift. These properties differ from that of a shift map, which increases the distance between points in a neighbourhood with each iteration.

4.8 Summary

We have discussed the basic properties of the hailstone function, and have established the following original results:

- i) The total stopping time of an integer X is bounded below by $\log_2 X$; that is, $\tau_-(X) \geq \log_2 X$.
- ii) The number of integers X having a total stopping time $\tau_-(X)=k$, for $k>0$ is bounded above by a value equal to the $(k+2)$ nd Fibonacci number.

In addition, we have shown that the claim of Agnes and Rasetti regarding the undecidability of the $3x+1$ problem is unjustified. We have also shown that the hailstone function does not act as a shift map on Sarkovsky's ordering of the natural numbers, contrary to the claim by the authors. Thus, the authors' claim that the hailstone function is chaotic over the natural numbers is also unjustified.

5. The Hailstone Map

5.1 Motivation

In the previous chapter, we discussed the dynamics of the hailstone function over the natural numbers and argued that it is not chaotic by any of the definitions given in Chapter 3. In this chapter, we discuss the properties of a map defined within the unit interval that is closely related to the hailstone function. This *hailstone map* was constructed by Agnes and Rasetti in [1]. We establish several properties of the hailstone map, and address the contention by Agnes and Rasetti that the map is chaotic and displays fractal orbits.

5.2 A dynamical system on the unit interval

The hailstone map is closely linked to the van der Corput function $C(n)$, which has seen much use in numerical analysis [68, 85]. The van der Corput function maps the natural numbers to a dense subset of the unit interval [67, 156]. This function takes the binary expansion of a natural number n and reverses the bits to create a *binary fraction*, as in

$$n = \sum_{i=0}^{L(n)-1} d_i 2^i = (d_{L(n)-1} \dots d_2 d_1 d_0)_2$$

$$C(n) = 0.d_0 d_1 d_2 \dots d_{L(n)-1}$$

where $L(n)$ is the length in bits of the binary expansion of n . Due to the countability of the natural numbers, the van der Corput function only maps to a countable subset of the unit interval. To identify the elements in the range of this function, we introduce the following

(5.2.1) DEFINITION. A *binary fraction* is a rational number in $[0,1)$ having a terminating binary expansion.

A binary fraction x is of the form $j \cdot 2^{-k}$ for some integers j and k ; we denote the set of binary fractions as Ψ_2 .

The van der Corput function $C: \mathbb{N} \rightarrow \Psi_2$ defines a one-to-one correspondence between the natural numbers and binary fractions [156]. The function acts on the binary expansion of a natural number X by "reflecting" it about the radix point:

$$X = \sum_{i=0}^{L-1} 2^i X_i; \quad X_i \in \{0,1\}$$

$$C(X) = \sum_{i=0}^{L-1} 2^{-(i+1)} X_i$$

$$X = X_{L-1} \dots X_1 X_0 \Rightarrow C(X) = 0.x_0 x_1 x_2 \dots x_{L-1}$$

where L denotes the length of the binary expansion of X . Observe that the length of the binary fraction $C(X)$ is equal to the length of the binary representation of X .

Let $C^{-1}: \Psi_2 \rightarrow \mathbb{N}$ denote the inverse of the van der Corput function, which maps binary fractions to natural numbers. Let us stress again that C^{-1} is not defined for rational numbers having non-terminating binary expansions. For example, the rational number $x=1/10$ has an infinite periodic binary expansion, given by $x=0.00011\ 0011\dots$. Only the *infinite* integer given by $X=\dots 1100\ 1100$ could map to this value.⁵² Thus X has no pre-image under C^{-1} . In what follows, let L denote the number of bits in the binary expansion of $x \in \Psi_2$.

(5.2.2) PROPOSITION. $C(2^k X) = 2^k C(X)$, for all $k \geq 0$.

PROOF.

Let $X_{L-1} \dots L_1 L_0$ be the binary representation of the number X . Now, multiplying an integer by 2^k is equivalent to shifting its binary representation k bits to the left. Thus the expansion of $2^k X$ consists of the digits in the expansion of X followed by k zeros, as in

⁵² Integers having an infinite number of digits are called *nonstandard* integers. An interesting discussion of nonstandard integers can be found in [144].

$X_{L-1}...X_1X_00...00$. The expansion of $C(2^kX)$ is given by $.00...0X_0X_1...X_{L-1}$. Observing that this is equivalent to $2^kC(X)$ completes the proof. ■

(5.2.3) DEFINITION. Let x and y be binary fractions. Then x and y are said to be *independent* with respect to C^{-1} if and only if $C^{-1}(x+y)=C^{-1}(x)+C^{-1}(y)$.

We observe that x and y are independent with respect to C^{-1} if and only if the binary expansions of x and y do not have any 1 bits in common. The following useful result gives a closed form for the inverse van der Corput function acting on binary fractions consisting of a block of 0's followed by a block of 1's.

(5.2.4) PROPOSITION. $C^{-1}(2^{-(j+i)}(2^i-1)) = 2^j(2^i-1)$, for $i, j \geq 0$.

PROOF.

Observe that the binary representation of the natural number (2^i-1) consists of i consecutive 1's; for example, $(2^3-1)=111_2$. Now $2^{-i}(2^i-1)$ is a binary fraction consisting of i 1's of the form $.1...1$. Thus $2^{-(j+i)}(2^i-1)$ is the binary fraction whose expansion consists of j 0's followed by i 1's; for example $2^{-(4+3)}(2^3-1)$ has the expansion $.0000111$. Now

$$\begin{aligned} x &= \sum_{i=0}^{L-1} 2^{-(i+1)} x_i \\ \rightarrow C^{-1}(x) &= \sum_{i=0}^{L-1} 2^i x_{L-1-i} \end{aligned}$$

Therefore the binary representation of $C^{-1}(2^{-(j+i)}(2^i-1))$ consists of i 1's followed by j 0's.

Observing that this is the representation of the natural number given by $2^j(2^i-1)$ completes

the proof. ■

We are now ready to define the hailstone map. Recall the definition of the hailstone function H from the previous chapter:

$$H: \mathbb{N} \rightarrow \mathbb{N}$$

$$H(X) = \begin{cases} \frac{X}{2} & \text{if } X \text{ is even} \\ 3X+1 & \text{if } X \text{ is odd} \end{cases}$$

(5.2.5) DEFINITION. The *hailstone map* is the function mapping the set of binary fractions to itself, given by

$$h: \Psi_2 \rightarrow \Psi_2$$

$$h = C \cdot H \cdot C^{-1}$$

$$h(x) = \begin{cases} 2x & \text{if } x < \frac{1}{2} \\ C(3C^{-1}(x)+1) & \text{if } x \geq \frac{1}{2} \end{cases}$$

Let the function $h_0(x)=2x$ represent the action of $h(x)$ for $0 \leq x < \frac{1}{2}$, and let $h_1(x)=C(3C^{-1}(x)+1)$ represent the action of $h(x)$ for $\frac{1}{2} \leq x < 1$. Figure 5.1 shows the graph of the hailstone map. For $x < \frac{1}{2}$ the function h_0 acts as a Bernoulli shift, mapping the set $\Psi_2[0, \frac{1}{2})$ to $\Psi_2[0, 1)$; these values of x correspond to the set of even natural numbers. A division by two (binary shift right) in the natural numbers corresponds to a multiplication by two (binary shift left) in Ψ_2 .

For $x \geq \frac{1}{2}$ the function h_1 has a complicated fractal-like structure, mapping the set $\Psi_2[\frac{1}{2}, 1)$ onto $\Psi_2[0, \frac{1}{2})$. The map effectively scatters these values of x back into the interval $[0, \frac{1}{2})$. The circled points correspond to the period three cycle $\langle 1, 4, 2 \rangle$ under the hailstone function.⁵³ These points are

$$(C(1), C(4)) = (0.1_2, 0.001_2) = (0.5, 0.125)$$

$$(C(4), C(2)) = (0.001_2, 0.01_2) = (0.125, 0.25)$$

$$(C(2), C(1)) = (0.01_2, 0.1_2) = (0.25, 0.5)$$

It is known [89] that hailstone sequences for all $n < 2^{40}$ eventually reach the cycle $\langle 1, 4, 2 \rangle$. This implies that, under the hailstone map, all binary fractions up to 40 bits in length have trajectories that eventually reach the cycle $\langle 0.5, 0.125, 0.25 \rangle$.

⁵³

Recall that almost all natural numbers enter the cycle $\langle 1, 4, 2 \rangle$ after a finite number of iterations of the hailstone function; the $3x+1$ Conjecture states that all natural numbers enter this cycle.

The Hailstone map

Figure 5.1 Hailstone map, first iterate

The hailstone map acts as a shift map for $x < \frac{1}{2}$, and has a fractal-like structure for $x \geq \frac{1}{2}$. The circled points represent the period three cycle.

In [1], Agnes and Rasetti make a number of claims regarding the properties of the hailstone map. The notation introduced in Chapter 5 to identify these claims and supporting arguments will be continued in the discussion to follow.

CLAIM A4 (p11):

We associate now an abstract one-dimensional dynamical system, defined as a map of the unit interval onto itself, for [the Ulam sequence]. More precisely, [this system] will turn out to map the subset Q_0 of the rationals less than one onto itself.

ARGUMENT:

None was provided by the authors.

DISCLAIM:

The "one-dimensional dynamical system" mentioned above corresponds to the hailstone map, defined earlier in this chapter. The "Ulam sequence" refers collectively to the set of all hailstone sequences.⁵⁴

The hailstone map is not defined for all rational numbers in the unit interval. This is because $h(x)$ is defined in terms of $C^{-1}(x)$, which is not defined for rational numbers having infinite periodic binary expansions. Hence h cannot map the subset of the rationals less than one into itself. ■

⁵⁴ Stanislaw Ulam was one of the researchers who popularized the $3x+1$ conjecture during the 1960's.

