

# **CHAOTIC DYNAMICS: FRACTALS, TILINGS AND SUBSTITUTIONS**

Geoffrey R. Goodson



This book is dedicated to my wife Joyce, my son Garth and my daughters Jacqui and Emma, whose love and support has been my greatest encouragement.

## Preface

Many of the most recent International Congresses of Mathematics have awarded Fields Medals to researchers in chaotic dynamics and related fields, indicating the importance of these areas. Dynamics has blossomed in the past 50 years, making it useful as a tool for demonstrating techniques to mathematics majors and for developing their general mathematical maturity. It is my hope that this book will provide interested students with an introduction to one-dimensional dynamical systems, giving them the tools necessary to succeed in more advanced courses on this topic. The early chapters of this book can be used as a stepping stone from the non-rigorous courses of freshman calculus, to the more advanced topics of real analysis, and topology.

Towson University is a liberal arts college and is part of the University of Maryland System. In my first years of teaching a course on dynamical systems, I based my lectures on the material of some of the existing text books which were then currently available, such as [122], [41], [65] and [32]. Each semester, I found myself changing the course content and exercises (frequently to meet the needs of my students). This led to the production of my own lecture notes (these notes owe a debt to the above mentioned books).

The content of this text arises primarily from lecture notes that I created over many years of teaching senior seminar type courses to final year students at Towson University, and also courses in the Towson University Applied Mathematics Graduate Program, and the Graduate Program in Mathematics Education. In the senior seminar course, students were taught the basics of one-dimensional dynamics, and were required to present a project at completion of the course. The later chapters of this book include many of the topics of these projects, for example, Sharkovsky's Theorem, as well as topics resulting from the independent study of some of my Master's students. With students in the Master's Program, I was generally able to move quickly through the earlier material and spend more time on advanced topics, the choice of which changed from semester to semester.

Students and instructors may find the following information useful. The first two chapters are an introduction to the theory of fixed points and periodic points, describing the behavior of maps under iteration. Newton's method, which is an important theme throughout the text, and elementary bifurcation theory, are discussed. These are reinforced with concrete examples and numerous exercises. Chapter 3 discusses Sharkovsky's Theorem with a proof of the special case of maps having points of period three (the Li-Yorke Theorem). The full proof of Sharkovsky's Theorem is left until Chapter 12. Chapters 4 and 5 lead the student to metric spaces, generalizing results that appear in Chapters 1 and 2, and including important examples such as the Cantor set and the shift map. In Chapters 6, and 7, the notions of chaos and

conjugacy are introduced. Chapters 8 through 14 continue with a study of conjugacy, the Schwarzian derivative, Newton's method, complex dynamics, Sharkovsky's Theorem and some two-dimensional dynamics. The latter third of the book is devoted to topological dynamics on compact spaces and an introduction to substitution dynamics. Throughout this book, I aim to develop the theory in a mathematically rigorous manner. The first 14 chapters (possibly omitting chapters 11 and 12), cover fairly standard topics in (mostly) one-dimensional dynamics, and should be accessible to upper level undergraduate students. The requirements from real analysis and topology (metric spaces), are developed as the material progresses. The text reinforces some of the theoretical results that students have encountered in calculus, such as the Intermediate Value and Mean Value Theorems. A subject such as chaotic dynamics requires a certain amount of mathematical sophistication. It is certainly an advantage for students to have a background in real analysis prior to taking the course, but it is not essential. I feel that this text allows students who have completed a freshman calculus sequence to be successful in a first course in dynamical systems, if this text is followed. Students who do have a previous background in real analysis, will certainly see the importance of real analysis and basic topology in mathematics. I believe that this is not frequently apparent to students in undergraduate studies.

Chapters 15 and 16 give an introduction to the theory of substitutions via examples, and we show how these can give rise to certain types of fractals and tilings. Subsequent chapters develop the rigorous mathematical theory of substitutions and Sturmian sequences. In order to give a rigorous account of symbolic dynamical systems in Chapters 18 and 19, Chapter 17 is devoted to topological dynamical systems, developing the necessary theory, and also introducing concepts related to the material that has preceded it. The final chapter touches on the Multiple Recurrence Theorem of Furstenburg and Weiss (topological version).

This is a mathematical text and I have not focused on applications. Most expositions of one-dimensional dynamical systems are non-rigorous at the undergraduate level or assume a level of sophistication above that of the upper level undergraduate or beginning graduate student. I believe that this book is suitable both for a first course in dynamical systems at the junior or senior level (or even at the sophomore level at some schools). It can also be used in a seminar course on substitution dynamics following a basic dynamics course, or as a supplement for projects in a standard course. A second course can be given to advanced undergraduates, either through independent study, or as a course utilizing the material of chapters 9, 12, 15-20 (also 14 if not done earlier).

Our study of substitutions is combinatorial and topological, avoiding any measure theory. Consequently, we do not touch on some important topics of current interest such as the spectral properties of substitutions (see [104]). It is hoped that our choice of topics will encourage the readers to continue their studies in some aspect of

dynamical systems, such as ergodic theory, topological dynamics, differentiable dynamics, symbolic dynamics, tiling dynamical systems or complex dynamics. We have included much more material than can be covered in a one semester course. A one semester course could consist of Chapters 1 through 14, possibly omitting Chapters 11 and 12. Depending on time constraints, and the level of the student, some or all of Chapters 9 and 13 might also be omitted. If necessary, various topics may be given to the students to read on their own or omitted altogether. These include Sections 2.8 concerning the tent map, 4.4 concerning diffeomorphisms on  $\mathbb{R}$ , and 6.6 giving the dependency of certain of the conditions in the definition of chaos. Section 14.7 is mostly of historical interest, giving the original proof due to Schröeder, of what is often called Cayley's Theorem, and may easily be omitted. Each chapter, and many of the sections are accompanied by exercises that aim to lead the student to a better understanding of the material. An asterisk \* is used to indicate more difficult problems. Much of the material in this text owes a debt to quite recent publications in the field of dynamical systems appearing in journals such as the *American Mathematical Monthly*, the *College Mathematics Journal*, *Mathematical Intelligencer* and *Mathematics Magazine*, and it is very valuable for the students to read some of this original material. Various internet resources have been used, such as Wolfram's MathWorld, some of these without citation because of difficulties in identifying the author. All of the figures in the text were created using Latex, or the computer algebra system Mathematica. The text was typeset with Latex. Computer algebra systems are indispensable tools for studying all aspects of this subject. We have sometimes used a computer algebra system to simplify complicated algebraic manipulations, but generally avoided its use when possible. In teaching this course, I have used Mathematica to illustrate the concepts. A supplement to this course, containing the Mathematica code used can be downloaded at: <http://pages.towson.edu/goodson/>

The lecture notes on which this book is based, benefited tremendously from being read by the students in my class. In particular, I would like to recognize three of my master's students: Nirmal Malapaka, Christopher Jones, and Albandary Alshahrami, all of whom produced Mathematica files that have enabled me to include many of the figures in this text.

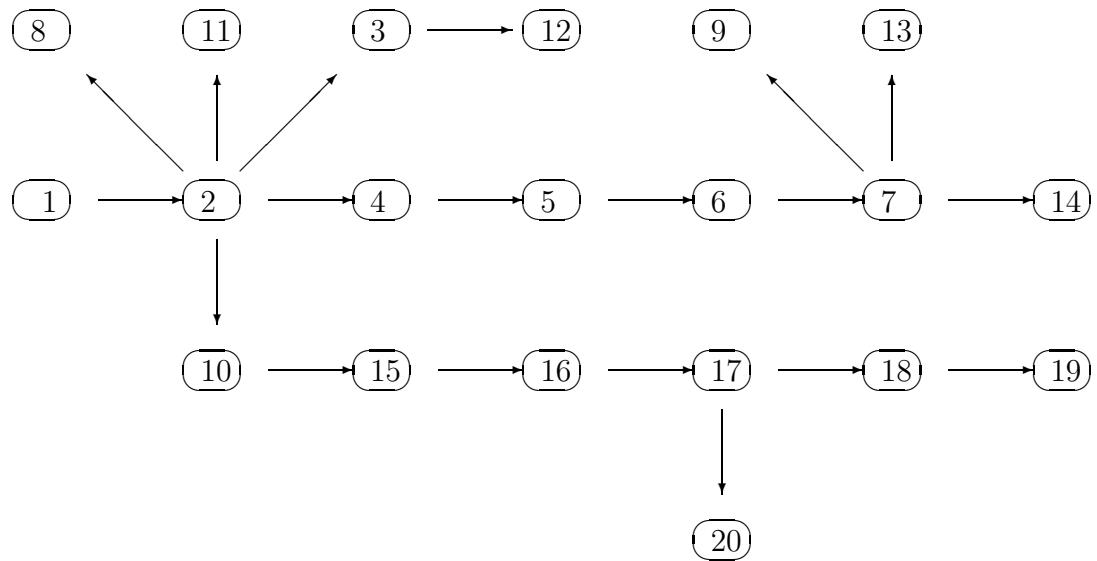
I would like to thank the teachers, who first introduced me to Ergodic Theory, and colleagues who were instrumental for arousing my interest in dynamical systems. These include William Parry, Peter Walters and Rufus Bowen at the University of Warwick, U.K., Dan Newton at the University of Sussex, U.K., Michael Sears and Harvey Keynes at the University of the Witwatersrand, Johannesburg and Dan Rudolph at the University of Maryland.

I am indebted to Towson University and my colleagues there, Angel Kumchev and Houshang Sohrab, for reading parts of this text and for their valuable input. I am grateful to the early reviewers (before this text was fully written), for their insight and excellent suggestions which have made this a better book. I would like to thank my

wife Joyce, who with a keen mathematical eye read this text. This project would have been impossible without her support, invaluable help and understanding. Finally, I would like to thank Kaitlin Leach at Cambridge University Press for her interest in this book, and for all the help from people at CUP in moving this book along.

I am solely responsible for any errors that may have occurred, and I welcome any comments from the readers of this book.

### Chapter Dependency.



Geoffrey Goodson  
 Mathematics Department  
 Towson University.  
 August, 2015  
 e-mail: ggoodson@towson.edu



## Contents

<b>Chapter 1. The Orbits of One-Dimensional Maps.</b>	1
1.1 Iteration of Functions and Examples of Dynamical Systems.	1
1.2 Newton's Method and Fixed Points.	10
1.3 Graphical Iteration.	18
1.4 The Stability of Fixed Points.	22
1.5 Non-hyperbolic Fixed Points.	30
<b>Chapter 2. Bifurcations and the Logistic Family.</b>	41
2.1 The Basin of Attraction.	41
2.2 The Logistic Family.	43
2.3 Periodic Points.	48
2.4 Periodic Points of the Logistic Map.	55
2.5 The Period Doubling Route to Chaos.	58
2.6 The Bifurcation Diagram and 3-Cycles of the Logistic Map.	59
2.7 The Tent Family $T_\mu$ .	69
2.8 The 2-Cycles and 3-Cycles of the Tent Family.	70
<b>Chapter 3. Sharkovsky's Theorem.</b>	75
3.1 Period Three Implies Chaos.	75
3.2 Converse of Sharkovsky's Theorem.	80
<b>Chapter 4. Dynamics on Metric Spaces.</b>	85
4.1 Basic Properties of Metric Spaces.	85
4.2 Dense Sets.	89
4.3 Functions Between Metric Spaces.	94
4.4 Diffeomorphisms of $\mathbb{R}$ .	101
<b>Chapter 5. Countability, Sets of Measure Zero and the Cantor Set.</b>	107
5.1 Countability and Sets of Measure Zero.	107
5.2 The Cantor Set.	112
5.3 Ternary Expansions and the Cantor Set.	115
5.4 The Tent Map for $\mu = 3$ .	119
5.5 A Cantor Set Arising From the Logistic Map $L_\mu$ , $\mu > 4$ .	121
<b>Chapter 6. Devaney's Definition of Chaos.</b>	125

6.1 The Doubling Map and the Angle Doubling Map.	126
6.2 Transitivity.	128
6.3 Sensitive Dependence on Initial Conditions.	130
6.4 The Definition of Chaos.	131
6.5 Symbolic Dynamics and the Shift Map.	135
6.6 For Continuous Maps, Sensitive Dependence is Implied by Transitivity and Dense Period Points.	139
<b>Chapter 7. Conjugacy of Dynamical Systems.</b>	143
7.1 Conjugate Maps.	143
7.2 Properties of Conjugate Maps and Chaos Through Conjugacy.	146
7.3 Linear Conjugacy.	152
<b>Chapter 8. Singer's Theorem.</b>	157
8.1 The Schwarzian Derivative Revisited.	157
8.2 Singer's Theorem.	161
<b>Chapter 9. Conjugacy, Fundamental Domains and the Tent Family.</b>	167
9.1 Conjugacy and Fundamental Domains.	167
9.2 Conjugacy, the Tent Map and Periodic Points of the Tent Family.	172
<b>Chapter 10. Fractals.</b>	179
10.1 Examples of Fractals.	179
10.2 An Intuitive Introduction to the Idea of Fractal Dimension.	181
10.3 Box Counting Dimension.	182
10.4 The Mathematical Theory of Fractals.	187
10.5 The Contraction Mapping Theorem and Self-Similar Sets.	189
<b>Chapter 11. Newton's Method for Real Quadratics and Cubics.</b>	197
11.1 Binary Representation of Real Numbers.	197
11.2 Newton's Method for Real Quadratic Polynomials.	199
11.3 Newton's Method for Real Cubic Polynomials.	201
11.4 The Cubic Polynomials $f_c(x) = (x + 2)(x^2 + c)$ .	203
<b>Chapter 12. Coppel's Theorem and a Proof of Sharkovsky's Theorem.</b>	209
12.1 Coppel's Theorem.	209
12.2 The Proof of Sharkovsky's Theorem.	213
12.3 The Completion of the Proof of Sharkovsky's Theorem.	218
<b>Chapter 13. Real Linear Transformations, the Hénon Map, and Hyperbolic Toral Automorphisms.</b>	223
13.1 Linear Transformations.	223
13.2 The Hénon Map.	231
13.3 Circle Maps Induced by Linear Transformations on $\mathbb{R}$ .	234