CLAIM A5 (p14):

[The] part of the map beyond $q=1/2$, whenever hit produces a random reinjection in the region $0 < q < 1/2$.

ARGUMENT:

None provided.

DISCLAIM:

Here, q represents a binary fraction in Ψ_2 . Whenever $q \geq 1/2$, the value of $h(q)$ is given by $h_1(q)$. This function is fully deterministic; hence its action is not random. In addition, the map is not even "pseudo-random" since there exists a strong correlation between the domain and range of the map h . For example, for all x values in the range $[1/2, 3/4)$, the values of $h(x)$ lie in the range $[0, 1/4)$. In contrast, for x values in the range $[3/4, 1)$, the values of $h(x)$ lie in the range $[1/4, 1/2)$. ■

Agnes and Rasetti [1] appear to define chaos as the existence of periodic points of all periods in a map of the unit interval, as described in the following

CLAIM A6 (p11):

Sarkowskii's theorem states that if a map has a periodic point of period k , it also has other periodic points whose periods are all the positive integers which follow k in the reordering of the field of natural numbers in which ascending odds come first, descending powers of two come last, and ascending odds times (ascending) powers of two in between. In particular, if a map has a periodic point of period three, it also has infinitely many other periodic points whose periods are all the natural integers; a condition

which implies chaos.

CLARIFICATION:

Sarkovsky's theorem does indeed state that the existence of a period-3 orbit guarantees the existence of orbits of all periods, but only for *continuous* real maps [34, 134]. Hence, Sarkovsky's theorem does not apply to the hailstone map.

The chaos of Li and Yorke, which occurs in continuous real maps having an orbit of period-3, implies not only a countable infinity of periodic orbits of all periods, but an uncountable infinity of aperiodic orbits as well [94]. ■

Although the authors correctly observe that these results cannot apply to the hailstone map, they still concluded that the map is chaotic:

CLAIM A7 (p12):

It is interesting that, even though for the [hailstone] map Sarkovskii's theorem cannot be assumed to hold, in that it requires that the map is at least piecewise smooth and monotonic - which is not the case, as will be shown in the sequel - nonetheless [the hailstone map] exhibits chaos.

DISCLAIM:

Certainly the hailstone map is not chaotic in the sense of Li and Yorke, as it is not a continuous map.⁵⁵ Even using the loose definition of "chaos" as being the existence of

⁵⁵ It can be shown that the hailstone map is not continuous with respect to the Euclidean metric. However, it is interesting to ask whether there exists *some other* metric for which the map is continuous.

periodic orbits of all periods, the authors do not demonstrate this behaviour in the hailstone map. Thus their claim that the hailstone map is chaotic in this sense is unjustified. ■

5.3 The arrowhead fractal

In this section we discuss the sense in which the graph of the hailstone map can be said to be fractal. Of course, this discussion is confined to the right hand side of the graph (with $x \geq \frac{1}{2}$), since the left hand side of the graph is given simply by the line $h_0(x) = 2x$.

Let A be the graph of the hailstone map in the interval $[\frac{1}{2}, 1)$; that is, $A = \{(x, h(x)) : x \in \Psi_2[\frac{1}{2}, 1)\}$, as shown in Figure 5.2. The set A exhibits a number of the fractal properties discussed in Chapter 2. In particular,

- A has a fine structure, with detail on arbitrarily small scales;
- A is totally disconnected (due to the discontinuity of $h_1(x)$ for $x \geq \frac{1}{2}$);
- A is apparently self-similar;
- A is defined in a simple way.

Henceforth, the set A shall be called the *arrowhead fractal*. We now discuss the extent to which the arrowhead fractal is self-similar. A can be partitioned into three sets A_1 , A_2 , and A_3 , where $A_1 = \{(x, h(x)) : x \in \Psi_2[\frac{1}{2}, 1)\}$, $A_2 = \{(x, h(x)) : x \in \Psi_2[5/8, 3/4)\}$, and $A_3 = \{(x, h(x)) :$

$x \in \Psi_2[1/2, 5/8)\}$ (see Figure 5.2). That is, $A = A_1 \cup A_2 \cup A_3$, and A_1 , A_2 , and A_3 are disjoint.

The apparent self-similarity of the arrowhead fractal may be captured through the construction of an iterated function system (IFS). Consider the following iterated function system, realizing the ratio list $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:

$$f_1: A \rightarrow A_1, \quad f_1((x, y)) = (\frac{1}{2}(1+x), \frac{1}{2}(\frac{1}{2}+y))$$

$$f_2: A \rightarrow A_2, \quad f_2((x, y)) = (\frac{1}{4}(2+x), \frac{1}{4}y)$$

$$f_3: A \rightarrow A_3, \quad f_3((x, y)) = (\frac{1}{4}(3-x), \frac{1}{4}(1-y))$$

The function f_1 maps the entire set A onto the half-scale copy A_1 in the upper right corner, as shown in Figure 5.2. The action of f_1 is to scale A down by one half, and translate it to the position $(3/4, 1/4)$.

Similarly, f_2 maps A onto the quarter-scale copy A_2 appearing at the bottom of the figure. The action of f_2 is to scale A down by one quarter, and translate it to the position $(5/8, 0)$. Finally, the function f_3 maps A onto the quarter-scale copy A_3 appearing on the left. The action of f_3 is to rotate A 180° about its central point, scale it down by one-quarter and then translate to the point $(1/2, 1/8)$.

The Arrowhead fractal

Figure 5.2 The arrowhead fractal

The sets A_1 , A_2 and A_3 are "sub-copies" of the arrowhead fractal.

Figure 5.3 shows the (self-similar) invariant set of this iterated function system. This figure is indistinguishable from the original graph of the hailstone map, giving strong evidence that the arrowhead fractal is self-similar. The self-similarity of the arrowhead fractal can be determined by examining each of the similarity functions f_1 , f_2 , and f_3 . If it can be determined that each $f_i: A \rightarrow A_i$ represents a one-to-one correspondence between A and A_i , then it will be established that A is self-similar. We shall first require the following

(5.3.1). LEMMA. Any function $f: S \times S \rightarrow T \times T$ of the form $f(x,y) = (ax+b, cy+d)$ $a, c \neq 0$ is injective. That is, for P and Q in $S \times S$, $f(P) = f(Q)$ implies $P = Q$.

PROOF.

Let $P_1 = (x,y)$ and $P_2 = (u,v)$ be two points in $S \times S$. Since $f(P_1) = f(P_2)$ we have $(ax+b, cy+d) = (au+b, cv+d)$, which implies that $(x,y) = (u,v)$. ■

Figure 5.3 Invariant set of the Arrowhead IFS

The IFS generating this graph makes no use of either the hailstone map or the hailstone function.

(5.3.2) PROPOSITION. f_1 is a one-to-one correspondence.

PROOF.

Since $f_1(x,y)$ is of the form $(ax+b, cy+d)$, by Lemma 8.4.2 it is injective. We need to prove that f_1 is surjective; that is, for each $Q \in A_1$ $\exists P \in A$: $f_1(P)=Q$. Let Q be a point in A_1 with $Q=(q,h(q))$, and let $P=(2q-1, 2h(q)-\frac{1}{2})$. Then $f_1(P)=(\frac{1}{2}(1+2q-1), \frac{1}{2}(1+2h(q)-1))=(q,h(q))=Q$. To complete the proof we just show that P is an element of A .

P is an element of A if and only if $h(2q-1)=2h(q)-\frac{1}{2}$. Recall that $h(2q-1)=CHC^{-1}(2q-1)$, and let $M=C^{-1}(q)$. Since $Q \in A_1$, $q \geq 3/4$ and the first two digits of its binary expansion are 11; thus $C^{-1}(2q-1)=\frac{1}{2}(M-1)$, and $HC^{-1}(2q-1)=H(\frac{1}{2}M-\frac{1}{2})=\frac{1}{2}(3M-1)$. Now $C^{-1}(2h(q)-\frac{1}{2})=\frac{1}{2}H(M)-1=\frac{1}{2}(3M-1)$, hence $h(2q-1)=2h(q)-\frac{1}{2}$, completing the proof. ■

(5.3.3) PROPOSITION. f_2 is a one-to-one correspondence.

PROOF.

The proof follows the same structure as that for f_1 . Since $f_2(x,y)$ is of the form $(ax+b, cy+d)$, it is injective. To prove that f_2 is surjective let Q be a point in A_2 with $Q=(q,h(q))$, and construct a $P \in A$ such that $f_2(P)=Q$. We claim that $P=(4q-2, 4h(q))$ is the required point. First, $f_2(P)=Q$ since $f_2(P)=(\frac{1}{4}(2+4q-2), \frac{1}{4}(4h(q)))=(q,h(q))=Q$.

P will be an element of A if and only if $h(4q-2)=4h(q)$. Observe that $h(4q-2)=CHC^{-1}(4q-2)$, and let $M=C^{-1}(q)$. Because $Q \in A_2$ $5/8 \leq q < 3/4$ and has a binary expansion of the form $q=0.10q_2q_3\dots$. Thus $C^{-1}(4q-2)=\frac{1}{4}M-\frac{1}{4}$, and $HC^{-1}(4q-2)=H(\frac{1}{4}M-\frac{1}{4})=\frac{1}{4}(3M+1)$. Now $C^{-1}(4h(q))=\frac{1}{4}H(M)-1=\frac{1}{4}(3M+1)$. Therefore $h(4q-2)=4h(q)$, and the assertion is

proved. ■

(5.3.4) PROPOSITION. f_3 is *not* a one-to-one correspondence.

PROOF.