13.4 Endomorphisms of the Torus.	235
13.5 Hyperbolic Toral Automorphisms.	238
<b>Chapter 14. Elementary Complex Dynamics.</b>	243
14.1 The Complex Numbers.	243
14.2 Analytic Functions in the Complex Plane.	244
14.3 The Dynamics of Polynomials and the Riemann Sphere.	249
14.4 The Julia Set.	254
14.5 The Mandelbrot Set $\mathcal{M}$ .	266
14.6 Newton's Method in the Complex Plane for Quadratics and Cubics.	273
14.7 Important Complex Functions.	281
<b>Chapter 15. Examples of Substitutions.</b>	291
15.1 One-dimensional Substitutions and the Thue-Morse Substitution.	291
15.2 The Toeplitz Substitution.	299
15.3 The Rudin-Shapiro Sequence.	301
15.4 Paperfolding Sequences.	304
<b>Chapter 16. Fractals Arising from Substitutions.</b>	311
16.1 A Connection Between the Morse Substitution and the Koch Curve.	311
16.2 Dragon Curves.	314
16.3 Fractals Arising from Two-Dimensional Substitutions	316
16.4 The Rauzy Fractal.	323
<b>Chapter 17. Compactness in Metric Spaces and an Introduction to Topological Dynamics.</b>	335
17.1 Compactness in Metric Spaces.	335
17.2 Continuous Functions on Compact Metric Spaces.	340
17.3 The Contraction Mapping Theorem for Compact Metric Spaces.	342
17.4 Basic Topological Dynamics.	343
17.5 Topological Mixing and Exactness.	353
<b>Chapter 18. Substitution Dynamical Systems.</b>	361
18.1 Sequence Spaces.	361
18.2 Languages.	367
18.3 Dynamical Systems Arising from Sequences.	369
18.4 Substitution Dynamics.	375
<b>Chapter 19. Sturmian Sequences and Irrational Rotations.</b>	381
19.1 Sturmian Sequences.	381
19.2 Sequences Arising From Irrational Rotations.	385
19.3 Cutting Sequences.	390
19.4 Sequences Arising from Irrational Rotations are Sturmian.	393
19.5 Semi-Topological Conjugacy Between $([0, 1], T_\alpha)$ and $(\overline{O(u)}, \sigma)$ .	397

xii	
19.6 The Three Distance Theorem.	400
Chapter 20. The Multiple Recurrence Theorem of Furstenberg and Weiss.	405
20.1 van der Waerden's Theorem.	405
Appendix A. Theorems from Calculus.	411
Appendix B. The Baire Category Theorem.	413
Appendix C. The Complex Numbers.	415
Appendix D. Weyl's Equidistribution Theorem.	417
Bibliography	419
Index	425

## CHAPTER 1

### The Orbits of One-Dimensional Maps.

In this chapter we introduce one-dimensional dynamical systems and analyze some elementary examples. A study of the iteration in Newton's method leads naturally to the notion of attracting fixed points for dynamical systems. Newton's method is emphasized throughout as an important motivation for the study of dynamical systems. The chapter concludes with various criteria for establishing the stability of the fixed points of a dynamical system.

#### 1.1 Iteration of Functions and Examples of Dynamical Systems.

Chaotic dynamical systems has its origins in Henri Poincaré's memoir on celestial mechanics and the three-body problem (1890's). Poincaré's memoir arose from his entry in a competition celebrating the 60th birthday of King Oscar of Sweden. His manuscript concerned the stability of the solar system and the question of how three bodies, with mutual gravitational interaction, behave. This was a problem that had been solved for two bodies by Isaac Newton. Although Poincaré was not able to determine exact solutions to the three-body problem, his study of the long term behavior of such dynamical systems resulted in a prize winning manuscript. In particular, he claimed that the solutions to the three-body problem (restricted to the plane) are stable, so that a solar system such as ours would continue orbiting more or less as it does, forever. After the competition, and when his manuscript was ready for publication, he noticed it contained a deep error which showed that instability may arise in the solutions. In correcting the error, Poincaré discovered chaos and his memoir became one of the most influential scientific publications of the past century [10]. Aspects of dynamical systems were already evident in the study of iteration in Newton's method for approximating the zeros of functions. The work of Cayley and Schroeder concerning Newton's method in the complex domain appeared during the 1880's, and interest in this new field of complex dynamics continued in the early 1900's with the work of Fatou and Julia. Their work lay dormant until the invention of the electronic computer. In the 1960's the subject exploded into life with the work of Sharkovsky and Li-Yorke on one-dimensional dynamics, and with Kolmogorov, Smale, Anosov

and others, with differentiable dynamics and ergodic theory. The advent of computer graphics allowed for the resurgence of complex dynamics and the depiction of fractals (Devaney and Mandelbrot).

This book is mainly concerned with one-dimensional dynamical systems for real and complex mappings and their connections with fractal geometry. We also treat certain symbolic dynamical systems in detail, in particular we look at substitution dynamical systems and the fractals they generate.

Dynamical systems is the study of how things change over time. Examples include the growth of populations, the change in the weather, radioactive decay, mixing of liquids such as the ocean currents, motion of the planets, the interest in a bank account. Some of these dynamical systems are well behaved and predictable, for example, if you know how much money you have in the bank today, it should be possible to calculate how much you will have next month (based on how much you deposit, interest rate etc.). However, some dynamical systems are inherently unpredictable and so are called chaotic. An example of this is weather forecasting, which is generally unreliable beyond predicting weather for the next three or four days. Intuition tells us that chaotic behavior will happen provided we have some degree of randomness in the system. However, chaos can happen even when the dynamical system is deterministic, that is, its future behavior is completely determined by its initial conditions. To quote Edward Lorenz, who was the first to realize that deterministic chaos is present in weather forecasting: Chaos is “when the present determines the future, but the approximate present does not approximately determine the future”. In theory, if we could measure exactly the weather at some instant in time at every point in the earth’s atmosphere, we could predict how it will behave in the future. But because we can only approximately measure the weather (wind speed and direction, temperature etc.), the future weather is unpredictable.

Throughout we use  $\mathbb{R}$  to denote the set of real numbers,  $\mathbb{Z} = \{\dots, -1, 0, 1, 2, 3, \dots\}$  is the set of integers,  $\mathbb{N} = \{0, 1, 2, \dots\}$  are the natural numbers and  $\mathbb{Z}^+ = \{1, 2, 3, \dots\}$  are the positive integers and  $\mathbb{Q}$  is the set of rational numbers.

Dynamical systems with continuously varying time, (which are called *flows*), arise from the solutions to differential equations. In this text, we will study *discrete dynamical systems*, arising from discrete changes in time. For example, we might model a population by measuring it daily. Suppose that  $x_n$  is the number of members of a population on day  $n$ , where  $x_0$  is the initial population. We look for a function

$f : \mathbb{R} \rightarrow \mathbb{R}$ , for which

$$x_1 = f(x_0), x_2 = f(x_1), \text{ and generally } x_n = f(x_{n-1}), n = 1, 2, \dots$$

This leads to the iteration of functions in the following way:

**Definition 1.1.1** Given  $x_0 \in \mathbb{R}$ , the *orbit* of  $x_0$  under  $f$  is the set

$$O(x_0) = \{x_0, f(x_0), f^2(x_0), \dots\},$$

where  $f^2(x_0) = f(f(x_0))$ ,  $f^3(x_0) = f(f^2(x_0))$ , and continuing indefinitely, so that

$$f^n(x) = f \circ f \circ f \circ \dots \circ f(x); \quad (n\text{-times composition}).$$

For each  $n \in \mathbb{N}$ , set  $x_n = f^n(x_0)$ , then  $x_1 = f(x_0)$ ,  $x_2 = f^2(x_0)$ , and in general

$$x_{n+1} = f^{n+1}(x_0) = f(f^n(x_0)) = f(x_n).$$

More generally,  $f$  may be defined on some subinterval  $I$  of  $\mathbb{R}$ , but in order for the iterates of  $x \in I$  under  $f$  to be defined, we need the *range* of  $f$  to be contained in  $I$ , so  $f : I \rightarrow I$  (both the *domain* and the *codomain* of  $f$  are the same set).

Thus we are studying the iterations of *one-dimensional maps*, (as opposed to higher dimensional maps  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $n > 1$ , some of which will be considered in Chapter 13).

**Definition 1.1.2** A (one-dimensional) *dynamical system* is a pair  $(I, f)$ , where  $f$  is a function  $f : I \rightarrow I$  and  $I$  is a subset of  $\mathbb{R}$ . Almost always,  $I$  will be a subinterval of  $\mathbb{R}$ , which includes the possibility that  $I = \mathbb{R}$ .

Often we will talk about the dynamical system  $f : I \rightarrow I$ , or just  $f$  when the domain is clear. Usually,  $f$  is assumed to be a continuous function, but we occasionally relax this requirement. For example,  $f : [0, 1] \rightarrow [0, 1]$ ,  $f(x) = x^2$  and  $g : [0, 1] \rightarrow [0, 1]$ ,  $g(x) = 2x$  if  $0 \leq x < 1/2$  and  $g(x) = 2x - 1$  if  $1/2 \leq x \leq 1$  are dynamical systems (the latter is not continuous), but  $h : [0, 2] \rightarrow [0, 4]$ ,  $h(x) = x^2$  is not a dynamical system, since the domain and codomain are different.

Given a dynamical system  $f$ , equations of the form  $x_{n+1} = f(x_n)$  are examples of *difference equations*. These arise from one-dimensional dynamical systems. For example,  $x_n$  may represent the number of bacteria in a culture after  $n$  hours, or the mass of radioactive material remaining after  $n$  days of an experiment. There is an obvious correspondence between one-dimensional maps and these difference

equations. For example, a difference equation commonly used for calculating square roots:

$$x_{n+1} = \frac{1}{2}(x_n + \frac{2}{x_n}),$$

corresponds to the function  $f(x) = \frac{1}{2}(x + \frac{2}{x})$ . If we start by setting  $x_0 = 2$  (or in fact any real number), and then find  $x_1, x_2, \dots$  etc., we get a sequence which rapidly approaches  $\sqrt{2}$  (see page 9 of Sternberg [122]). One of the issues we examine in this chapter is how this difference equation arises and its usefulness in calculating square roots.

### Examples of Dynamical Systems 1.1.3

**1. The Trigonometric Functions.** Consider the iterations of the trigonometric function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = \sin(x)$ . Select  $x_0 \in \mathbb{R}$  at random, e.g.,  $x_0 = 2$  and set  $x_{n+1} = \sin(x_n)$ ,  $n = 0, 1, 2, \dots$ . What happens to  $x_n$  as  $n$  increases? One way to investigate this type of dynamical system is to use a graphing utility: enter **Sin(2)**, followed by **ENTER**, and then **Sin(ANSWER)**, and then continue this process. You will need to do this many times to get a good idea of what is happening. It may be easier to use a computer algebra system to carry out the computations.

Now replace the sine function with the cosine function and repeat the process. How do we explain what appears to be happening in each case? These are questions that will be answered in this chapter.

**2. Linear Maps.** These are possibly the simplest dynamical systems for modeling population growth and also the easiest to deal with from a dynamical point of view, since we can obtain a clear description of their long term behavior. Every linear map  $f : \mathbb{R} \rightarrow \mathbb{R}$  is of the form  $f(x) = a \cdot x$  for some  $a \in \mathbb{R}$ . Suppose that  $x_n$  = size of a population at time  $n$ , with the property

$$x_{n+1} = a \cdot x_n,$$

for some constant  $a > 0$ . This is an example of a *linear model* for the growth of the population.

If the initial population is  $x_0 > 0$ , then  $x_1 = ax_0$ ,  $x_2 = ax_1 = a^2x_0$ , and in general  $x_n = a^n x_0$  for  $n = 0, 1, 2, \dots$ . This is the exact solution (or *closed form solution*) to the difference equation  $x_{n+1} = a \cdot x_n$ . Clearly  $f(x) = ax$  is the corresponding dynamical system. We can use the solution to determine the long term behavior of the population:

The sequence  $(x_n)$  is very well behaved since:

- (i) if  $a > 1$ , then  $x_n \rightarrow \infty$  as  $n \rightarrow \infty$ ,
- (ii) if  $0 < a < 1$  then  $x_n \rightarrow 0$  as  $n \rightarrow \infty$  (i.e., the population becomes extinct),
- (iii) if  $a = 1$ , then the population remains unchanged.