We show that f_3 is not surjective. Given a point $Q=(q,h(q))$ in A_3 ; its pre-image $f_3^{-1}(Q)$ is $P=(3-4q,1-4h(q))$. We claim that P is not necessarily in A by giving a counter-example. Choose $Q=(1/2,1/8)$ to be the tip of the arrowhead in A_3 . Now $P=f_3^{-1}(Q)=(3-4(1/2),1-4/8)=(1,1/2)$; in fact, P is the tip of the arrowhead in A_1 , but it is not an element of A since hailstone map is not defined for the value 1. ■

In fact, because the hailstone map is "discontinuous to the left" *none* of the points in A correspond, under f_3 of any points in A_3 . This result shows that the arrowhead fractal is not *strictly* self-similar. We now consider the *Hausdorff dimension* of the set A , which is claimed by Agnes and Rasetti [1] to have the value 1:

CLAIM A8 (p12):

It is immediate to guess from the figure that ... it is indeed a discrete fractal curve, in which a characteristic arrow-head shape is repeated by self-similarity infinitely many times, and whose Hausdorff dimension is as infinitely close to 1 as the density of the rationals in the field of reals.

ARGUMENT:

None provided by the authors.

DISCLAIM:

The Hausdorff dimension of any countable set is zero, as established in Chapter 2. Since A is a countable set (its points are in one-to-one correspondence with the odd natural numbers), its Hausdorff dimension is zero. ■

Box-counting dimension

We have shown that the similarity dimension and Hausdorff dimensions of A are both zero. However, we will show that the box-counting dimension of A is non-zero. Recall that the box-counting dimension of a set F is defined as

$$\dim_B F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta}$$

where $N_\delta(F)$ is the number of boxes of width δ required to cover F [43].

To determine the box-counting dimension of A it is necessary to calculate the number of boxes required to cover the set at a given scale. To this end, we define a *full-scale box* to be a box of side length $\frac{1}{2}$, a *half-scale box* to be a box of side-length $\frac{1}{4}$, and so forth.

(5.3.5) PROPOSITION. The arrowhead fractal has box-counting dimension one.

PROOF.

Consider the number of boxes required to cover A at a particular scale. The set can be covered by one full-scale box, by four half-scale boxes, by sixteen quarter-scale boxes, and so forth. In general the set can be covered by 2^k boxes at scale 2^{-k} . Thus the box-counting dimension is given by

$$\dim_H A = \lim_{k \rightarrow \infty} \frac{\log 2^k}{-\log 2^{-k}} = \lim_{k \rightarrow \infty} \frac{k \log 2}{k \log 2} = 1$$

which completes the proof. ■

Although the arrowhead fractal can be covered by 2^n boxes at scale 2^{-n} , the number of *sub-copies* of the fractal at scale 2^{-n} is not equal to 2^n , as shown in the following

(5.3.6) PROPOSITION. The number of sub-copies appearing in the arrowhead fractal at a scale of 2^{-n} is given by

$$S_n = \frac{4}{3} 2^{n-1} - \frac{1}{3} (-1)^{n-1}$$

PROOF.

Notice that the arrowhead fractal is made up of one copy at half-scale plus two copies at quarter-scale, or three copies at quarter scale and two copies at eighth-scale. Consider a sub-copy of the image at scale 2^{-n} . This is made up of one sub-copy at scale $2^{-(n+1)}$, plus two sub-copies at scale $2^{-(n+2)}$. Now the copies at scale $2^{-(n+2)}$ originate from two sources: from the copies at scale 2^{-n} (each contributing two), and from the copies at scale $2^{-(n+1)}$

(each contributing one). Letting S_n be the number of sub-copies at scale 2^n , we have

$$S_0 = 1$$

$$S_1 = 1$$

$$S_{n+1} = S_n + 2S_{n-1}$$

Solving this linear recurrence⁵⁶ yields $S_n = (4 \cdot 2^{n-1} - (-1)^{n-1})/3$, which completes the proof. ■

Does the hailstone map have fractal orbits?

Agnes and Rasetti [1] use the fractal structure of the hailstone map to suggest that the map has fractal trajectories (orbits).

CLAIM A9 (p5):

The dynamical systems are not Hamiltonian and their orbits turn out to be fractals (i.e. self-similar discontinuous varieties).

CLARIFICATION:

The authors considered three dynamical systems, of which the hailstone map is the only one of interest here. Hamiltonian systems are "energy-conserving", and correspond in this context to invertible maps. Non-Hamiltonian systems correspond in this context to non-invertible maps (such as the hailstone map).

⁵⁶ For example, see [56, 158].

DISCLAIM.

The trajectory of a point x under the hailstone map is the sequence

$$\text{Tr}(x) = \langle x, h(x), h^2(x), h^3(x), \dots \rangle$$

Now the trajectory of x is fractal if the set of points $\{h^i(x); i=0,1,2,\dots\}$ has a fractal structure. In order for this to happen, $\text{Tr}(x)$ must have an infinite number of distinct points. It appears, due to the probable truth of the $3x+1$ conjecture, that all trajectories under the hailstone map are eventually periodic, and so contain a finite number of distinct values. Hence the claim is unjustified if the conjecture is true. ■

Is the arrowhead fractal a strange attractor?

Recall that an *attractor* is defined to be an attracting invariant subset of the phase space of a dynamical system. A *strange attractor* is an attractor that is either chaotic, or fractal. We show that the arrowhead fractal is not a strange attractor, as follows.

The phase space of the hailstone map is the set of binary fractions between zero and one. It is *not* the set of points in the unit square, or even that subset of the unit square defined by the arrowhead fractal. Hence, the arrowhead fractal cannot be an attractor under the hailstone map since it does not exist within the phase space of the system. In fact, the only known invariant sets under the hailstone map are Ψ_2 , and the period three cycle $\langle .1_2, .001_2, .01_2 \rangle$.

This shows that a function that has a fractal graph does not necessarily have fractal orbits. Conversely, a function that has a smooth graph may admit fractal orbits; the tent map is an example; its trajectories can take the form of Cantor sets in the unit interval, as discussed in Chapter 3.

5.4 Summary

We have shown that the hailstone map is not chaotic in the sense of Li and Yorke, on the basis of three arguments: first, since the hailstone map is not piecewise-continuous (at least with respect to the Euclidean metric), the theorem of Li and Yorke does not apply. Second, since the map is defined only over a discrete set, the map cannot possess an uncountable number of aperiodic orbits. Third, since any periodic orbits of period greater than three must have many thousands of terms, the hailstone map cannot be said to have orbits of all periods.

It is important to note that the dynamics of the hailstone map are not equivalent to the dynamics of the hailstone function, as the two systems are not topologically conjugate. This is due to the fact that the van der Corput function (being discontinuous at every point) does not constitute a homeomorphism between Ψ_2 and $\mathbb{N}$.

Finally, we have shown that the Hausdorff dimension of the fractal appearing in the graph of the hailstone map is in fact 0, and not 1 as claimed by the authors.

PART III: SEARCHING FOR DISCRETE CHAOS

6. Conclusions

We have presented the major concepts underlying the theory of fractals and chaos. Our investigation into the hailstone numbers revealed that reports by Agnes and Rasetti [1] regarding the existence of chaos in the system are unjustified. We summarize our thesis results in the following sections.

6.1 On the theory of fractals and chaos

FRACTALS: A *fractal* is loosely defined to be a set of points having a non-trivial geometric structure at all length scales. The scales involved can be either arbitrarily large, or arbitrarily small. Fractal sets often have some metric property that is also scale-invariant, and which is characterized by the *fractal dimension* of the set. This property is typically related to the size of the set, measured in some sense. Finite sets cannot be fractal in the strict sense of the term; however, finite sets can be fractal across a finite range of scales.

SENSITIVE DEPENDENCE: A dynamical system is said to have *sensitive dependence on initial conditions* if any uncertainty in the system state grows at an exponential rate (on average) for some non-trivial period of time. In practice, the effect of sensitive

dependence is to make the system unpredictable.

TOPOLOGICAL SENSITIVITY: We found it useful to define the term *topological sensitivity* to mean sensitive dependence in the sense of Devaney [34]. We say, then, that a system has topological sensitivity if, in every neighbourhood of the system there exist trajectories that eventually diverge from each other by some minimal distance, although not necessarily at an exponential rate.

TOPOLOGICAL CONJUGACY: Two dynamical systems are said to be *topologically conjugate* if there exists a homeomorphism (continuous one-to-one correspondence) that equates the actions of both systems over their respective domains. Topological conjugacy is an important concept in dynamical systems theory, wherein the dynamical properties of one system can be determined by showing it to be topologically conjugate to another system whose properties are already known. Two systems that are topologically conjugate are said to be equivalent in terms of their dynamics.

SYMBOLIC DYNAMICS: The *symbolic dynamics* of a system refers to the action of a dynamically equivalent system over a space of infinite symbol sequences. It is often the case that the dynamics of a system can be shown to be equivalent to a shift map, which acts on a symbol sequence by discarding the first element and shifting the remaining elements to the left one position. The dynamics of the shift map are well understood.

CHAOS: Many authors define a chaotic dynamical system as being one that simply has sensitive dependence on initial conditions. Other authors impose additional restrictions on the definition, usually with regard to the periodic structure of the system. Still others require that a system be topologically conjugate to a shift map. Unfortunately, there is no general agreement on the definition of chaos.

6.2 On the hailstone numbers

The hailstone numbers are a fully discrete dynamical system defined over the natural numbers. Although governed by a very simple evolution law, the system produces sequences that behave in an erratic manner. Thus the hailstone numbers provide a fine example of a simple deterministic system whose evolution is effectively unpredictable.

THE $3x+1$ CONJECTURE: The $3x+1$ conjecture states that the hailstone trajectory of every natural number eventually reaches the value 1. The $3x+1$ problem is the problem of determining whether the trajectory for an arbitrary number ever reaches the value 1. In our case study of the hailstone numbers, the authors Agnes and Rasetti [1] claimed that the $3x+1$ problem is undecidable by comparing it to the halting problem of a Turing machine. However, we came to the following

CONCLUSION: *The claim by Agnes and Rasetti that the $3x+1$ problem is undecidable is*

unjustified.