**3. Affine maps.** These are functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  of the form  $f(x) = ax + b$ , ( $a \neq 0$ ), for constants  $a$  and  $b$ . Consider the iterates of such maps:

$$\begin{aligned} f^2(x) &= f(ax + b) = a(ax + b) + b = a^2x + ab + b, \\ f^3(x) &= a^3x + a^2b + ab + b, \\ f^4(x) &= a^4x + a^3b + a^2b + ab + b, \end{aligned}$$

and generally

$$f^n(x) = a^n x + a^{n-1}b + a^{n-2}b + \cdots + ab + b.$$

Let  $x_0 \in \mathbb{R}$  and set  $x_n = f^n(x_0)$ , then we have

$$\begin{aligned} x_n &= a^n x_0 + (a^{n-1} + a^{n-2} + \cdots + a + 1)b \\ &= a^n x_0 + b \left( \frac{a^n - 1}{a - 1} \right), \quad \text{if } a \neq 1, \end{aligned}$$

or

$$x_n = \left( x_0 + \frac{b}{a - 1} \right) a^n + \frac{b}{1 - a}, \quad \text{if } a \neq 1,$$

is the closed form solution. Here we have used the formula for the sum of a finite geometric series:

$$\sum_{k=0}^{n-1} r^k = \frac{r^n - 1}{r - 1},$$

when  $r \neq 1$ . If  $a = 1$ , the solution is  $x_n = x_0 + nb$ .

We can use these equations to determine the long term behavior of  $x_n$ . We see that:

- (i) if  $|a| < 1$  then  $a^n \rightarrow 0$  as  $n \rightarrow \infty$ , so that

$$\lim_{n \rightarrow \infty} x_n = \frac{b}{1 - a},$$

- (ii) if  $a > 1$ , then  $\lim_{n \rightarrow \infty} x_n = \infty$  for  $b, x_0 > 0$ ,
- (iii) if  $a = 1$ , then  $\lim_{n \rightarrow \infty} x_n = \infty$  if  $b > 0$ .

The limit does not exist if  $a \leq -1$ .

### 1.1.4 Recurrence Relations.

Many sequences can be defined *recursively* by specifying the first few terms, and then stating a general rule which specifies how to obtain the  $n$ th term from the  $(n - 1)$ th term (or other additional terms), and using mathematical induction to see that the sequence is “well defined” for every  $n \in \mathbb{N}$ . For example,  $n! = n\text{-factorial}$  can be defined in this way by specifying  $0! = 1$ , and  $n! = n \cdot (n - 1)!$ , for  $n \in \mathbb{Z}^+$ . The *Fibonacci sequence* ( $F_n$ ), can be defined by setting

$$F_0 = 0, \quad F_1 = 1 \quad \text{and} \quad F_{n+2} = F_{n+1} + F_n, \quad \text{for } n \geq 0,$$

so that  $F_2 = 1$ ,  $F_3 = 2$ , giving the sequence  $0, 1, 1, 2, 3, 5, 8, 13, 21, \dots$ .

The orbit of a point  $x_0 \in \mathbb{R}$  under a function  $f$  is then defined recursively as follows:

$$x_n = f(x_{n-1}), \quad \text{for } n \in \mathbb{Z}^+,$$

with a given starting value  $x_0$ . The principle of mathematical induction tells us that  $x_n$  is defined for every  $n \geq 0$ , since it is defined for  $n = 0$ . Assuming it has been defined for  $k = n - 1$  then  $x_n = f(x_{n-1})$  defines it for  $k = n$ .

Ideally, given a recursively defined sequence  $(x_n)$ , we would like to have a specific formula for  $x_n$  in terms of elementary functions (a so called *closed form solution*). This is often very difficult, or impossible to achieve. In the case of affine maps and certain logistic maps, there is a closed form solution. One can use these ideas to study certain problems as illustrated in the following examples.

**Example 1.1.5** An amount  $\$ T$  is deposited in your bank account at the end of each month. The interest is  $r\%$  per month. Find the amount  $A(n)$  accumulated at the end of  $n$  months (assume  $A(0) = T$ ).

**Answer.**  $A(n)$  satisfies the difference equation

$$A(n+1) = A(n) + A(n)r/100 + T, \quad \text{where } A(0) = T,$$

or

$$A(n+1) = A(n)(1 + r/100) + T.$$

Setting  $x_0 = T$ ,  $a = 1 + r/100$  and  $b = T$  in the formula of Example 1.1.3 (3), the solution is

$$\begin{aligned} A(n) &= (1 + r/100)^n T + T \left( \frac{(1 + r/100)^n - 1}{1 + r/100 - 1} \right) \\ &= (1 + r/100)^n T + 100 \frac{T}{r} ((1 + r/100)^n - 1). \end{aligned}$$

### 1.1.6 The Logistic Map.

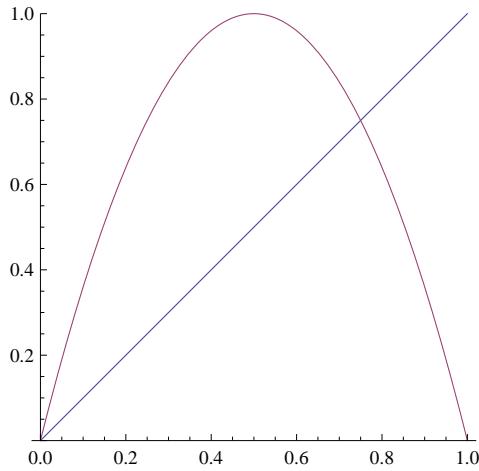
In the late 1940's, John von Neumann proposed that the map  $f(x) = 4x(1 - x)$  could be used as a pseudo-random number generator. Maps of this type were amongst the first to be studied using electronic computers. Paul Stein and Stanislaw Ulam did an extensive computer study of  $f(x)$  and related maps in the early 1950's, but much about these maps remained mysterious.

More generally, maps of the form

$$L_\mu : \mathbb{R} \rightarrow \mathbb{R}, \quad L_\mu(x) = \mu x(1 - x),$$

were proposed to model a certain type of population growth (see the work of Robert May [87]). Here  $\mu$  is a real parameter which is fixed. Note that if  $0 < \mu \leq 4$ , then  $L_\mu$  is a dynamical system of the interval  $[0, 1]$ , i.e.  $L_\mu : [0, 1] \rightarrow [0, 1]$ . For example, when  $\mu = 4$ ,  $L_4(x) = 4x(1 - x)$ , with  $L_4([0, 1]) = [0, 1]$  with graph given in the figure below. If  $\mu > 4$ ,  $L_\mu$  is no longer a dynamical system of  $[0, 1]$  as  $L_\mu([0, 1])$  is not a subset of  $[0, 1]$ .

Historically, population biologists were interested in those values of  $\mu$  that give rise to stable populations after long term iteration. However, we shall see that as  $\mu$  becomes close to 4, the long term behavior becomes highly unstable. The chaotic nature of this behavior was first pointed out by James Yorke in 1975. During a visit to Yorke at the University of Maryland, Robert May mentioned that he did not understand what happens to  $L_\mu$  as  $\mu$  approaches 4. Shortly after this, the seminal works of Li-Yorke ([84], 1975) and May ([87], 1976), appeared.



The logistic map with  $\mu = 4$ .

**Remark 1.1.7** It is conjectured that closed form solutions for the difference equation arising from the logistic map are only possible when  $\mu = -2$ ,  $\mu = 2$  or  $\mu = 4$  (see Exercises 1.1 # 3 for the cases where  $\mu = 2$ ,  $\mu = 4$  and # 13 for the case where  $\mu = -2$ , and also [128] for a discussion of this conjecture).

### Exercises 1.1

1. If  $L_\mu(x) = \mu x(1 - x)$  is the logistic map, calculate  $L_\mu^2(x)$  and  $L_\mu^3(x)$ .
2. Use Example 1.1.3 for affine maps to find the solutions to the difference equations:
  - (i)  $x_{n+1} - \frac{x_n}{3} = 2$ ,  $x_0 = 2$ ,
  - (ii)  $x_{n+1} + 3x_n = 4$ ,  $x_0 = -1$ .
3. A *logistic difference equation* is one of the form  $x_{n+1} = \mu x_n(1 - x_n)$  for some fixed  $\mu \in \mathbb{R}$ . Find exact, (closed form) solutions, to the following logistic difference equations:
  - (i)  $x_{n+1} = 2x_n(1 - x_n)$ . (Hint: Use the substitution  $x_n = (1 - y_n)/2$  to transform the equation into a simpler equation that is easily solved).
  - (ii)  $x_{n+1} = 4x_n(1 - x_n)$ . (Hint: Set  $x_n = \sin^2(\theta_n)$  and simplify to get an equation that is easily solved).
4. You borrow  $\$P$  at  $r\%$  per annum, and pay off  $\$M$  at the end of each subsequent month. Write down a difference equation for the amount owing  $A(n)$  at the end of each month (so  $A(0) = P$ ). Solve the equation to find a closed form for  $A(n)$ . If  $P = 100,000$ ,  $M = 1000$  and  $r = 4$ , after how long will the loan be paid off?

5. At 70.5 years of age, you have  $\$A$  invested in a pre-tax retirement account. It is earning interest at  $r_1\%$  per annum. The tax laws require you to take out  $r_2\%$  per annum of what is remaining in the account ( $r_2 > r_1$ , where in practice  $r_2 = 3.65$ ). How much is remaining after  $n$  years?

Solve this problem with  $\$A = \$500,000$ ,  $r_1 = 3\%$ ,  $r_2 = 3.65\%$  and  $n = 15$  years.

6. If  $T_\mu(x) = \begin{cases} \mu x; & 0 \leq x < 1/2 \\ \mu(1-x); & 1/2 \leq x \leq 1 \end{cases}$ , show that  $T_\mu$  is a dynamical system of  $[0, 1]$ , for  $\mu \in (0, 2]$

7. Let  $f(x) = x^2 + bx + c$ . Give conditions on  $b$  and  $c$  for  $f : [0, 1] \rightarrow [0, 1]$  to be a dynamical system. (Hint: Recall that the maximum and minimum values of a continuous function defined on a closed interval  $[a, b]$  occur either at the end points, or where  $f'(x) = 0$ , or where  $f'(x)$  does not exist).

8. Determine whether the functions  $f$  defined below can be considered as dynamical systems  $f : I \rightarrow I$ :

(a)  $f(x) = x^3 - 3x$ , (i)  $I = [-1, 1]$ , (ii)  $I = [-2, 2]$ .

(b)  $f(x) = 2x^3 - 6x$ , (i)  $I = [-1, 1]$ , (ii)  $I = \left[-\sqrt{\frac{7}{2}}, \sqrt{\frac{7}{2}}\right]$ .

9. If  $f_\mu(x) = \mu x^2 \frac{1-x}{1+x}$ , show that for  $0 < \mu < (5\sqrt{5} + 11)/2$ ,  $f_\mu$  is a dynamical system of  $[0, 1]$ .

10. For the following functions, find  $f^2(x)$ ,  $f^3(x)$  and a general formula for  $f^n(x)$ :

(i)  $f(x) = x^2$ , (ii)  $f(x) = |x + 1|$ , (iii)  $f(x) = \begin{cases} 2x; & 0 \leq x < 1/2 \\ 2x - 1; & 1/2 \leq x < 1. \end{cases}$

11. Use mathematical induction to show that if  $f(x) = \frac{2}{x+1}$ , then

$$f^n(x) = \frac{2^n(x+2) + (-1)^n(2x-2)}{2^n(x+2) - (-1)^n(x-1)}.$$

12. (a) The *tribonacci sequence* ( $T_n$ ) is a generalization of the Fibonacci sequence, defined recursively by

$$T_0 = 0, T_1 = 0, T_2 = 1, \quad T_{n+1} = T_n + T_{n-1} + T_{n-2}, \quad n \geq 2.$$

Write down the first 10 terms of  $T_n$ .

- (b) Let  $(F_n)$  be the Fibonacci sequence. Set  $v_n = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}$ , and  $F = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ . Show that  $v_{n+1} = F \cdot v_n$ ,  $n \geq 0$ .

- (c) Find a matrix  $T$  such that if  $(T_n)$  is the tribonacci sequence, and  $w_n = \begin{pmatrix} T_{n+2} \\ T_{n+1} \\ T_n \end{pmatrix}$ , then  $w_{n+1} = T \cdot w_n$ ,  $n \geq 0$ .

- 13\*. Show that a closed form solution to the logistic difference equation when  $\mu = -2$  is given by

$$x_n = \frac{1}{2} [1 - f[r^n f^{-1}(1 - 2x_0)]], \quad \text{where } r = -2 \quad \text{and} \quad f(\theta) = 2 \cos\left(\frac{\pi - \sqrt{3}\theta}{3}\right).$$

(Hint: Set  $x_n = \frac{1 - f(\theta_n)}{2}$  and use steps similar to 3(ii)).

## 1.2 Newton's Method and Fixed Points.

Isaac Newton (1669) and Joseph Raphson (1690) gave special cases of what we now call Newton's method, with the modern version being given by Thomas Simpson in 1740. Newton's method is an algorithm for rapidly finding the approximate values of zeros of functions.