HAILSTONE FUNCTION. All known hailstone trajectories enter the same period-three cycle, and no other periodic trajectories are known to exist. A theorem due to Terras [152] implies that the $3x+1$ conjecture is probably true in the sense that the set of numbers violating the conjecture has density zero in the natural numbers. The probable truth of the $3x+1$ conjecture implies that the likelihood of there being any non-periodic trajectories is effectively zero.

CONCLUSION: *The claim by Agnes and Rasetti that the hailstone function is chaotic is unjustified.*

HAILSTONE FUNCTION AND THE FIBONACCI NUMBERS: An original result establishing a relationship between the hailstone function and the Fibonacci numbers places an upper bound on the number of integers having a given total stopping time. In particular, the number of integers having a total stopping time equal to k is bounded above by F_{k+2} , where F_{k+2} is the $(k+2)$ nd Fibonacci number.

TOTAL STOPPING TIMES: An original result establishes a lower bound on the number of iterations of the hailstone function required for the trajectory of a natural number X to encounter the value 1, thereby becoming periodic. In particular, the trajectory of a natural number X requires at least $\log_2 X$ steps before reaching 1.

SARKOVSKY'S THEOREM AND PERIOD-THREE: The existence of a period-three orbit in the hailstone map does not justify the claim that the map is chaotic. In particular, because the hailstone map is not piece-wise continuous, Sarkovsky's theorem and the theorem of Li and Yorke do not apply.

CONCLUSION: *Period-three does not imply chaos in the hailstone numbers.*

SARKOVSKY'S ORDERING AND THE SHIFT MAP. A superficial resemblance between the action of the hailstone function over Sarkovsky's ordering of the natural numbers and the shift map does not justify the claim that the hailstone function is equivalent to a shift map. In fact, we have the following

CONCLUSION. *The hailstone function does not act as a shift map on Sarkovsky's ordering of the natural numbers.*

HAILSTONE MAP: Due to its construction, the hailstone map is equivalent to the hailstone function in terms of its trajectory structure. However, the domains of the two systems are not homeomorphic; thus the hailstone function and hailstone map are not topologically conjugate.

CONCLUSION. *The hailstone function and hailstone map are not dynamically equivalent.*

THE ARROWHEAD FRACTAL: The graph of the hailstone map does indeed contain a fractal, which we have called the *arrowhead fractal*. Because the hailstone map is only defined over a discrete set the Hausdorff dimension of the arrowhead fractal is zero (although it does have box-counting dimension one). Also, because the graph of the hailstone map lies outside the phase space of the system, this fractal cannot represent a strange attractor.

CONCLUSION. *The arrowhead fractal is not a strange attractor.*

6.3 On fractals and chaos in fully discrete systems

From our research into the literature and our investigation of the hailstone numbers, we make the following conclusions regarding the existence of fractals and chaos in fully discrete dynamical systems.

DISCRETE FRACTALS: The properties of fractal sets depend on the existence of non-trivial structure at all scales, but does not require a space-continuum. The fractal dimension of a set is related to the way in which measures of the set size vary with scale. Measures that do not require an uncountable number of points in the set can yield non-zero dimensions for discrete sets. Thus we conclude that fractals can indeed be discrete.

SENSITIVE DEPENDENCE IN DISCRETE SYSTEMS: Provided the definition of the uncertainty

in the state of a dynamical system does not require a space-continuum, the growth in uncertainty can be measured as a function of the system evolution. If uncertainty grows at an exponential rate (on average) from some non-trivial period of time, then it can be said that the system has sensitive dependence on initial conditions. This applies to infinite systems having a dense set of points, in which neighbouring points can be arbitrarily close. However, the same can be said for finite systems, in which there is a lower bound on the distance between points. It can then be said that a finite system has sensitive dependence on initial conditions if this minimal uncertainty grows at an exponential rate (on average) for some non-trivial period of time.

CONCLUSION: *Sensitive dependence on initial conditions can occur in fully discrete dynamical systems.*

PERIODIC POINTS IN DISCRETE SYSTEMS: Chaos can occur in fully discrete dynamical systems in the form of sensitive dependence on initial conditions. Some definitions of chaos require that a chaotic system possess periodic points of all periods. Since this condition can exist in discrete systems, a corresponding definition of discrete chaos can be made.

CONCLUSION: *Discrete chaos can be defined to encompass sensitive dependence together with an infinity of periodic points.*

APERIODIC ORBITS IN DISCRETE SYSTEMS: Some definitions of chaos, particularly the definition due to Li and Yorke [94], require that a system possess an uncountable number of aperiodic points. This is impossible in a discrete (countable) system, however it is possible to have a *countably* infinite number of aperiodic orbits in a discrete system.⁵⁷

CONCLUSION: *Discrete chaos can be reasonably defined to require the existence of a countably infinite number of aperiodic points.*

THREE DEGREES OF EQUIVALENCE: We have observed that there are three basic types of equivalence between dynamical systems:

1. *structural equivalence*, by which a one-to-one correspondence can be established between the domains of each system;
2. *functional equivalence*, wherein the two systems are spatially equivalent, and by which the actions of both systems over their respective domains coincide, and;
3. *dynamical equivalence*, wherein the two systems are functionally equivalent, and by which the limit sets of each system correspond.

The distinction between functional and dynamical equivalence has implications

⁵⁷ A fully discrete system has been proposed by Kevin McGuire, possessing a countably infinite set of aperiodic integer sequences, where the *i*-th sequence contains the *i*-th powers of all the prime numbers.

with respect to sensitive dependence. Since sensitive dependence is a metric property, (it depends on the distance measured between points) it is not sufficient that two systems be functionally equivalent in order for the existence of sensitive dependence in one system to imply its existence in the other.

6.4 Recommendations for further research

In this section we present a variety of directions for further research into fractals and chaos in discrete systems.

1. Find a topological conjugacy to map the hailstone function into another system that is more amenable to analysis. One possibility is to use the technique of symbolic dynamics to establish an equivalence between the hailstone function and an appropriate function (preferably a shift map) over a sequence space.
2. Determine if the hailstone map exhibits sensitive dependence on initial conditions.
3. Determine the extent to which the van der Corput function induces fractal structures when mapping systems over the integers into systems within the unit interval.

4. Determine whether the $3x+1$ problem is undecidable.
5. Prove that the closure of the arrowhead fractal is a self-similar compact invariant set of the corresponding iterated function system. Investigate possible links between the contractions in the iterated function system and the hailstone map itself.
6. Pursue the existing link between the Hailstone numbers and the Fibonacci numbers, with an eye to tightening the upper bound on the number of integers having a given total stopping time.

APPENDIX A: Concepts from Metric Topology

In this Appendix, those basic concepts from topology that are used in the thesis, particularly in Chapter 3 (Fractals and Dimension) are defined. Further information on metric spaces and topology can be found in [34, 39, 43]. A brief discussion of measures can also be found in Chapter 9 of Hardy and Wright [70]. Unless otherwise specified, we assume that all limits and infima have well-defined values.

A.1 Metric spaces

(A.1.1) DEFINITION. A *metric space* is a set S together with a distance function $d: S \times S \rightarrow [0, \infty)$ such that

$$\begin{aligned}d(x, y) = 0 &\Leftrightarrow x = y; \\d(x, y) &= d(y, x) \\d(x, z) &\leq d(x, y) + d(y, z)\end{aligned}$$

The distance function d is called a *metric* on S . When the existence of the metric function is implied, the set S itself is referred to as a metric space.

(A.1.2) DEFINITION. The *Euclidean metric* for points $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$

is defined as

$$|y-x| = \left(\sum_{i=1}^n (y_i - x_i)^2 \right)^{\frac{1}{2}}$$

This is the intuitive notion of the distance between points. A *Euclidean metric space* is a subset of $\mathbb{R}^n$ endowed with the Euclidean metric.

(A.1.3) DEFINITION. The *diameter* of a metric space S , written $|S|$, is the least upper bound on the distance between any two points in S :

$$|S| = \sup \{ |y-x| : x, y \in S \}$$

A set with finite diameter is said to be *bounded*; a set with infinite diameter is said to be *unbounded*.

(A.1.4) DEFINITION. The *open ball* of radius r , centred at the point x is the set

$$B_r(x) = \{y \in S : |y-x| < r\}$$

The *closed ball* of radius r , centred at x is the set

$$\bar{B}_r(x) = \{y \in S : |y-x| \leq r\}$$

(A.1.5) DEFINITION. An *interior point* of the set S is any point x for which $\exists r > 0 : B_r(x) \subseteq S$. The union of all open balls in S is the *interior* of S , and is written $\text{int}(S)$.

(A.1.6) DEFINITION. A *boundary point* of the set S is any point x for which every open ball centred on x contains points in S and points not in S . That is, a boundary point belongs both to the closure of S and to the closure of its complement $\mathbb{R}^n - S$. The *boundary* of S is given by $\partial S = \bar{S} - \text{int}(S)$.

(A.1.7) DEFINITION. An *open set* is one for which every point is an interior point.

(A.1.8) DEFINITION. An *accumulation point* x of a subset $A \subseteq S$ is a point for which every open ball centred on x contains points in S other than x .

(A.1.9) DEFINITION. A *closed set* is one that contains all of its accumulation points. The *closure* of a set $A \subseteq S$ is the set A together with all of its accumulation points:

$$\bar{A} = A \cup \{x \in S : B_r(x) \cap A \neq \emptyset \text{ for all } r > 0\}$$

(A.1.10) DEFINITION. The set $A \subseteq B$ is said to be *dense* in B if the closure of A is equal to B .

(A.1.11) EXAMPLE. Let $Q[0,1]$ be the set of rationals in the unit interval; that is, $Q[0,1] = \{x \in \mathbb{Q} : 0 \leq x \leq 1\}$. The closure of $Q[0,1]$ includes all of the irrational points in $[0,1]$, hence

$$\overline{Q[0,1]} = [0,1]$$

A.2 Compact metric spaces

(A.2.1) DEFINITION. A *sequence* $\langle x_i \rangle$ in a space S is a possibly infinite list of values $\langle x_0, x_1, \dots \rangle$ where each $x_i \in S$ for $i \geq 0$.