Given a differentiable function  $f : \mathbb{R} \rightarrow \mathbb{R}$  and under suitable conditions, Newton's method allows us to find good approximations to zeros of  $f(x)$ , i.e., approximate solutions to the equation  $f(x) = 0$ . The idea is to start with a first approximation

$x_0$ , and look at the tangent line to  $f(x)$  at the point  $(x_0, f(x_0))$ . Suppose this line intersects the  $x$ -axis at  $x_1$ , then

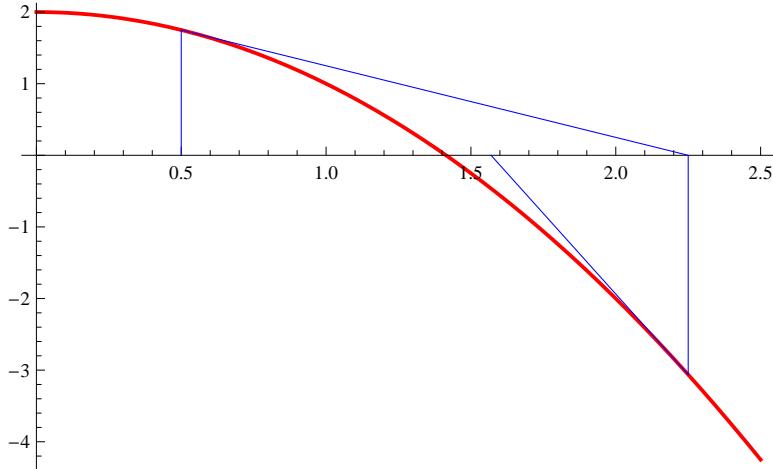
$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \quad \text{if } f'(x_0) \neq 0.$$

If our initial guess  $x_0$  is close enough to the zero,  $x_1$  will be a better approximation to the zero. Repeat the process with the tangent line to  $f(x)$  at  $(x_1, f(x_1))$ . At the  $n + 1$ th stage we obtain

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},$$

an algorithm in the form of a difference equation, where  $x_0$  is a first approximation to a zero of  $f(x)$ . The corresponding real function is

$$N_f(x) = x - \frac{f(x)}{f'(x)}, \quad (\text{the Newton function}).$$



The first two approximations for solving  $f(x) = 2 - x^2 = 0$ , starting with  $x_0 = .5$ .

For example, if  $f(x) = x^2 - a$ , then  $f'(x) = 2x$  and

$$N_f(x) = x - \frac{x^2 - a}{2x} = \frac{1}{2} \left( x + \frac{a}{x} \right),$$

so that

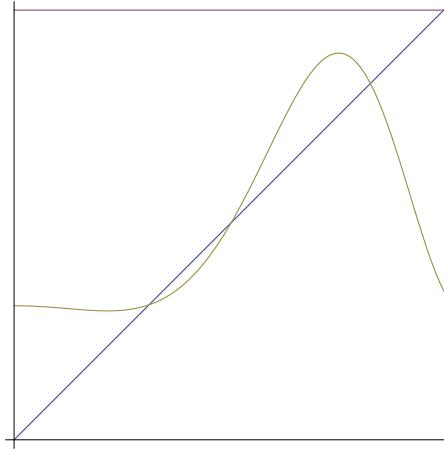
$$x_{n+1} = \frac{1}{2} \left( x_n + \frac{a}{x_n} \right),$$

giving the difference equation we mentioned in Section 1.1 that is used for approximating  $\sqrt{2}$  when  $a = 2$ .

Note that if  $f(x) = 0$ , then  $N_f(x) = x - \frac{f(x)}{f'(x)} = x$  and conversely, if  $N_f(x) = x$  then  $f(x) = 0$ . This suggests that points where a function  $g(x)$  satisfies  $g(x) = x$  are important. This leads to:

**Definition 1.2.1** Let  $f : I \rightarrow I$ , where  $I$  is a subinterval of  $\mathbb{R}$ . A point  $c \in I$ , for which  $f(c) = c$ , is called a *fixed point* of  $f$ .

A fixed point of  $f$  is a point where the graph of  $f(x)$  intersects the line  $y = x$ . We denote the set of fixed points of  $f$  by  $\text{Fix}(f)$ .



Fixed points occur where the graph of  $f(x)$  intersects the line  $y = x$ .

### Examples 1.2.2

1. Suppose that  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^2$ , then  $x^2 = x$  gives  $x(x - 1) = 0$ , so has fixed points  $c = 0$  and  $c = 1$ , and  $\text{Fix}(f) = \{0, 1\}$ .
2. If  $f(x) = x^3 - x$ , then  $x^3 - x = x$  gives  $x(x^2 - 2) = 0$ , so the fixed points are  $c = 0$  and  $c = \pm\sqrt{2}$ , and  $\text{Fix}(f) = \{0, \pm\sqrt{2}\}$ .
3. We are interested in the logistic map  $L_\mu(x) = \mu x(1 - x) = \mu x - \mu x^2$  for  $0 < \mu \leq 4$ , since for these parameter values  $\mu$ ,  $L_\mu$  is a dynamical system of  $[0, 1]$ . If  $\mu > 4$ ,  $L_\mu(x) > 1$  for some values of  $x$  in  $[0, 1]$ , and further iterates will go to  $-\infty$ .

Solving  $L_\mu(x) = x$  gives  $x = 0$  or  $x = 1 - 1/\mu$ . If  $0 < \mu \leq 1$ , then  $1 - 1/\mu \leq 0$ , so  $c = 0$  is the only fixed point in  $[0, 1]$ .

The logistic map  $L_4(x) = 4x(1 - x) = 4x - 4x^2$ ,  $0 \leq x \leq 1$ , has the properties:  $L_4(0) = 0$ , (a fixed point),  $L_4(1) = 0$ , a maximum at  $x = 1/2$  (with  $L_4(1/2) = 1$ ). Solving  $L_4(x) = x$  gives  $4x - 4x^2 = x$ , so  $4x^2 = 3x$ , and the fixed points are  $c = 0$  and  $c = 3/4$ .

This map also has what we call *eventual fixed points*:

**Definition 1.2.3** We say that  $x^* \in \mathbb{R}$  is an *eventual fixed point* of  $f(x)$  if there exists a fixed point  $c$  of  $f(x)$  and  $r \in \mathbb{Z}^+$  satisfying  $f^r(x^*) = c$ , but  $f^s(x^*) \neq c$  for  $0 \leq s < r$ .

For  $L_4(x) = 4x(1-x)$ ,  $L_4(1) = 0$  and  $L_4(0) = 0$ , so that  $c = 1$  is eventually fixed. Also  $L_4(1/4) = 3/4$ , so  $c = 1/4$  is eventually fixed, as is  $c = (2 + \sqrt{3})/4$ . We can check that there are many other eventually fixed points.

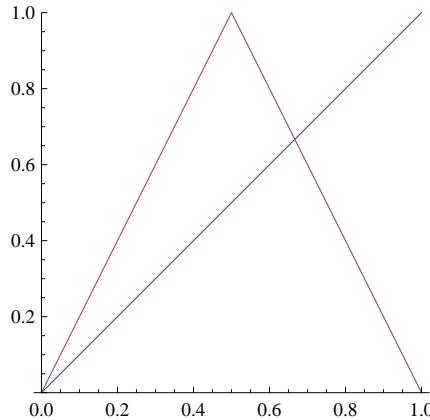
#### Example 1.2.4 The Tent Map.

Define a function  $T : [0, 1] \rightarrow [0, 1]$  by

$$T(x) = 1 - 2|x - 1/2| = \begin{cases} 2x; & 0 \leq x \leq 1/2 \\ 2(1-x); & 1/2 < x \leq 1. \end{cases}$$

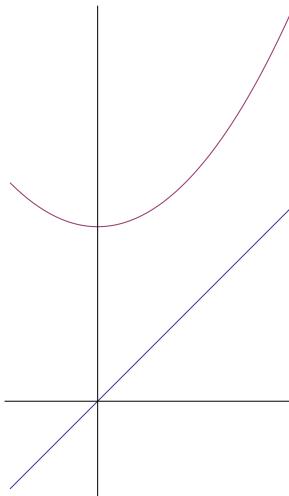
$T$  is called the *tent map*. The fixed points are given by  $T(0) = 0$  and  $T(2/3) = 2/3$ .

Since  $T(1/4) = 1/2$ ,  $T(1/2) = 1$  and  $T(1) = 0$ , we see that  $x = 1/4, 1/2, 1$  are eventually fixed. It is not difficult to see that there are (infinitely) many other eventually fixed points.



The tent map intersects  $y = x$  where  $x = 0$  and  $x = 2/3$ .

**Example 1.2.5** This example shows that some maps do not have fixed points.  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^2 + 1$  has no fixed points, since the equation  $x^2 + 1 = x$  has no real solution.



$f(x) = x^2 + 1$  does not intersect the line  $y = x$ .

**Example 1.2.6** Let  $c \in \mathbb{R}$  and  $f_c : \mathbb{R} \rightarrow \mathbb{R}$  be defined by  $f_c(x) = x^2 + c$ . We ask for which values of  $c$  does  $f_c$  have a fixed point? What are the corresponding values of the fixed point(s)? If we graph  $f_c$  using a computer algebra system and a “manipulate” type plot, we can see that when  $c = 1$  there are no fixed points (Example 1.2.5), but as  $c$  decreases, at some value of  $c$ ,  $f_c$  intersects the line  $y = x$ . The values of  $c$  where this happens may be explicitly determined (see Exercises 1.2 # 1).

Our next result is a type of *fixed point theorem*. The proof requires the use of the Intermediate Value Theorem. Many of our proofs use standard results from calculus, including the Mean Value Theorem, (Rolle's Theorem), the Monotone Sequence Theorem and some completeness properties of the real numbers. Implicit in the proofs of Theorems 1.2.7 and 1.2.9 is the fact that if  $f : [a, b] \rightarrow \mathbb{R}$  is a continuous function, then  $f([a, b])$ , the range of  $f$ , is also an interval of the form  $[c, d]$ . These results are discussed in Appendix A. Fixed point theorems are of particular interest in dynamical systems as the fixed points may give information about the long term behavior of the function.

**Theorem 1.2.7** *Let  $I = [a, b]$ ,  $a < b$ , and suppose  $f : I \rightarrow I$  is a continuous function. Then  $f(x)$  has a fixed point  $c \in I$ .*

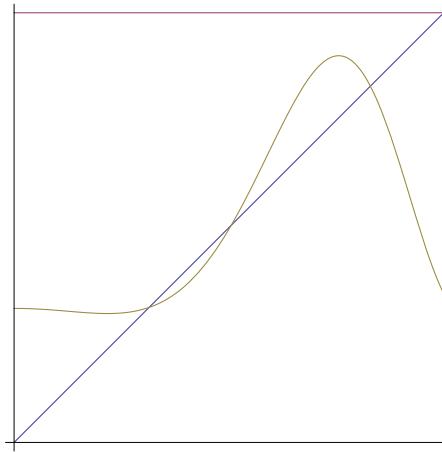
**Proof.** Set  $g(x) = f(x) - x$ , a continuous function. We may assume that  $f(a) \neq a$  and  $f(b) \neq b$ , for otherwise we have nothing to prove, so we must have  $f(a) > a$  and  $f(b) < b$ .

It follows that

$$g(a) = f(a) - a > 0, \quad \text{and} \quad g(b) = f(b) - b < 0.$$

Since  $g(x)$  is positive at  $a$  and negative at  $b$ , the Intermediate Value Theorem ensures the existence of  $c \in (a, b)$  with  $g(c) = 0$ , i.e.,  $f(c) = c$ , or  $c$  is a fixed point of  $f(x)$ .

□

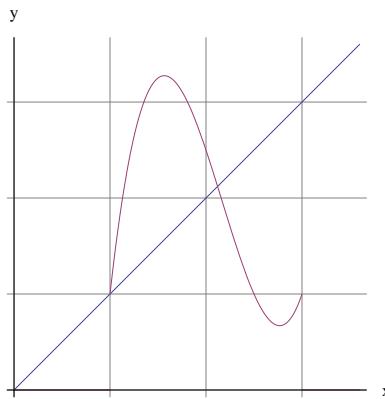


The graph of  $f(x)$  always intersects the line  $y = x$ .

**Remark 1.2.8** Theorem 1.2.7 is an example of an existence theorem. It says nothing about how to find the fixed point(s), where they are, or how many there are. It tells us that if  $f(x)$  is a continuous function on an interval  $I$  with  $f(I) \subseteq I$ , then  $f(x)$  has a fixed point in  $I$ . Another important example of an existence theorem says that if  $f(I) \supseteq I$ , then  $f(x)$  has a fixed point in  $I$ , as we shall now demonstrate:

**Theorem 1.2.9** *Let  $f : I \rightarrow \mathbb{R}$  ( $I = [a, b]$ ,  $a < b$ ), be a continuous function with  $f(I) \supseteq I$ . Then  $f(x)$  has a fixed point in  $I$ .*

**Proof.** As before, set  $g(x) = f(x) - x$ . There exists  $c_1 \in [a, b]$  with  $f(c_1) < c_1$  (in fact  $f(c_1) < a < c_1$ ). Also there is  $c_2 \in [a, b]$  with  $f(c_2) > c_2$  (because  $f([a, b])$  is also a closed and bounded interval).

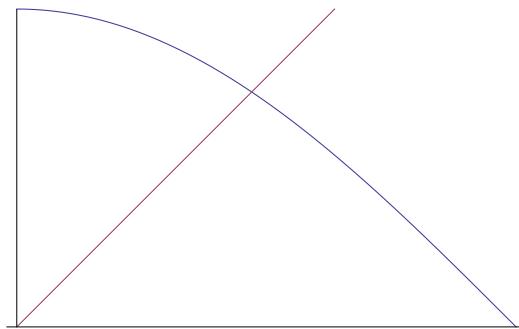


If  $f(I) \supseteq I$ , then  $f$  has a fixed point.

Then  $g(c_1) < 0$  and  $g(c_2) > 0$ , and since  $g(x)$  is a continuous function, it follows by the Intermediate Value Theorem that there exists  $c \in I$ , ( $c_1 < c < c_2$  or  $c_2 < c < c_1$ ), with  $g(c) = 0$ ;  $f(c) = c$ .

□

**Example 1.2.10** It is often difficult to find fixed points explicitly. For example, suppose that  $f(x) = \cos x$ , then we apply the Intermediate Value Theorem to  $g(x) = \cos x - x$ . Since  $g(0) = 1 > 0$  and  $g(\pi/2) = -\pi/2 < 0$ ,  $g$  has a zero  $x = c$ ,  $0 < c < \pi/2$ . Then  $c$  is a fixed point of  $f$ :  $f(c) = c$ , but we have no method for finding an exact value of  $c$ . An approximation to  $c$  can be found by applying Newton's method to  $g(x)$ . Convergence is very rapid and we see that  $c = .739085\dots$ , approximately.



$f(x) = \cos x$  has a fixed point in  $[0, \pi/2]$ .

## Exercises 1.2

1. Give conditions on  $b$  and  $c$  for the map  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^2 + bx + c$  to have a fixed point. Use these conditions to show that  $f_c(x) = x^2 + c$  has a fixed point provided  $c \leq 1/4$ .
  
2. Show that if  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a cubic function, then  $f$  always has a fixed point.
  
3. Find all fixed points and eventual fixed points of the map  $f(x) = 1 - |x|$ . (Hint: Look at the graphs of  $f$  and  $f^2$ ).
  
4. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be such that for some  $n \in \mathbb{Z}^+$ , the  $n$ th iterate of  $f$  has a unique fixed point  $c$  (i.e.,  $f^n(c) = c$  and  $c$  is unique). Show that  $c$  is a fixed point of  $f$ .
  
5. Use a computer algebra system to find how the following functions behave when the given point is iterated (comment on what appears to be happening in each case):
  - (i)  $L_1(x) = x(1 - x)$ , with starting point  $x_0 = .75$ .
  - (ii)  $L_2(x) = 2x(1 - x)$ , with starting point  $x_0 = .1$ .
  - (iii)  $L_3(x) = 3x(1 - x)$ , with starting point  $x_0 = .2$ .
  - (iv)  $L_{3.2}(x) = 3.2x(1 - x)$ , with starting point  $x_0 = .95$ .
  - (v)  $f(x) = \sin(x)$ , with starting point  $x_0 = 9.5$ .
  - (vi)  $g(x) = \cos(x)$ , with starting point  $x_0 = -15.3$ .
  
6. Consider the eventual fixed points of the logistic map  $L_\mu : [0, 1] \rightarrow [0, 1]$ ,  $L_\mu(x) = \mu x(1 - x)$ , for  $0 < \mu < 4$ .
  - (a) Show that there are no eventual fixed points associated with the fixed point  $x = 0$ , other than  $x = 1$ .
  - (b) Show that for  $1 < \mu \leq 2$ , the only eventual fixed point associated with the fixed point  $x = 1 - 1/\mu$  is  $x = 1/\mu$ .

- (c) Show that there are additional eventual fixed points associated with  $x = 1 - 1/\mu$  when  $2 < \mu < 3$ .
- (d) Investigate the eventual fixed points of the logistic map when  $\mu = 5/2$ .

7. (a) Let  $f(x) = (1 + x)^{-1}$ . Find the fixed points of  $f$  and show that there are no points  $c$  with  $f^2(c) = c$  and  $f(c) \neq c$  (period 2-points). Note that  $f(-1)$  is not defined, but points that get mapped to  $-1$  belong to the interval  $[-2, -1]$  and are of the form

$$\nu_n = -\frac{F_{n+1}}{F_n}, \quad n \geq 1,$$

where  $(F_n)$ ,  $n \geq 0$ , is the Fibonacci sequence (see 1.1.4). Note that  $\nu_n \rightarrow -r$  as  $n \rightarrow \infty$ , where  $-r$  is the negative fixed point of  $f$  (see [22] for more details).

(b) If  $x_0 \in (0, 1]$ , set  $x_n = \frac{F_{n-1}x_0 + F_n}{F_nx_0 + F_{n+1}}$ . Use mathematical induction to show that

$$x_{n+1} = f(x_n) = \frac{F_nx_0 + F_{n+1}}{F_{n+1}x_0 + F_{n+2}}.$$

Deduce that as  $n \rightarrow \infty$ ,  $x_n \rightarrow 1/r$ , the positive fixed point of  $f$ .

8. Let  $f : S \rightarrow S$  be a dynamical system, where  $S = \{a_1, a_2, \dots, a_p\}$  is a finite subset of  $\mathbb{R}$ . Show that every point  $x_0 \in S$  is eventually periodic. This exercise shows that dynamical systems on finite sets are not very complicated.

9. Give an example of a one-to-one and onto function  $f : [0, 1] \rightarrow [0, 1]$ , with no fixed point.

### 1.3 Graphical Iteration.

In this section we introduce the notions of *attracting* and *repelling* fixed points from a graphical point of view. The formal definitions are given in the next section.

To understand the behavior of a function  $f(x)$  under iteration, it is useful to follow the iterates at  $x_0$  using a process called *graphical iteration*. The resulting graphs are sometimes called *web diagrams*.

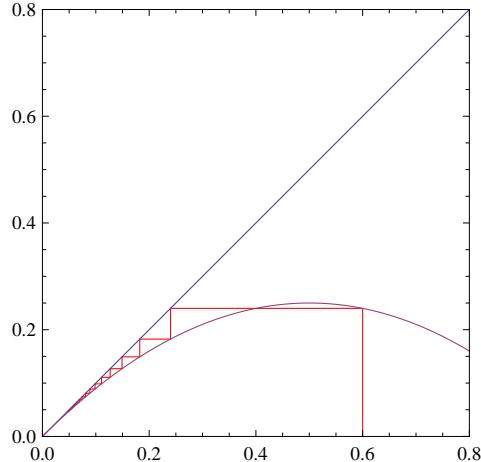
We start at  $x_0$  on the  $x$ -axis and draw a line vertically to the graph of  $f(x)$ . We then move horizontally to the line  $y = x$ , then vertically to the graph, and continue

in this way:

$$(x_0, 0) \rightarrow (x_0, f(x_0)) \rightarrow (f(x_0), f(x_0)) \rightarrow (f(x_0), f^2(x_0)) \rightarrow (f^2(x_0), f^2(x_0)) \rightarrow \dots$$

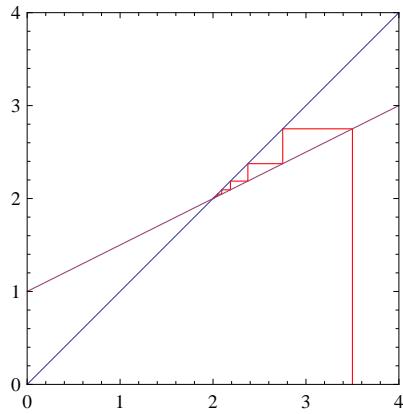
We will see that in some examples the iterations converge to a fixed point. In others  $f^n(x_0) \rightarrow \infty$ , whilst in others still,  $f^n(x_0)$  oscillates between different points, or behaves in a quite unpredictable way.

**Example 1.3.1**  $f(x) = x(1 - x)$ . In this example, an examination of graphical iteration seems to suggest that the orbits of any point in  $[0, 1]$  approach the fixed point  $x = 0$ . The point  $x = 0$  illustrates the notion of *attracting* fixed point, to be defined in Section 1.4.



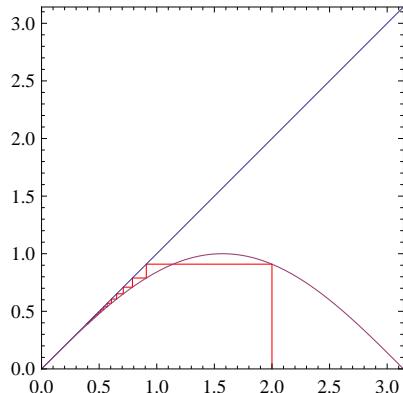
Iterating  $f(x) = x(1 - x)$ , starting with  $x_0 = 0.6$ .

**Example 1.3.2** Let  $f(x) = x/2 + 1$ . This is an affine transformation with  $a = 1/2$ ,  $b = 1$ . According to what we saw in Example 1.1.3, the iterates should converge to  $\frac{b}{1-a} = 2$ , since  $|a| < 1$ . What is actually happening is that  $c = 2$  is an attracting fixed point of  $f(x)$  with the property that it attracts all members of  $\mathbb{R}$  (said to be *globally attracting*, so  $f^n(x) \rightarrow 2$  as  $n \rightarrow \infty$ , for all  $x \in \mathbb{R}$ ).

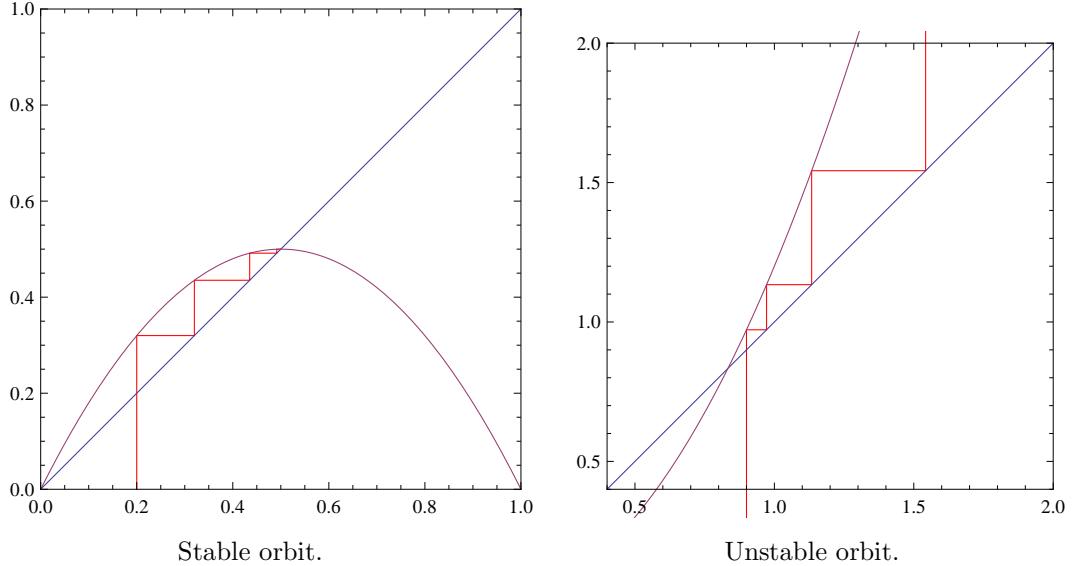


Graphical iteration for  $f(x) = x/2 + 1$ , starting with  $x_0 = 3.5$ .

**Example 1.3.3** We see from the graph of  $f(x) = \sin x$  that  $c = 0$  seems to be an attracting fixed point. We shall show later that this fixed point is actually globally attracting:  $f^n(x) \rightarrow 0$  as  $n \rightarrow \infty$ , for every  $x \in \mathbb{R}$ . A computer algebra system suggests that this is true, but notice that the convergence to 0 is very slow.



$f(x) = \sin x$  with a fixed point at  $x = 0$ .



Two basic situations arise from the iterations near a fixed point (together with some variations of these):

- (i) Stable orbit, where graphical iterations approaches a fixed point (called an *asymptotically stable fixed point*).
- (ii) Unstable orbit, where graphical iterations move away from a fixed point (called a *repelling fixed point*).

In addition, we notice that if the sequence  $x_n = f^n(x_0)$  converges to some point  $c$  as  $n \rightarrow \infty$ , then  $c$  is a fixed point.

**Proposition 1.3.4** *If  $f : I \rightarrow I$  is a continuous function on an interval  $I$  and  $\lim_{n \rightarrow \infty} f^n(x_0) = c \in I$ , then  $f(c) = c$  (i.e., if the orbit converges to a point  $c$ , then  $c$  is a fixed point of  $f$ ) .*

**Proof.** Clearly  $\lim_{n \rightarrow \infty} f^n(x_0) = c \Rightarrow f(\lim_{n \rightarrow \infty} f^n(x_0)) = f(c)$ , and since  $f$  is continuous,

$$\lim_{n \rightarrow \infty} f^{n+1}(x_0) = f(c).$$

We also have  $c = \lim_{n \rightarrow \infty} f^{n+1}(x_0)$ , so  $c = f(c)$  by the uniqueness of limit.  $\square$

### 1.4 The Stability of Fixed Points.

In the last section we gave an intuitive idea about what it means for a fixed point to be attracting or repelling. In order to give criteria for fixed points to be attracting/repelling, we need a rigorous definition of these notions.

**Definition 1.4.1** Let  $I$  be a subinterval of  $\mathbb{R}$ ,  $f : I \rightarrow I$  a function with  $c \in I$  a fixed point:  $f(c) = c$ .

(i)  $c$  is a *stable fixed point* if for all  $\epsilon > 0$ , there exists  $\delta > 0$  such that if  $x \in I$  and  $|x - c| < \delta$ , then  $|f^n(x) - c| < \epsilon$  for all  $n \in \mathbb{Z}^+$ .

If this does not hold,  $c$  will be called *unstable*.

(ii)  $c$  is an *attracting fixed point* if there is a real number  $\eta > 0$  such that if  $x \in I$ , and  $|x - c| < \eta$ , then  $\lim_{n \rightarrow \infty} f^n(x) = c$ .

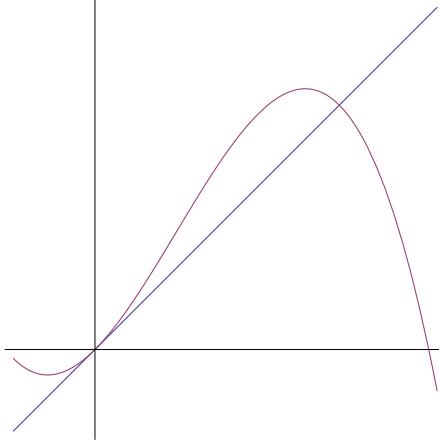
(iii)  $c$  is an *asymptotically stable fixed point* if it is both stable and attracting.