(A.2.2) DEFINITION. A *subsequence* of $\langle x_i \rangle$ is a sequence of the form

$$\langle x_{k_0}, x_{k_1}, \dots \rangle; \quad x_{k_i} \in \langle x_i \rangle, \quad 0 < k_0 < k_1 < \dots$$

(A.2.3) DEFINITION. The *limit* of a sequence $\langle x_i \rangle$ is the number x for which there exists $K \in \mathbb{N}$ such that $i \geq K$ implies $|x_i - x| < \epsilon$ for all $\epsilon > 0$. If the limit x exists, the sequence $\langle x_i \rangle$ is said to *converge* to x ; otherwise the sequence is said to *diverge*.

(A.2.4) DEFINITION. The number x is *cluster point* of the sequence $\langle x_i \rangle$ if for every $K \geq 0$ there exists $i > K$ for which $|x_i - x| < \epsilon$ for all $\epsilon > 0$.

(A.2.5) DEFINITION. A sequence $\langle x_i \rangle$ is said to be a *Cauchy sequence* if for every $\epsilon > 0$ there exist $K \geq 0$ and $L \geq 0$ such that $|x_i - x_j| < \epsilon$ for all $i \geq K, j \geq L$.

(A.2.6) DEFINITION. A metric space S is said to be *complete* if every Cauchy sequence in S converges to a point in S . Equivalently, S is complete if the limit of every convergent sequence in S is a point in S .

(A.2.7) PROPOSITION. The point $x \in S$ is an accumulation point of S if and only if x is the limit of some sequence in S .

PROOF. See [Edgar, 1990]. ■

(A.2.8) PROPOSITION. A metric space is complete if and only if it is closed.

PROOF. [Edgar, 1990].

($\Rightarrow$) Let S be a complete metric space; then S contains the limit of every convergent sequence of S . Therefore contains all of its accumulation points and is closed.

($\Leftarrow$) Let S be a closed set; then S contains all of its accumulation points. Therefore it contains the limit points of all convergent sequences in S . Hence S is a complete metric space. ■

(A.2.9) DEFINITION. A metric space S is said to be *compact* if every sequence in S has at least one cluster point in S . Equivalently, S is compact if it is a closed and bounded subset of a complete metric space.

The terms "closed" and "complete" have nearly the same meaning, except that we say that sets are closed, whereas metric spaces are complete. In contrast, the term "compact" can be applied to both sets and metric spaces.

(A.2.10) DEFINITION. Let S be a subset of $\mathbb{R}^n$. Then S is said to be *connected* if it cannot be covered by two disjoint open sets.

(A.2.11) DEFINITION. The largest subset of S containing a point x is called the *connected component* of x .

(A.2.12) DEFINITION. Let S be a subset of $\mathbb{R}^n$. Then S is *totally disconnected* if the connected component of every point x consists of just that point.

A.3 Lebesgue measure

The notion of measure is central to topology, and hence to fractal geometry. The measure of a set S is a numerical value used to represent the 'size' of S in some sense. In particular, we have the following

(A.3.1) DEFINITION. [Falconer, 1990]. Let μ be a function mapping subsets of $\mathbb{R}^n$ to $\mathbb{R}$. Then μ is said to be a *measure* on $\mathbb{R}^n$ provided

- (a) $\mu(\emptyset) = 0$
- (b) $\mu(A) \leq \mu(B)$ if $A \subseteq B$
- (c) If $A_1, A_2, \dots$ is a countable sequence of sets then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

with equality if the A_i are disjoint.

The simplest example of a measure μ is the *counting measure*, where $\mu(A)$ is the number of points in A if A is finite, and ∞ otherwise.

In this section we define the notion of *Lebesgue measure* which roughly corresponds to the intuitive notions of the length, area, and volume of a set. Only a general definition is given here, and we avoid those details that do not contribute to the discussion later on.⁵⁸ Further details can be found in [39, 43].

(A.3.2) DEFINITION. The *linear Lebesgue measure* $L^1(S)$ of a set S is

$$L^1(S) = \inf \left\{ \sum_{i=1}^{\infty} (b_i - a_i) : S \subset \bigcup_{i=1}^{\infty} [a_i, b_i] \right\}$$

where $[a, b]$ is a small interval that intersects S .

The Lebesgue measure is calculated by considering all coverings of S by countable collections of intervals, and taking the smallest (limiting) total length possible. In the case where A is itself an interval, or a countable collection of disjoint intervals, then $L^1(S)$ is simply the length of S .

(A.3.3) PROPOSITION.[†] All discrete (countable) subsets of $\mathbb{R}$ have Lebesgue measure zero.

PROOF.

⁵⁸ In particular, the discussion is restricted to the case where the Lebesgue inner measure and Lebesgue outer measure are equal. Also, all subsets of $\mathbb{R}$ are assumed to be Borel sets.

3 of/de 3

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

PIONEERS IN METHYLENE BLUE TESTING SINCE 1974

10800 COUNTY ROAD 9, BIRMINGHAM, AL 35207 USA
TEL: 817-426-7887 FAX: 817-426-7887 TLE: 817-426-7887

Let $S \subset \mathbb{R}$ be a discrete set. Let δ_i denote an interval of length δ' covering $x_i \in S$, for some small $\delta > 0$. Then S is a subset of the (countable) union of these intervals:

$$S \subset \bigcup_{i=1}^{\infty} \delta_i$$

Given $\delta > 0$, let n be the greatest integer such that $\delta < (1/n)$. Now the total length of the intervals δ_i goes to zero as δ goes to zero:

$$\sum_{i=1}^{\infty} \delta_i = \sum_{i=1}^{\infty} \delta' < \sum_{i=1}^{\infty} \left(\frac{1}{n}\right)' = \frac{1}{n-1} \rightarrow 0 \text{ as } n \rightarrow \infty$$

completing the proof. ■

(A.3.4) DEFINITION. The *n-dimensional Lebesgue measure* of a set S is the infimum of all countable coverings of S composed of n -dimensional volume elements A_i :

$$L^n(S) = \inf \left\{ \sum_{i=1}^{\infty} \text{vol}^n(A_i) : S \subset \bigcup_{i=1}^{\infty} A_i \right\}$$

where $\text{vol}^n(A_i)$ is the n -dimensional volume of the compact set A_i .

The n -dimensional Lebesgue measure can be thought of as the extension of the usual notions of n -dimensional volume to arbitrary sets. For example, $L^0(S)$ represents the number of points in S . In general, $L^1(S)$ can be written as $\text{length}(S)$, $L^2(S)$ as $\text{area}(S)$, $L^3(S)$ as $\text{vol}(S)$, and $L^n(S)$ as $\text{vol}^n(S)$. We remark that all discrete subsets of $\mathbb{R}^n$ have n -dimensional Lebesgue measure zero; the proof is analogous to that for L^1 .

A.4 Continuous functions

(A.4.1) DEFINITION. A function $f:S \rightarrow T$ is said to be *continuous* if $\forall \epsilon > 0 \exists \delta > 0: |y-x| < \delta \Rightarrow |f(y)-f(x)| < \epsilon$.

(A.4.2) DEFINITION. A function $f:S \rightarrow T$ is said to be *injective*, or *one-to-one*, if $f(x)=f(y)$ implies $x=y$.

(A.4.3) DEFINITION. A function $f:S \rightarrow T$ is said to be *surjective*, or *onto*, if $\forall y \in T \exists x \in S: f(x)=y$.

(A.4.4) DEFINITION. A *bijective* function is one that is both injective and surjective. A bijective function is also called a *one-to-one correspondence*.

(A.4.5) DEFINITION. A function $f:S \rightarrow T$ is said to be a *homeomorphism* if it is a bijection that is continuous in both directions. Two spaces are said to be *homeomorphic* if there exists a homeomorphism from one space to the other.

(A.4.6) DEFINITION. [Devaney, 1986]. Let $f:A \rightarrow A$, and $g:B \rightarrow B$ be two maps. Then f and g are said to be *topologically conjugate* if there exists a homeomorphism $h:A \rightarrow B$ such that $h \circ f = g \circ h$.

(A.4.7) DEFINITION. A property of a set is said to be a *topological property* if it is invariant under homeomorphisms.

A.5 Notes on dimension

(A.5.1) DEFINITION. The *open r -neighbourhood* of a set contains those points y that are within distance r of some point x in A :

$$N_r(A) = \{y: |y-x| < r, \text{ for some } x \in A\}$$

(A.5.2) DEFINITION. Let A and B be two non-empty compact sets. Then the *Hausdorff distance* between A and B is defined as

$$D(A,B) = \inf \{r: A \subseteq N_r(B), B \subseteq N_r(A)\}$$

(A.5.3) PROPOSITION. [Edgar]. Let A_n be a sequence of non-empty compact subsets of the metric space S . If A_n converges to A in the Hausdorff metric, then every sequence $\langle x_i \rangle$ for which $x_i \in A_i$ converges to a point $x \in A$. ■

It is interesting to note that the s -dimensional Hausdorff measure is equal to the s -dimensional Lebesgue measure, to within a constant factor. In particular, the Lebesgue form uses the s -dimensional volume element in the sum, whereas the Hausdorff form uses

only the s -th power of the diameter in the sum. For completeness, we state this relationship in the following theorem (without proof) from [42]:

(A.5.4) THEOREM. [Falconer, 1985]. If $F \subset \mathbb{R}^n$, then $L^s(F) = c_n H^s(F)$, where

$$c_n = \frac{\pi^{\frac{1}{2}n}}{2^n \left(\frac{1}{2}\right)!}$$

In particular, $c_1 = 1$ and $c_2 = \pi/4$. ■

(A.5.5) PROPOSITION. [Falconer, 1985]. The Hausdorff dimension of the triadic Cantor set is $\log 2 / \log 3$.

PROOF.

Let C be the triadic Cantor set and let C_k be the k -th approximant to C , as shown in Figure 3.1. Now, C_k contains 2^k disjoint intervals labelled C_{kj} , where $j = 1, 2, \dots, 2^k$. Each of the intervals C_{kj} is of size 3^{-k} , and the minimum distance between adjacent intervals is 3^{-k} .

We establish an upper bound on the s -dimensional Hausdorff measure of C at $s = \log 2 / \log 3$. Observe that the family of basic intervals $\{C_{kj}\}$ in the approximant C_k covers all of C . Thus we have

$$H_{3^{-k}}^s \leq \sum_{j=1}^{2^k} |C_{kj}|^s = 2^k |3^{-k}|^s = 2^k (3^s)^{-k}$$

Setting $s = \log 2 / \log 3$ gives

$$H_{3^{-k}}^s(C) \leq 2^k(2)^{-k} = 1$$

$$\Rightarrow H^s(C) \leq \lim_{k \rightarrow \infty} H_{3^{-k}}^s(C) = 1$$

which results in the desired upper bound.

The lower bound can be established by a counting argument. Let $\{U_i\}$ be any finite cover of C . Without loss of generality it can be assumed that each U_i are closed sub-intervals of $[0,1]$; for it is only the diameter $|U_i|$ that matters. Given a particular U_i , let k be an integer such that

$$3^{-(k+1)} \leq |U_i| < 3^{-k}$$

Then U_i can intersect at most one basic interval of C_k . If $j \geq k$ then U_i intersects at most

$$2^{j-k} = 2^j 3^{-jk} = 2^j ((3)3^{-(k+1)})^j \leq 2^j 3^j |U_i|^j$$

basic intervals in C_j . That is, $2^j 3^j |U_i|^j$ is an upper bound (independent of k) on the number of basic intervals of size 3^{-j} that are intersected by U_i . Now choose j large enough so that this upper bound is valid for each U_i . Then the total number of basic intervals of size 3^{-j} intersected by all the U_i is bounded above by

$$\sum_i 2^j 3^j |U_i|^j$$

But the total number is exactly 2^j , therefore

$$\sum_i 2^j 3^j |U_i|^j \geq 2^j$$

$$\Rightarrow \sum_i |U_i|^j \geq 3^{-j} = \frac{1}{2}$$

So the Hausdorff measure $H^s(C)$ is non-zero and finite for $s = \log 2 / \log 3$; therefore the value

of $H^s(C)$ is infinite for $s < \log 2 / \log 3$, and zero for $s > \log 2 / \log 3$. The existence of a jump discontinuity in the Hausdorff measure for this value of s implies that the dimension of C is $\log 2 / \log 3$. ■

(A.5.6) THEOREM. [Edgar, 1990]. Let $(r_1, r_2, \dots, r_n)$ be a contracting ratio list. Then there is a unique non-negative number s satisfying

$$\sum_{i=1}^n r_i^s = 1$$

The number s is 0 if and only if $n=1$. ■

(A.5.7) THEOREM. [Edgar, 1990]. Let S be a complete metric space, let $(r_1, r_2, \dots, r_n)$ be contracting ratio list, and let $(f_1, f_2, \dots, f_n)$ be an iterated function system of similarities in S that realizes the ratio list. Then there is a unique non-empty compact invariant set for the iterated function system. ■

REFERENCES

- [1] C. Agnes and M. Rasetti; Complexity, undecidability and chaos: A class of dynamical systems with fractal orbits, in: *Workshop on Chaos and Complexity*, eds. R. Livi, S. Ruffo, S. Ciliberto and M. Buiatti (World Scientific, Singapore, 1987) p. 3.
- [2] V.M. Alekseev and M.V. Yakobson, Symbolic Dynamics and Hyperbolic Dynamic Systems, *Physics Reports* 75 (1981) 287.
- [3] Z. Aleksic, Estimating the embedding dimension, *Physica D* 52 (1991) 362.
- [4] P.W. Anderson, ed., Santa Fe Institute Studies in the Sciences of Complexity, V: *The Economy as an Evolving Complex System* Addison Wesley, Redwood City, 1989.
- [5] D.K. Arrowsmith and C.M. Place, *An Introduction to Dynamical Systems* Cambridge University Press, Cambridge, 1990.
- [6] H. Atmanspacher, H. Scheingraber and W. Voges, Global scaling properties of a chaotic attractor reconstructed from experimental data, *Physical Review A* 37 (1988) 1314.
- [7] G.L. Baker and J.P. Gollub, *Chaotic Dynamics: An Introduction* Cambridge University Press, Cambridge, UK, 1990.
- [8] M. Barnsley, *Fractals Everywhere* Academic Press, Boston, 1988.
- [9] M.F. Barnsley and S.G. Demko, eds., Notes and Reports in Mathematics in Science and Engineering, 2: *Chaotic Dynamics and Fractals* Academic Press, San Diego, CA, 1986.
- [10] C. Beck and G. Roepstorff, Effects of Phase Space Discretization on the Long-time Behaviour of Dynamical Systems, *Physica D* 25 (1987) 173.
- [11] J.S. Birman and R.F. Williams, Knotted Periodic Orbits in Dynamical Systems - I: Lorenz's Equations, *Topology* 22 (1983) 47.
- [12] G.S. Boolos and R.C. Jeffrey, *Computability and Logic*, 3rd ed. Cambridge University Press, Cambridge, UK, 1989.

- [13] W.E. Boyce and R.C. DiPrima, *Elementary Differential Equations and Boundary Value Problems*, 3rd ed. John Wiley & Sons, New York, NY, 1977.
- [14] K. Briggs, An improved method for estimating Liapunov exponents of chaotic time series, *Physics Letters A* 151(1 Suppl. 2) (1990) 27.
- [15] F. Cajori, Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World, 2: *Principia: Vol. II The System of the World*, Revised ed. University of California Press, Berkeley, CA, 1934.
- [16] G.J. Chaitin, Information-theoretic computational complexity, *IEEE Transactions on Information Theory* IT-20 (1974) 10.
- [17] G.J. Chaitin, Information-theoretic limitations of formal systems, *Journal of the ACM* 21 (1974) 403.
- [18] G.J. Chaitin, Randomness and mathematical proof, *Scientific American* 232(5,May) (1975) 47.
- [19] G.J. Chaitin, Algorithmic information theory, *IBM Journal of Research and Development* 21 (1977) 350.
- [20] T. Clarke, Bounded Turing Machines are Chaotic, *IJCAI-91 Proceedings* (1991).
- [21] J.H. Conway; Unpredictable iterations, in: *Proceedings of the Number Theory Conference, Boulder, Colorado* (1972) p. 49.
- [22] J.H. Conway; FRACTRAN: A simple universal programming language for arithmetic, in: *Open Problems in Communication and Computing*, eds. T.M. Cover and B. Gopinath (Springer-Verlag, New York, NY, 1987) p. 3.
- [23] E.M. Coven, I. Kan and J.A. Yorke, Pseudo-orbit Shadowing in the Family of Tent Maps, (1986).
- [24] J.P. Crutchfield, J.D. Farmer, N.H. Packard and R.S. Shaw, Chaos, *Scientific American*(December) (1986) 46.
- [25] K. Culik, L.P. Hurd and S. Yu, Computation Theoretic Aspects of Cellular Automata, *Physica D* 45 (1990) 357.
- [26] K. Culik, J. Pachl and S. Yu, Department of Computer Science Technical Reports, CS-87-47: *On the Limit Sets of Cellular Automata* University of Waterloo, Waterloo, 1987.

- [27] N.J. Cutland, *Computability: An Introduction to Recursive Function Theory* Cambridge University Press, Cambridge, 1980.
- [28] A.G. Darbyshire and T.J. Price; Phase Portraits from Chaotic Time Series, in: *Fractals and Chaos*, eds. A.J. Crilly, R.A. Earnshaw and H. Jones (Springer-Verlag, New York, 1991) p. 248.
- [29] M.J. Davis and R.S. MacKay, Markov shifts in the Hénon family, *Physica D* 52 (1991) 171.
- [30] M.D. Davis and E.J. Weyuker, *Computer Science and Applied Mathematics: Computability, Complexity, and Languages* Academic Press, Orlando, 1983.
- [31] P.J. Davis and R. Hersh, *Descartes' Dream: The World According to Mathematics* Houghton Mifflin, Boston, MA, 1986.
- [32] F.M. Dekking, Recurrent sets, *Advances in Mathematics* 44 (1982) 78.
- [33] D.S. Dendrinos and M. Sonis, *Applied Mathematical Sciences, 86: Chaos and Socio-Spatial Dynamics* Springer-Verlag, New York, 1990.
- [34] R.L. Devaney, Addison-Wesley Series in Nonlinearity, 1: *An Introduction to Chaotic Dynamical Systems*, 2nd ed. Addison-Wesley, Redwood City, CA, 1989.
- [35] R.L. Devaney, *Chaos, Fractals, and Dynamics: Computer Experiments in Mathematics* Addison-Wesley, Menlo Park, CA, 1990.
- [36] A.K. Dewdney, Leaping into Lyapunov Space, *Scientific American* 265(3,September) (1991) 178.
- [37] Jean-P. Eckmann, Roads to turbulence in dissipative dynamical systems, *Reviews of Modern Physics* 53(4,October Suppl. 1) (1981) 643.
- [38] Jean-P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, *Reviews of Modern Physics* 57(3) (1985) 617.
- [39] G.A. Edgar, *Undergraduate Texts in Mathematics: Measure, Topology, and Fractal Geometry* Springer-Verlag, New York, NY, 1990.
- [40] S. Ellner, Estimating attractor dimensions from limited data: a new method, with error estimates, *Physics Letters A* 133 (1988) 128.
- [41] S. Ellner, A.R. Gallant, D. McCaffrey and D. Nychka, Convergence rates and data requirements for Jacobian-based estimates of Lyapunov exponents from data,

Physics Letters A 153(6 Suppl. 7) (1991) 357.