**Remark 1.4.2** The graphical iterations of the Section 1.3 suggest that if a fixed point  $c$  of  $f$  has the property:  $|f'(c)| < 1$ , then  $c$  is an asymptotically stable fixed point. This is essentially what the next theorem tells us. In particular, if  $f'(x)$  is continuous near  $x = c$ , then  $|f'(x)| < 1$  for  $x$  close to  $c$ .

If  $c$  is an unstable fixed point, we can find  $\epsilon > 0$  and  $x$  arbitrarily close to  $c$  such that some iterate of  $x$ , say  $f^n(x)$ , satisfies  $|f^n(x) - c| > \epsilon$ . This happens when  $|f'(c)| > 1$  as the next theorem shows. A fixed point can be stable without being attracting, and it can be attracting without being stable.

**Definition 1.4.3** A fixed point  $c$  of  $f$  is *hyperbolic* if  $|f'(c)| \neq 1$ . If  $|f'(c)| = 1$  it is *non-hyperbolic*. The reasons for these names becomes apparent when one looks at the fixed points of maps  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ .

Thus if  $c$  is a non-hyperbolic fixed point, then  $f'(c) = 1$  or  $f'(c) = -1$ , so the graph of  $f(x)$  either meets the line  $y = x$  tangentially, or at  $90^\circ$ :



$f(x) = -2x^3 + 2x^2 + x$  has both types of non-hyperbolic fixed points.

In the next theorem we show that the stability of hyperbolic fixed points is easy to determine. The proof uses the Mean Value Theorem (see Appendix A):

**Theorem 1.4.4** *Let  $f : I \rightarrow I$  be a differentiable function with a continuous first derivative (we say that  $f$  is of class  $C^1$ ).*

- (i) *If  $c$  is a fixed point of  $f(x)$  with  $|f'(c)| < 1$ , then  $c$  is asymptotically stable. The iterates of points close to  $c$ , converge to  $c$  geometrically (i.e., there is a constant  $0 < \lambda < 1$  for which  $|f^n(x) - c| < \lambda^n|x - c|$  for all  $n \in \mathbb{Z}^+$ , and for all  $x \in I$  sufficiently close to  $c$ ).*
- (ii) *If  $c$  is a fixed point of  $f(x)$  for which  $|f'(c)| > 1$ , then  $c$  is a repelling fixed point of  $f$ .*

**Proof.** (i) We may assume that  $I$  is an open interval with  $c \in I$ . Suppose  $|f'(c)| < \lambda < 1$  for some  $\lambda > 0$ . Using the continuity of  $f'(x)$ , there exists an open interval  $J \subset I$  with  $c \in J$  and  $|f'(x)| < \lambda < 1$  for all  $x \in J$ .

By the Mean Value Theorem, if  $x \in J$  there exists  $a \in J$ , lying between  $x$  and  $c$ , satisfying

$$f'(a) = \frac{f(x) - f(c)}{x - c},$$

so that

$$|f(x) - c| = |f'(a)||x - a| < \lambda|x - c|,$$

(i.e.,  $f(x)$  is closer to  $c$  than  $x$  was).

Repeating this argument with  $f(x)$  replacing  $x$  gives

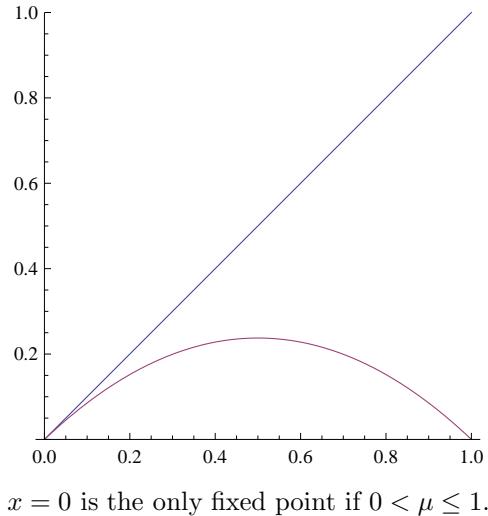
$$|f^2(x) - c| < \lambda^2|x - c|, \dots, |f^n(x) - c| < \lambda^n|x - c|.$$

Since  $\lambda^n \rightarrow 0$  as  $n \rightarrow \infty$ , it follows that  $f^n(x) \rightarrow c$  as  $n \rightarrow \infty$ .

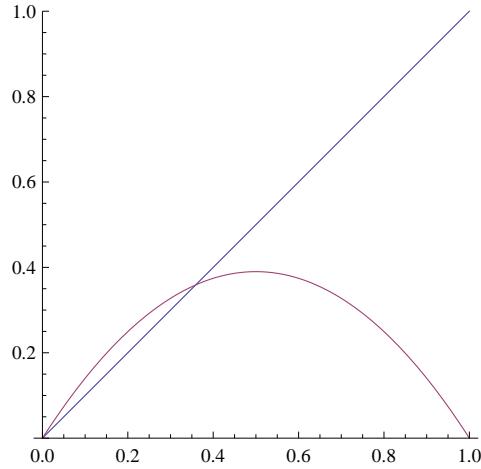
The proof of (ii) is similar. □

**Example 1.4.5** We have seen that the logistic map  $L_\mu(x) = \mu x(1 - x) = \mu x - \mu x^2$  has fixed points  $x = 0$  and  $x = 1 - 1/\mu$ . If  $0 < \mu \leq 1$ , then  $1 - 1/\mu \leq 0$ , so  $c = 0$  is the only fixed point in  $[0, 1]$ .

$L'_\mu(x) = \mu - 2\mu x$ , and  $L'_\mu(0) = \mu$  so 0 is an attracting fixed point when  $0 < \mu < 1$ . When  $\mu = 1$  it is a non-hyperbolic fixed point.



If  $\mu > 1$ , then 0 and  $1 - 1/\mu$  are both fixed points in  $[0, 1]$ , but now 0 is repelling.



For  $1 < \mu < 3$ , there are two fixed points, with  $x = 0$  repelling and  $x = 1 - 1/\mu$  attracting.

Also,

$$L'_\mu(1 - 1/\mu) = \mu - 2\mu(1 - 1/\mu) = 2 - \mu,$$

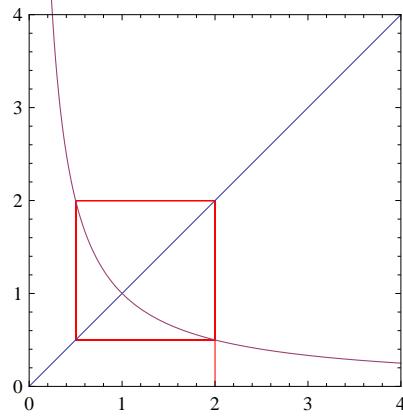
so that

$$|L'_\mu(1 - 1/\mu)| = |2 - \mu| < 1 \text{ if and only if } 1 < \mu < 3.$$

In this case  $x = 1 - 1/\mu$  is an attracting fixed point, and repelling for  $\mu > 3$ . Note that  $L'_1(0) = 1$  and  $L'_3(2/3) = 1$  so we have non-hyperbolic fixed points for these values of  $\mu$ .

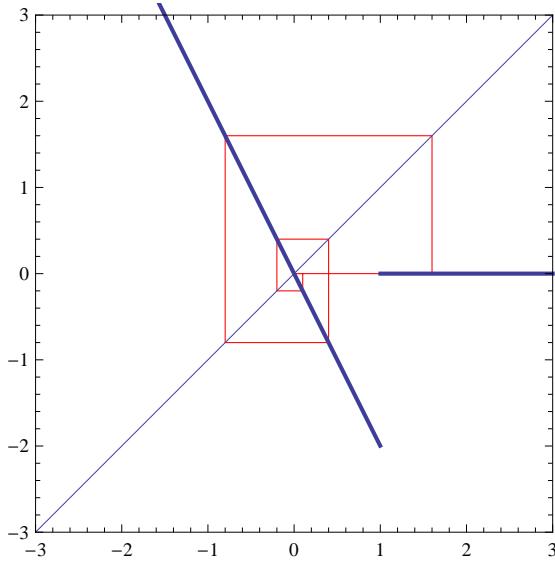
In Section 1.5 we will examine the stability of the non-hyperbolic fixed points  $x = 0$  and  $x = 2/3$ , when  $\mu = 1$  and  $\mu = 3$  respectively.

**Example 1.4.6** Let  $f(x) = 1/x$ . The two fixed points  $x = \pm 1$  are non-hyperbolic since  $f'(x) = -1/x^2$ , so  $f'(\pm 1) = -1$ . However, both fixed points are stable, but not attracting since  $f^2(x) = f(1/x) = x$ , for all  $x \neq 0$ . Points close to  $\pm 1$  move neither closer to  $\pm 1$ , nor further away.



The fixed point  $c = 1$  is neither attracting nor repelling.

**Example 1.4.7** Consider the function  $f_a(x) = \begin{cases} -2x; & x < a \\ 0; & x \geq a \end{cases}$ , where  $a > 0$ . Then  $c = 0$  is an unstable (repelling) fixed point of  $f_a$  which is attracting i.e.,  $\lim_{n \rightarrow \infty} f_a^n(x) = 0$  for all  $x \in \mathbb{R}$ . Recall that such a fixed point is called *globally attracting*. Note that the function  $f_a$  is not continuous at  $x = a$ . It has been shown by Sedaghat [114] that a continuous mapping of the real line cannot have an unstable fixed point that is globally attracting.



Points close to  $x = 0$  initially move away from 0, but are ultimately mapped directly to 0.

#### Example 1.4.8 Newton's Method Revisited.

We examine the question of why Newton's method is so successful at finding the approximate value of zeros of a function  $f(x)$ . Why do the iterates of the Newton function converge so rapidly to a root of  $f$  for most choices of an initial guess? Suppose that  $f : I \rightarrow I$  is a function whose zero  $c$  is to be approximated using Newton's method. The Newton function is  $N_f(x) = x - f(x)/f'(x)$ , where we are assuming  $f'(c) \neq 0$  and that  $f''(x)$  exist in an open interval containing  $c$ . Notice that since  $f(c) = 0$ ,  $N_f(c) = c$ , i.e.,  $c$  is a fixed point of  $N_f$ . Consider  $N'_f(c)$ :

$$N'_f(x) = 1 - \frac{f'(x)f'(x) - f(x)f''(x)}{[f'(x)]^2} = \frac{f(x)f''(x)}{[f'(x)]^2},$$

so that

$$N'_f(c) = \frac{f(c)f''(c)}{[f'(c)]^2} = 0,$$

since  $f(c) = 0$ .

It follows that  $|N'_f(c)| = 0 < 1$ , so that  $c$  is an attracting (asymptotically stable) fixed point for  $N_f$ , and in particular

$$\lim_{n \rightarrow \infty} N_f^n(x_0) = c,$$

provided  $x_0$ , the first approximation to  $c$ , is sufficiently close to  $c$ .

**Definition 1.4.9** A fixed point  $c$  of  $f(x)$  is said to be *super-attracting* if  $f'(c) = 0$ . This gives a very fast convergence to the fixed-point for nearby points.

**Remark 1.4.10** Suppose that  $f'(c) = 0$ , then  $N_f(x)$  is not defined at  $x = c$ , since the quotient  $f(x)/f'(x)$  is not defined there. If  $f(x)$  can be written as  $f(x) = (x - c)^k h(x)$  where  $h(c) \neq 0$  and  $k \in \mathbb{Z}^+$  (for example if  $f(x)$  is a polynomial with a multiple root), then we have

$$\frac{f(x)}{f'(x)} = \frac{(x - c)^k h(x)}{k(x - c)^{k-1} h(x) + (x - c)^k h'(x)} = \frac{(x - c)h(x)}{kh(x) + (x - c)h'(x)} = 0,$$

when  $x = c$ , so that  $N_f(x)$  has a removable discontinuity at  $x = c$  (removed by setting  $N_f(c) = c$ ). Now find  $N'_f(x)$ . Then we can show that  $N'_f(x)$  has a removable discontinuity at  $x = c$ , which can be removed by setting  $N'_f(c) = (k - 1)/k$ , giving  $|N'_f(c)| < 1$  (see Exercises 1.4 # 9).

We summarize the above in the next theorem. The requirement that  $f$  be a polynomial may be weakened, to give a more general result.

**Theorem 1.4.11** Let  $f : I \rightarrow I$  be a polynomial function and  $I$  an interval containing  $c$ , a zero of  $f(x)$ . Then  $x = c$  is a super-attracting fixed point of the Newton function  $N_f$ , if and only if  $f'(c) \neq 0$ .

**Proof.** If  $f'(c) \neq 0$ , then  $N'_f(c) = f(c)f''(c)/[f'(c)]^2 = 0$ , since  $f(c) = 0$ .

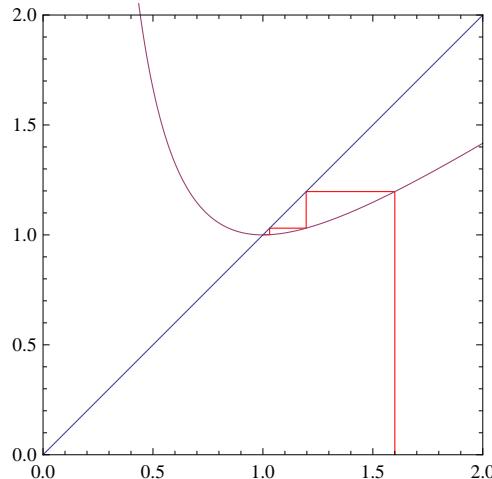
Conversely, suppose that  $f(c) = 0$  and  $f'(c) = 0$ . We can write  $f(x) = (x - c)^k h(x)$ , where  $h(c) \neq 0$ , and  $k > 1$ . The discussion in Remark 1.4.10 shows that  $N_f(c) = c$  and  $N'_f(c) = (k - 1)/k \neq 0$ , contradicting  $c$  being a super-attracting fixed point. It follows that  $f'(c) \neq 0$ .