- [42] K.J. Falconer, Cambridge Tracts in Mathematics, 85: *The Geometry of Fractal Sets* Cambridge University Press, Cambridge, 1985.
- [43] K.J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications* John Wiley & Sons, Chichester, UK, 1990.
- [44] D. Farmer, T. Toffoli and S. Wolfram, eds., *Cellular Automata: Proceedings of an Interdisciplinary Workshop* North-Holland, Amsterdam, 1984.
- [45] W.J. Firth; Instabilities and chaos in lasers and optical resonators, in: *Chaos*, ed. A.V. Holden (Princeton University Press, Princeton, NJ, 1986) p. 135.
- [46] M. Fleischmann, D.J. Tildesley and R.C. Ball, Proceedings of the Royal Society of London: *Fractals in the Natural Sciences* Princeton University Press, Princeton, N.J., 1989.
- [47] J. Ford; Directions in Classical Chaos, in: Directions in Condensed Matter Physics, 3: *Directions in Chaos*, Vol. 1, ed. B. Hao (World Scientific, Singapore, 1988) p. 1.
- [48] J. Ford; What is Chaos, That We Should Be Mindful Of It?, in: *The New Physics*, ed. P. Davies (Cambridge University Press, Princeton?, 1989) p. 348.
- [49] G.W. Frank, T. Lookman, M.A.H. Nerenberg, C. Essex, J. Lemieux and W. Blume, Chaotic Time Series Analysis of Epileptic Seizures, *Physica D* 46 (1990) 427.
- [50] M. Gardner, The random number Ω bids fair to hold the mysteries of the universe, *Scientific American* 241(5, November) (1979) 20.
- [51] L. Garner, On the Collatz $3n+1$ problem, *Proceedings of the American Mathematical Society* 82 (1981) 19.
- [52] I. Garrido and C. Simo; Some ideas about strange attractors, in: *Lecture Notes in Physics*, Vol. 179, ed. L. Garrido (Springer-Verlag, Berlin, 1983) p. 1.
- [53] M. Gerhardt and H. Schuster, A Cellular Automaton Describing the Formation of Spatially Ordered Structures in Chemical Systems, *Physica D* 36 (1989) 209.
- [54] L. Glass and M. Mackey, *From Clocks to Chaos* Princeton University Press, Princeton, 1988.

- [55] W.H. Gottschalk and G.A. Hedlund, American Mathematical Society Colloquium Publications, 36: *Topological Dynamics* American Mathematical Society, Providence, RI, 1955.
- [56] R.L. Graham, D.E. Knuth and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science* Addison-Wesley, Reading, MA, 1989.
- [57] P. Grassberger, Chaos and Diffusion in Deterministic Cellular Automata, *Physica D* 10 (1984) 52.
- [58] P. Grassberger and Procaccia, Characterization of strange attractors, *Physical Review Letters* 31(6) (1983) 346.
- [59] P. Grassberger, *Journal of Statistical Physics* 26 (1981) 173.
- [60] P. Grassberger, Estimating the fractal dimensions and entropies of strange attractors, in: *Chaos*, ed. A.V. Holden (Princeton University Press, Princeton, NJ, 1986) p. 291.
- [61] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, *Physica D* 9 (1983) 189.
- [62] C. Grebogi, E. Ott, S. Pelikan and J.A. Yorke, Strange attractors that are not chaotic, *Physica D* 13 (1984) 261.
- [63] J. Guckenheimer, Symbolic Dynamics and Relaxation Oscillations, *Physica D* 1 (1980) 227.
- [64] M.C. Gutzwiller, Interdisciplinary Applied Mathematics, 1: *Chaos in Classical and Quantum Mechanics* Springer-Verlag, New York, NY, 1990.
- [65] M.C. Gutzwiller, Quantum Chaos, *Scientific American* 266(1, January) (1992) 78.
- [66] J. Hadamard, Les surfaces à courbes opposées et leurs lignes géodésiques, *Journal Mathématique Pures et Appliquées* 4 (1898) 27.
- [67] J.H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numerische Mathematik* 2 (1960) 84.
- [68] J.M. Hammersley and D.C. Handscomb, *Monte Carlo Methods* Methuen, London, 1964.
- [69] B. Hao, *Elementary Symbolic Dynamics and Chaos in Dissipative Systems* World Scientific, Singapore, 1989.

- [70] G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*, 4th ed. Oxford University Press, Oxford, UK, 1960.
- [71] K.P. Harikrishnan and V.M. Nandakumaran, Evidence for the existence of Sarkovskii ordering in a two-dimensional map, *Physics Letters A* 133 (1988) 305.
- [72] B. Hayes, Computer recreations: on the ups and downs of the hailstone numbers, *Scientific American* 250 (1984) 10.
- [73] W. Heisenberg, *Encounters with Einstein: And Other Essays on People, Places, and Particles* Princeton University Press, Princeton, NJ, 1983.
- [74] W. Heisenberg; The Uncertainty Principle, in: *The World of Mathematics*, Vol. 2, ed. J.R. Newman (Tempus, Redmond, WA, 1988) p. 1029.
- [75] H.G.E. Hentschel and I. Procaccia, The Infinite Number of Generalized Dimensions of Fractals and Strange Attractors, *Physica D* 8 (1983) 435.
- [76] C. Herring and J.I. Palmore, Random Number Generators are Chaotic, *SIGPLAN Notices* 24(11) (1989) 76.
- [77] T. Higuchi, Relationship Between the Fractal Dimension and the Power Law Index for a Time Series: A Numerical Investigation, *Physica D* 46 (1990) 254.
- [78] A.V. Holden and M.A. Muhamad; A graphical zoo of strange and peculiar attractors, in: *Chaos*, ed. A.V. Holden (Princeton University Press, Princeton, NJ, 1986) p. 15.
- [79] P. Holmes, 'Strange' phenomena in dynamical systems and their physical implications, *Applied Mathematical Modelling* 1 (1977) 362.
- [80] J.E. Hutchinson, Fractals and self-similarity, *Indiana University Mathematics Journal* 30 (1981) 713.
- [81] E.A. Jackson, *Perspectives on Nonlinear Dynamics*, Vol. 1 Cambridge University Press, Cambridge, 1989.
- [82] E.A. Jackson, *Perspectives on Nonlinear Dynamics*, Vol. 2 Cambridge University Press, Cambridge, 1990.
- [83] E. Jen, Aperiodicity in one-dimensional cellular automata, *Physica D* 45 (1990) 3.
- [84] H. Jones; Fractals Before Mandelbrot: A Selective History, in: *Fractals and*

Chaos, eds. A.J. Crilly, R.A. Earnshaw and H. Jones (Springer-Verlag, New York, 1991) p. 277.

- [85] A. del Junco and J.M. Steele, Hammersley's law for the Van der Corput sequence: an instance of probability theory for pseudorandom numbers, *The Annals of Probability* 7 (1979) 267.
- [86] W.G. Kelly and A.C. Peterson, *Difference Equations: An Introduction with Applications* Academic Press, Boston, MA, 1991.
- [87] M. Kline, *Mathematical Thought from Ancient to Modern Times*, Vol. 1 Oxford University Press, Oxford, 1990.
- [88] M. Kline, *Mathematical Thought from Ancient to Modern Times*, Vol. 3 Oxford University Press, Oxford, 1990.
- [89] J.C. Lagarias, The $3x+1$ problem and its generalizations, *American Mathematical Monthly* 92(1) (1985) 3.
- [90] C.G. Langton; Life at the Edge of Chaos, in: Santa Fe Institute Studies in the Sciences of Complexity, X: *Artificial Life II*, eds. C.G. Langton, C. Taylor, J.D. Farmer and S. Rasmussen (Addison-Wesley, Redwood City, CA, 1992) p. 41.
- [91] P.S. Laplace, *A Philosophical Essay on Probabilities*, 6th ed., New York, 1917.
- [92] H.A. Lauwerier; One-dimensional iterative maps, in: *Chaos*, ed. A.V. Holden (Princeton University Press, Princeton, NJ, 1986) p. 39.
- [93] M. Li and P.M.B. Vitányi; Kolmogorov Complexity and its Applications, in: *Handbook of Theoretical Computer Science*, Vol. A, ed. J. van Leeuwen (North-Holland, Amsterdam, 1990) p. 187.
- [94] Tien-Y. Li and J.A. Yorke, Period three implies chaos, *American Mathematical Monthly* 82 (1975) 985.
- [95] Tien-Y. Li and J.A. Yorke, *Transactions of the American Mathematical Society* 235 (1978) 183.
- [96] E.N. Lorenz, Deterministic Nonperiodic Flow, *Journal of the Atmospheric Sciences* 20 (1963) 130.
- [97] E.N. Lorenz; On the Prevalence of Aperiodicity in Simple Systems, in: *Lecture Notes in Mathematics, 755: Global Analysis*, eds. M. Grmela and J.E. Marsden (Springer-Verlag, Berlin, 1978) p. 53.

- [98] E.N. Lorenz, Computational chaos - a prelude to computational instability, *Physica D* 35 (1989) 299.
- [99] B.B. Mandelbrot, How long is the coast of Britain? Statistical self-similarity and fractional dimension, *Science* 156(5 May) (1967) 636.
- [100] B.B. Mandelbrot, *The Fractal Geometry of Nature* W.H. Freeman and Company, New York, 1983.
- [101] R.M. May, Simple Mathematical Models with Very Complicated Dynamics, *Nature* 261(10 June) (1976) 459.
- [102] R.M. May; Chaos and the Dynamics of Biological Populations, in: *Dynamical Chaos*, eds. M.V. Berry, I.C. Percival and N.O. Weiss (Princeton University Press, Princeton, NJ, 1987) p. 27.
- [103] G. Mayer-Kress, ed., *Dimensions and Entropies in Chaotic Systems* Springer-Verlag, Berlin, 1986.
- [104] J.L.J. McCauley and J.I. Palmore, Computable Chaotic Orbits, *Physics Letters A* 115 (1986) 433.
- [105] P. Mirowski, From Mandelbrot to chaos in economic theory, *Southern Economic Journal*(October) (1990).
- [106] C. Moore, Unpredictability and undecidability in dynamical systems, *Physical Review Letters* 64(20) (1990) 2354.
- [107] J.L. Morris, *Computational Methods in Elementary Numerical Analysis* John Wiley & Sons, Chichester, UK, 1983.
- [108] M. Morse, *Symbolic Dynamics* The Institute for Advanced Study, Princeton, NJ, 1966.
- [109] J.R. Newman; Commentary on Pierre Simon de Laplace, in: *The World of Mathematics*, 2nd ed., Vol. 2, ed. J.R. Newman (Tempus, Redmond, WA, 1988) p. 1291.
- [110] Y. Oono, Period other than 2^n implies chaos, *Progress of Theoretical Physics* 59 (1978) 1028.
- [111] Y. Oono and M. Osikawa, Chaos in Nonlinear Difference Equations I, *Progress of Theoretical Physics* 64 (1980) 54.

- [112] A.R. Osborne and A. Provenzale, Finite correlation dimension for stochastic systems with power-law spectra, *Physica D* 35 (1989) 357.
- [113] E. Ott, Strange attractors and chaotic motions of dynamical systems, *Reviews of Modern Physics* 53(4,October Suppl. 1) (1981) 655.
- [114] N.H. Packard, J.P. Crutchfield, J.D. Farmer and R.S. Shaw, Geometry from a time series, *Physical Review Letters* 45(9) (1980) 712.
- [115] J. Palmore and C. Herring, Computer Arithmetic, Chaos and Fractals, *Physica D* 42 (1990) 99.
- [116] J.I. Palmore and J.L. McCauley, Shadowing by computable chaotic orbits, *Physics Letters A* 122 (1987) 399.
- [117] T.S. Parker and L.O. Chua, *Practical Numerical Algorithms for Chaotic Systems* Springer-Verlag, New York, 1989.
- [118] Heinz-O. Peitgen and P.H. Richter, *The Beauty of Fractals: Images of Complex Dynamical Systems* Springer-Verlag, Berlin, 1986.
- [119] I.C. Percival; Chaos in Hamiltonian Systems, in: *Dynamical Chaos*, eds. M.V. Berry, I.C. Percival and N.O. Weiss (Princeton University Press, Princeton, NJ, 1987) p. 131.
- [120] C.A. Pickover; Patterns in Hailstone $(3n+1)$ Numbers, in: *Computers, Pattern, Chaos and Beauty* (St. Martin's Press, New York, NY, 1990) p. 185.
- [121] H. Poincaré, *Science and Method* Dover, New York, NY, 1952.
- [122] H. Poincaré; Why Space Has Three Dimensions, in: *Mathematics and Science: Last Essays*, ed. J.W. translator Bolduc (Dover, New York, 1963) p. 25.
- [123] D. Pottinger; Data Visualization Techniques for Nonlinear Systems, in: *Fractals and Chaos*, eds. A.J. Crilly, R.A. Earnshaw and H. Jones (Springer-Verlag, New York, 1991) p. 259.
- [124] O.E. Rössler, Chaotic oscillations: an example of hyperchaos, *AMS Lectures in Applied Mathematics* 17 (1979) 141.
- [125] O.E. Rössler, An equation for hyperchaos, *Physics Letters A* 71 (1979) 155.
- [126] O.E. Rössler, The chaotic hierarchy, *Z. Naturforsch* 38a (1983) 788.

- [127] J.-C. Roux, R.H. Simoyi and H.L. Swinney, Observation of a strange attractor, *Physica D* 8 (1983) 257.
- [128] D. Ruelle, Strange attractors, *Mathematical Intelligencer* 2 (1980) 126.
- [129] D. Ruelle, Small random perturbations of dynamical systems and the definition of attractors, *Communications of Mathematical Physics* 82 (1981) 137.
- [130] D. Ruelle, Deterministic chaos: the science and the fiction, *Proceedings of the Royal Society of London* 427A (1990) 241.
- [131] D. Ruelle, *Chance and Chaos* Princeton University Press, Princeton, NJ, 1991.
- [132] D. Ruelle and S. Isola, *Lezioni Lincee: Chaotic Evolution and Strange Attractors* Cambridge University Press, Cambridge, 1989.
- [133] D. Ruelle and F. Takens, On the nature of turbulence, *Communications of Mathematical Physics* 20 (1971) 167.
- [134] A.N. Sarkovsky, Coexistence of the cycles of a continuous mapping on the line into itself, *Ukrainskii matematicheskii zhurnal* 16(1) (1964) 61.
- [135] C. Series, Symbolic dynamics for geodesic flows, *Acta Mathematica* 146(1-2) (1981) 103.
- [136] C. Series; Symbolic dynamics and Diophantine equations, in: *Number Theory and Dynamical Systems*, eds. M.M. Dodson and J.A.G. Vickers (Cambridge University Press, Cambridge, UK, 1989) p. 49.
- [137] R. Shaw, Strange attractors, chaos and information flows, *Z. Naturforsch* 36a (1981) 80.
- [138] G.F. Simmons, *International Series in Pure and Applied Mathematics: Differential Equations: with Applications and Historical Notes* McGraw-Hill, New York, NY, 1972.
- [139] Y.G. Sinai, Dynamical systems with elastic reflections, *Russian Mathematical Surveys* 25 (1970) 137.
- [140] A. Sinha and C. Jayaprakash, A numerical study of one-dimensional cellular automata, *Physica D* 39 (1989) 352.
- [141] S. Smale, Differentiable Dynamical Systems, *Bulletin of the American Mathematical Society* 73 (1967) 747.

- [142] C. Sparrow; The Lorenz equations, in: *Chaos*, ed. A.V. Holden (Princeton University Press, Princeton, NJ, 1986) p. 111.
- [143] C. Sparrow, *Applied Mathematical Sciences*, 41: *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors* Springer-Verlag, New York, 1982.
- [144] I. Stewart, *Opus Books: The Problems of Mathematics* Oxford University Press, Oxford, UK, 1987.
- [145] I. Stewart, *Does God Play Dice: The New Mathematics of Chaos* Penguin Books, London, 1989.
- [146] K. Svozil, Constructive Chaos by Cellular Automata and Possible Sources of an Arrow of Time, *Physica D* 45 (1990) 420.
- [147] H.L. Swinney, Observations of order and chaos in nonlinear systems, *Physica D* 7 (1983) 3.
- [148] H.L. Swinney and J.P. Gollub, Characterization of hydrodynamic strange attractors, *Physica D* 18 (1986) 448.
- [149] S. Takahashi, Cellular Automata and Multifractals: Dimension Spectra of Linear Cellular Automata, *Physica D* 45 (1990) 36.
- [150] F. Takens, Detecting Strange Attractors in Turbulence, *Lecture Notes in Mathematics* 898 (1981) 366.
- [151] R. Terras, A stopping time problem on the positive integers, *Acta Arithmetica* 30 (1976) 241.
- [152] R. Terras, On the existence of a density, *Acta Arithmetica* 35 (1979) 101.
- [153] J. Theiler, Estimating the fractal dimension of chaotic time series, *Lincoln Laboratory Journal* 3(1) (1990) 63.
- [154] C.J. Tricot, Two Definitions of Fractal Dimension, *Mathematical Proceedings of the Cambridge Philosophical Society* 91 (1982) 57.
- [155] E. Ugalde and J. Urías, Cellular Automata on Fractals, *IJCAI-91 Proceedings* (1991).
- [156] J.C. Van der Corput, Verteilungsfunktionen, *Proc. Ned. Akad. v. Wet.* 38 (1935) 813.

- [157] J.A. Vastano, T. Russo and H.L. Swinney, Bifurcation to Spatially Induced Chaos in a Reaction-Diffusion System, *Physica D* 46 (1990) 23.
- [158] N. Vilenkin, *Combinatorial Mathematics For Recreation*, 1st ed. Mir, Moscow, 1972.
- [159] S. Wiggins, Texts in Applied Mathematics, 2: *Introduction to Applied Nonlinear Dynamical Systems and Chaos* Springer-Verlag, New York, NY, 1990.
- [160] S.J. Willson, Growth Rates and Fractional Dimensions in Cellular Automata, *Physica D* 10 (1984) 69.
- [161] S.J. Willson, Cellular Automata Can Generate Fractals, *Discrete Applied Mathematics* 8 (1984) 91.
- [162] S.J. Willson, Computing Fractal Dimensions for Additive Cellular Automata, *Physica D* 24 (1987) 190.
- [163] S.J. Willson, The Equality of Fractional Dimensions for Certain Cellular Automata, *Physica D* 24 (1987) 179.
- [164] J. Wisdom; Two lectures on chaotic dynamics in the solar system, in: NATO ASI Series B: Physics, 176: *Nonlinear Evolution and Chaotic Phenomena*, eds. G. Gallavotti and P.F. Zweifel (Plenum Press, New York, 1988) p. 185.
- [165] A. Wolf; Quantifying chaos with Lyapunov exponents, in: *Chaos*, ed. A.V. Holden (Princeton University Press, Princeton, NJ, 1986) p. 273.
- [166] S. Wolfram, Statistical Mechanics of Cellular Automata, *Reviews of Modern Physics* 55(3,July) (1983) 601.
- [167] S. Wolfram, *Theory and Applications of Cellular Automata* World Scientific, Singapore, 1986.
- [168] M.V. Yakobson, *Communications of Mathematical Physics* 81 (1981) 39.

END

2/3·12·92

FIN