□

**Example 1.4.12** Suppose  $f(x) = x^3 - 1$ , then  $f(1) = 0$  and

$$N_f(x) = x - f(x)/f'(x) = x - \left( \frac{x^3 - 1}{3x^2} \right) = \frac{2x}{3} + \frac{1}{3x^2},$$

so  $N_f(1) = 1$  and  $N'_f(x) = \frac{2}{3} - \frac{2}{3x^3}$ , giving  $N'_f(1) = 0$ . We observe from graphical iterations, that since the graph is very flat when  $x = 1$ , we have very fast convergence to the fixed point of  $N_f$ . This is why Newton's method is such a good algorithm for finding zeros of functions.



Very fast convergence to the fixed point.

### Exercises 1.4

- Find the fixed points and determine their stability for the function  $f(x) = \frac{6}{x} - 1$ .

2. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ . If  $f'(x)$  exists with  $f'(x) \neq 1$  for all  $x \in \mathbb{R}$ , prove that  $f$  has at most one fixed point (Hint: Use the Mean Value Theorem).
3. For the family of quadratic maps  $f_c(x) = x^2 + c$ ,  $x \in [0, 1]$ , use a computer algebra system to give graphical iteration (web plots) for the values shown (use 20 iterations):
- (i)  $c = 1/2$ , starting point  $x_0 = 1$ ,
  - (ii)  $c = 1/4$ , starting point  $x_0 = .1$ ,
  - (iii)  $c = 1/8$ , starting point  $x_0 = .7$ .
4. Let  $S_\mu(x) = \mu \sin(x)$ ,  $0 \leq x \leq 2\pi$ ,  $0 < \mu \leq \pi$  and  $C_\mu(x) = \mu \cos(x)$ ,  $-\pi \leq x \leq \pi$  and  $-\pi \leq \mu \leq \pi$ ,  $\mu \neq 0$ .
- (a) Show that  $S_\mu$  has a super-attracting fixed point at  $x = \pi/2$ , when  $\mu = \pi/2$ .
  - (b) Find the corresponding values for  $C_\mu$  having a super-attracting fixed point.
5. Show that the map  $f(x) = \frac{2}{x+1}$  has no periodic points of period  $n > 1$ . (Hint: Use the closed formula in Exercise 1.1.6).
6. Show that if  $f(x) = x + 1/x$  and  $x > 0$ , then  $f^n(x) \rightarrow \infty$  as  $n \rightarrow \infty$ . (Hint:  $f$  has no fixed points in  $\mathbb{R}$  and it is continuous with  $0 < f(x) < f^2(x) < \dots$ , so suppose the limit exists and obtain a contradiction).
7. Let  $N_f$  be the Newton function of the map  $f(x) = x^2 + 1$ . Clearly there are no fixed points of the Newton function as there are no zeros of  $f$ . Show that there are points  $c$  where  $N_f^2(c) = c$  (called *period 2-points* of  $N_f$ ).
8. (a) Suppose that  $f(c) = f'(c) = 0$  and  $f''(c) \neq 0$ . If  $f''(x)$  is continuous at  $x = c$ , show that the Newton function  $N_f(x)$  has a removable discontinuity at  $x = c$  (Hint: Apply L'Hopital's rule to  $N_f$  at  $x = c$ ).

(b) If in addition,  $f'''(x)$  is continuous at  $x = c$  with  $f'''(c) \neq 0$ , show that  $N'_f(c) = 1/2$ , so that  $x = c$  is not a super-attracting fixed point in this case.

(c) Check the above for the function  $f(x) = x^3 - x^2$  with  $c = 0$ .

9. Continue the argument in Remark 1.4.10 (generalizing the last exercise), to show that the derivative of the Newton function  $N_f$  has a removable discontinuity at  $x = c$ , which can be removed by setting  $N'_f(c) = (k - 1)/k$ .

10. Let  $f$  be a twice differentiable function with  $f(c) = 0$ . Show that if we find the Newton function of  $g(x) = f(x)/f'(x)$ , then  $x = c$  will be a super-attracting fixed point for  $N_g$ , even if  $f'(c) = 0$  (this is called *Halley's method*).

11. Let  $f(x) = \begin{cases} 4x; & 0 \leq x < 1/4 \\ 2 - 4x; & 1/4 \leq x < 1/2 \\ 0; & 1/2 \leq x \leq 1 \end{cases}$ . Note that  $f$  is a continuous function.

Use graphical analysis to show that  $x = 0$  is a repelling fixed point, but  $f^n(x) \rightarrow 0$  as  $n \rightarrow \infty$ , for all  $x \in [1/2, 1]$ . Is  $x = 0$  globally attracting? Use graphical analysis to determine the set  $\{x \in [0, 1] : f^n(x) \rightarrow 0, \text{ as } n \rightarrow \infty\}$ .

12. Write Definition 1.4.1 (of a stable fixed point) as a quantified statement (using  $\forall$  and  $\exists$  and other logical symbols). Deduce the negation of the statement.

13. Let  $f : \mathbb{R} \rightarrow (0, \infty)$  be differentiable everywhere with  $f'(x) < 0$  for all  $x \in \mathbb{R}$ . Show that the Newton function  $N_f$  has no fixed points or periodic points.

## 1.5 Non-hyperbolic Fixed Points.

We have established criteria for the stability of hyperbolic fixed points. In this section we give some criteria for the stability of non-hyperbolic fixed points. Our first theorem deals with the case of a fixed point  $c$  for  $f$  with  $f'(c) = 1$ . We end the chapter by considering what happens when  $f'(c) = -1$ . Let  $I$  be a subinterval of  $\mathbb{R}$  and  $c \in I$ . If  $f : I \rightarrow I$  has a continuous first derivative and  $f(c) = c$ , where  $|f'(c)| < 1$ , then we have seen that  $c$  is an asymptotically stable fixed point of  $f$ . If  $f(x) = \sin x$ ,

then  $f(0) = 0$ , and  $|f'(0)| = 1$ , a non-hyperbolic fixed point, so stability is unclear. However, graphical iteration suggests that the basin of attraction of  $f$  is all of  $\mathbb{R}$ , or  $c = 0$  is an asymptotically stable fixed point. Before considering non-hyperbolic fixed points in detail, let us prove the last statement analytically:

**Proposition 1.5.1** *The fixed point  $c = 0$  of  $f(x) = \sin x$  is globally attracting and stable.*

**Proof.** Notice that  $c = 0$  is the only fixed point of  $\sin x$ . To see this, if  $\sin c = c$  then we cannot have  $|c| > 1$  since  $|\sin x| \leq 1$  for all  $x$ . If  $0 < c \leq 1$ , the Mean Value Theorem implies there exists  $a \in (0, c)$  with

$$f'(a) = \frac{f(c) - f(0)}{c - 0} = \frac{\sin c}{c}.$$

Since  $0 < \cos a < 1$  for  $a \in (0, 1)$ , we have  $\sin c = c \cos a < c$ . In a similar way we see that there is no fixed point  $c$  with  $-1 \leq c < 0$ .

To show that  $c = 0$  is a globally attracting fixed point of  $f(x)$ , let  $x \in \mathbb{R}$ , where we may assume that  $-1 \leq x \leq 1$ , since this will be the case after the first iteration.

First, suppose that  $0 < x \leq 1$ , then note that  $0 < f'(x) < 1$  on this interval. By the Mean Value Theorem, there exists  $a \in (0, x)$  with

$$f'(a) = \frac{f(x) - f(0)}{x - 0}, \quad \text{or} \quad 0 < f(x) = f'(a)x < x.$$

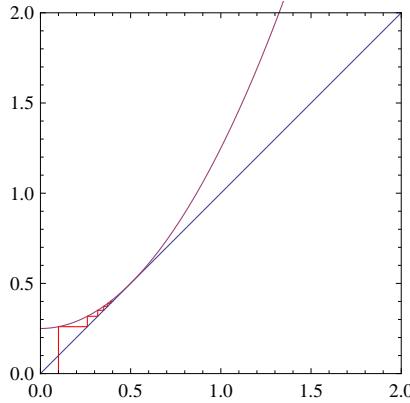
Continuing like this way we obtain  $0 < f^2(x) < f(x)$ , and then

$$0 < f^n(x) < f^{n-1}(x) < \dots < f(x) < x,$$

so we have a decreasing sequence  $x_n = f^n(x)$  bounded below by 0. It follows that this sequence converges, so it must converge to a fixed point (by Proposition 1.3.4). As  $c = 0$  is the only fixed point,  $f^n(x) \rightarrow 0$  as  $n \rightarrow \infty$ . A similar argument can be used if  $-1 \leq x < 0$ .

□

**Example 1.5.2** It is possible for the fixed point to be unstable with a one-sided stability (called *semi-stable*). For example, consider  $f(x) = x^2 + 1/4$  which has the single (non-hyperbolic) fixed point  $c = 1/2$ . This fixed point is stable on the left, but unstable on the right.



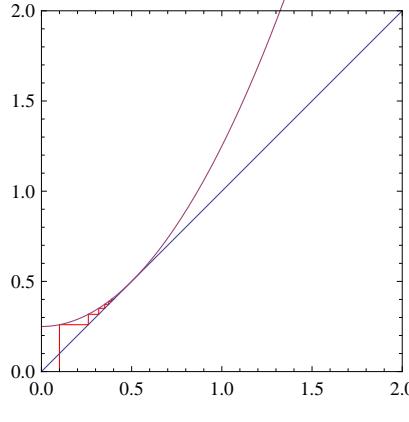
If  $f(x) = x^2 + 1/4$ , then  $f^n(x_0) \rightarrow 1/2$  for  $x_0 < 1/2$ , and  $f^n(x_0) \rightarrow \infty$  for  $x_0 > 1/2$ .

Stability and asymptotic stability on the right and left of a fixed point may be defined in the obvious way. Our next theorem gives some criteria for non-hyperbolic fixed points  $x = c$  of the type where  $f'(c) = 1$ , to be asymptotically stable/unstable, and some criteria for semi-stability. In Theorem 1.5.7, we treat the case where  $f'(c) = -1$ .

**Theorem 1.5.3** *Let  $c$  be a non-hyperbolic fixed point of  $f(x)$  with  $f'(c) = 1$ . If  $f'(x)$ ,  $f''(x)$  and  $f'''(x)$  are continuous at  $x = c$ , then:*

- (i) *If  $f''(c) \neq 0$ ,  $c$  is semi-stable. More specifically:*
  - (a) *if  $f''(c) > 0$ , we have one-sided asymptotic stability on the left of  $c$ ,*
  - (b) *if  $f''(c) < 0$ , we have one-sided asymptotic stability on the right of  $c$ .*
- (ii) *If  $f''(c) = 0$  and  $f'''(c) > 0$ ,  $c$  is unstable.*
- (iii) *If  $f''(c) = 0$  and  $f'''(c) < 0$ ,  $c$  is asymptotically stable.*

**Proof.** (i)(a) If  $f'(c) = 1$  then  $f(x)$  is tangential to  $y = x$  at  $x = c$ . Suppose that  $f''(c) > 0$ . Then  $f(x)$  is concave up at  $x = c$  and the graph of  $f(x)$  must look like the following:



$f(x)$  is concave up near  $x = c$ .

The graph suggests stability on the left and instability on the right, and this is what we now show.

Since the various derivatives are continuous, and  $f''(c) > 0$ , we must have  $f''(x) > 0$  in some small interval  $(c - \delta, c + \delta)$  surrounding  $c$ . In particular, the derivative function  $f'(x)$  must be increasing on that interval, so that since  $f'(c) = 1$ ,

$$f'(x) < 1 \quad \text{for all } x \in (c - \delta, c), \quad \text{and} \quad f'(x) > 1 \quad \text{for all } x \in (c, c + \delta).$$

Also, from the continuity of  $f'(x)$ , we may assume that  $f'(x) > 0$  on  $(c - \delta, c + \delta)$ .

By the Mean Value Theorem applied to  $f(x)$  on the interval  $[x, c] \subset (c - \delta, c)$ , there exists  $q \in (x, c)$  with

$$f'(q) = \frac{f(c) - f(x)}{c - x}.$$

Since  $0 < f'(q) < 1$ ,  $f(c) = c$  and  $c > x$ , we have

$$0 < \frac{f(c) - f(x)}{c - x} < 1,$$

or

$$x < f(x) < c.$$

Repeating this argument gives  $f(x) < f^2(x) < c$ , and more generally we see that the sequence  $f^n(x)$  is increasing and bounded above by  $c$ , so must converge to a fixed point. There can be no other fixed point in this interval, for if there were another, say  $d \neq c$ , the Mean Value Theorem gives  $f'(q_1) = 1$  for some  $q_1 \in (x, c)$ , a contradiction. Consequently,  $f^n(x)$  converges to  $c$ , and so  $c$  is stable on the left.

On the other hand, if  $[c, x] \subset [c, c + \delta]$ , then applying the Mean Value Theorem gives some  $q \in (c, x)$  with

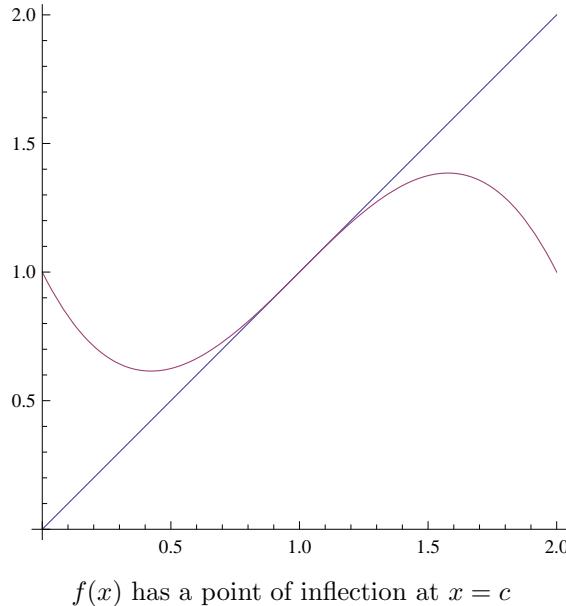
$$f'(q) = \frac{f(x) - f(c)}{x - c} > 1, \quad \text{so} \quad f(x) > x > c,$$

since  $x - c > 0$ . The point  $x$  moves away from  $c$  under iteration, so the fixed point is unstable on the right.

(i)(b) If  $f''(c) < 0$ , then the graph of  $f(x)$  is concave down at  $x = c$ , and we use an argument similar to that in (i)(a).

The proof of (ii) is similar to that of (iii), and so is omitted.

(iii)  $f'''(c) < 0$ ,  $f''(c) = 0$  and  $f'(c) = 1$ . We will show that we have a point of inflection at  $x = c$  as in the following graph, clearly suggesting that  $c$  is an asymptotically stable fixed point:



By the second derivative test,  $f'(x)$  has a local maximum at  $x = c$  (the continuous function  $f'(x)$  is concave down). It follows that

$$f'(x) < 1 \quad \text{for all } x \in (c - \delta, c + \delta), x \neq c,$$

for some  $\delta > 0$ . Alternatively, since  $f'''(x)$  is continuous near  $x = c$  and  $f'''(c) < 0$ , we deduce  $f''(x) > 0$  for  $x \in (c - \delta, c)$ , and  $f'(x)$  is increasing on  $(c - \delta, c)$ , and

$f''(x) < 0$  for  $x \in (c, c + \delta)$ , and  $f'(x)$  is decreasing on  $(c, c + \delta)$ , giving  $f'(x) < 1$  for  $x \in (c - \delta, c + \delta)$ ,  $x \neq c$ .

We now use an argument similar to that in (i)(a): Let  $x \in (c, c + \delta)$ , then there exists  $a \in (c, x)$  with

$$f'(a) = \frac{f(x) - f(c)}{x - c} < 1,$$

so that  $f(x) < x$ . Continue in this way to obtain a decreasing sequence  $(f^n(x))$ , bounded below. Similarly if  $x \in (c - \delta, c)$  we get  $x < f(x)$  to obtain an increasing sequence bounded above.

□

**Examples 1.5.4** 1. Returning to the function  $f(x) = \sin x$ , we see that  $f'(0) = 1$ ,  $f''(0) = 0$  and  $f'''(0) = -1$ , so the conditions of Theorem 1.5.3 (iii) are satisfied and we conclude that  $x = 0$  is an asymptotically stable fixed point.

2. If  $f(x) = \tan x$ , then  $f'(x) = \sec^2 x$ , so  $f'(0) = 1$ ,  $f''(x) = 2\sec^2 x \tan x$  and  $f''(0) = 0$ .  $f'''(x) = 4\sec^2 x \tan^2 x + 2\sec^4 x$ ,  $f'''(0) = 2 > 0$ . Thus Theorem 1.5.3 (ii) applies, and the fixed point  $x = 0$  is unstable.
3. For  $f(x) = x^2 + 1/4$ , with  $f(1/2) = 1/2$ ,  $f'(1/2) = 1$ , we can apply Theorem 1.5.3 (i)(a) to see that we have stability on the left, but not on the right.

**The Schwarzian Derivative 1.5.5** How do we treat the case where  $f'(c) = -1$  at the fixed point? Here the notion of *negative Schwarzian derivative*  $Sf(x)$  plays a role:

**Definition 1.5.6** The *Schwarzian derivative*  $Sf(x)$  of  $f(x)$  is the function

$$Sf(x) = \frac{f'''(x)}{f'(x)} - \frac{3}{2} \left[ \frac{f''(x)}{f'(x)} \right]^2.$$

$Sf(x)$  is defined when  $f'''(x)$  exists and  $f'(x) \neq 0$ . If  $f'(x) = -1$ , then

$$Sf(x) = -f'''(x) - \frac{3}{2}[f''(x)]^2.$$

The next theorem gives some criteria for the stability of non-hyperbolic fixed points  $x = c$  when  $f'(c) = -1$ . The idea is to show that the conditions of Theorem 1.5.3 apply to the function  $g(x) = f^2(x)$ , where we see that  $Sf(x)$  arises naturally as the third derivative of  $g$ . The Schwarzian derivative was first discovered by Joseph

Lagrange in 1781, but was named in honor of Hermann Schwartz by Cayley (see [95]).

**Theorem 1.5.7** Suppose that  $c$  is a fixed point of  $f(x)$  and  $f'(c) = -1$ . If  $f'(x)$ ,  $f''(x)$  and  $f'''(x)$  are continuous at  $x = c$  then:

- (i) If  $Sf(c) < 0$ ,  $x = c$  is an asymptotically stable fixed point.
- (ii) If  $Sf(c) > 0$ ,  $x = c$  is an unstable fixed point.

**Proof.** (i) Set  $g(x) = f^2(x)$ , then  $g(c) = c$ . We see that if  $c$  is asymptotically stable with respect to  $g$ , then it is asymptotically stable with respect to  $f$ . Now

$$g'(x) = \frac{d}{dx}(f(f(x))) = f'(f(x)) \cdot f'(x),$$

so that  $g'(c) = f'(c) \cdot f'(c) = (-1)(-1) = 1$ .

Let us apply Theorem 1.5.3 (iii) to the function  $g(x)$ . Now

$$g''(x) = f'(f(x)) \cdot f''(x) + f''(f(x)) \cdot [f'(x)]^2,$$

thus

$$g''(c) = f'(c)f''(c) + f''(c)[f'(c)]^2 = 0, \quad \text{since } f'(c) = -1.$$

Also,

$$g'''(x) = f''(f(x)) \cdot f'(x) \cdot f''(x) + f'(f(x)) \cdot f'''(x) + f'''(f(x))[f'(x)]^3 + f''(f(x)) \cdot 2f'(x)f''(x).$$

Therefore

$$\begin{aligned} g'''(c) &= [f''(c)]^2(-1) - f'''(c) - f'''(c) + 2f''(c)(-1)f''(c) \\ &= -2f'''(c) - 3[f''(c)]^2 \\ &= 2Sf(c) < 0, \end{aligned}$$

and the result follows from Theorem 1.5.3 (iii).

- (ii) Follows now from Theorem 1.5.3 (ii). □

**Remark 1.5.8** The above proof shows how the Schwarzian derivative arises as the derivative of  $g = f \circ f = f^2$ . In the case where  $f'(c) = -1$ , it follows that

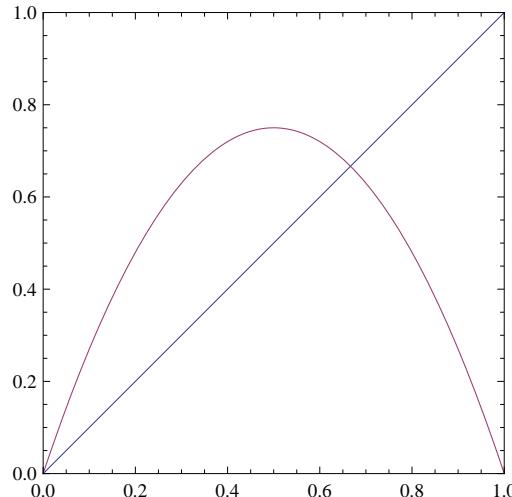
$$g''(c) = 0 \quad \text{and} \quad Sf(c) = \frac{1}{2}g'''(c).$$

**Example 1.5.9** For the logistic map  $L_\mu(x) = \mu x(1 - x)$  we have

$$L'_\mu(x) = \mu - 2\mu x, \quad L''_\mu(x) = -2\mu \quad \text{and} \quad L'''_\mu(x) = 0.$$

When  $\mu = 1$ ,  $x = 0$  is the only fixed point and Theorem 1.5.3 (i)(b) shows that  $x = 0$  is asymptotically semi-stable (attracting on the right). However, we regard this as a stable fixed point of  $L_\mu : [0, 1] \rightarrow [0, 1]$ , since points to the left of 0 are not in the domain of  $L_\mu$ .

When  $\mu = 3$ ,  $c = 2/3$  is fixed and  $L'_\mu(2/3) = -1$ , giving a non-hyperbolic fixed point. However,  $Sf(2/3) = 0 - \frac{3}{2}[6]^2 < 0$  (negative Schwarzian derivative), so by Theorem 1.5.7 (i),  $x = 2/3$  is asymptotically stable.



When  $\mu = 3$ ,  $L_3(2/3) = -1$  and  $c = 2/3$  is an asymptotically stable fixed point. As  $\mu$  increases beyond 3, the slope at the fixed point  $c$  becomes greater than 1 (in absolute value), and the fixed point becomes unstable.

### Exercises 1.5

1. Find the fixed points of the following maps and use the appropriate theorems to determine whether they are asymptotically stable, semi-stable or unstable:

$$(i) f(x) = \frac{x^3}{2} + \frac{x}{2}, \quad (ii) f(x) = \arctan x, \quad (iii) f(x) = x^3 + x^2 + x,$$

$$(iv) f(x) = x^3 - x^2 + x, \quad (v) f(x) = \begin{cases} 3x/4; & x \leq 1/2 \\ 3(1-x)/4; & x > 1/2 \end{cases}$$

2. Consider the family of quadratic maps  $f_c(x) = x^2 + c$ ,  $x \in \mathbb{R}$ .
- Use the theorems of Section 1.5 to determine the stability of the hyperbolic fixed points, for all possible values of  $c$ .
  - Find any values of  $c$  so that  $f_c$  has a non-hyperbolic fixed point, and determine the stability of these fixed points.
3. (a) Show that  $f(x) = -2x^3 + 2x^2 + x$  has two non-hyperbolic fixed points and determine their stability.
- (b) If  $x = 0$  and  $x = 1$  are non-hyperbolic fixed points for  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = ax^3 + bx^2 + cx + d$ , find all possible values of  $a, b, c$  and  $d$ .
- (c) Write down the function  $f(x)$  in each case for (b) above, and determine the stability of the fixed points.
4. Let  $f(x) = x + \mu \cos x$ , where  $\mu > 0$ .
- Show that  $f$  has infinitely many fixed points.
  - Find the range of  $\mu > 0$  which give rise to stable fixed points.
  - Determine the nature of the fixed points when  $\mu = 2$ .
5. Let  $f(x) = x^2 + ax + b$ , and let  $p$  be a fixed point of  $f$ .
- Give conditions on  $a$  and  $b$  so that  $f'(p) = 1$ .
  - Show that if  $f'(p) = 1$ , then  $p$  is semistable.
  - Show that if  $f'(p) = -1$ , then  $p$  is asymptotically stable.
6. Find the Schwarzian derivative of  $f(x) = e^x$ , and  $g(x) = \sin(x)$ , and show that they are always negative.

7. If  $f(x) = \frac{ax+b}{cx+d}$ ,  $a, b, c, d \in \mathbb{R}$ , then  $f$  is called a *linear fractional transformation*. Show that its Schwarzian derivative is  $Sf(x) = 0$  for all  $x$  in its domain.

8. If  $f : \mathbb{R} \rightarrow \mathbb{R}$  is defined by  $f(x) = \begin{cases} x \sin(1/x); & x \neq 0 \\ 0; & x = 0 \end{cases}$ .

- (a) Find the fixed points of  $f$  and show that for  $x \neq 0$  they are non-hyperbolic.
- (b) Show that  $x = 0$  is not an *isolated fixed point* (i.e., there are other fixed points arbitrarily close to 0). Is  $x = 0$  a stable, attracting or repelling fixed point? (Note that  $f'(0)$  is not defined).
  
- 9. Let  $f(x)$  be a polynomial with  $f(c) = c$ . (Recall that a polynomial  $p(x)$  has  $(x - c)^2$  as a factor if and only if both  $p(c) = 0$  and  $p'(c) = 0$ .)
- (i) If  $f'(c) = 1$ , show that  $(x - c)^2$  is a factor of  $g(x) = f(x) - x$ .
- (ii) If  $|f'(c)| = 1$ , show that  $(x - c)^2$  is a factor of  $h(x) = f^2(x) - x$  (i.e., if  $f(x)$  has a non-hyperbolic fixed point  $c$ , then  $c$  is a repeated root of  $f^2(x) - x$ ).
- (iii) Show in the case that  $f'(c) = -1$ , we actually have that  $(x - c)^3$  is a factor of  $h(x) = f^2(x) - x$ .
- (iv) Check that (iii) holds for the non-hyperbolic fixed point  $x = 2/3$ , of the logistic map  $L_3(x) = 3x(1 - x)$ .
- (v) Check that (i), (ii) and (iii) hold for the (non-hyperbolic) fixed points of the polynomial  $f(x) = -2x^3 + 2x^2 + x$ .

10. Consider the converse of the statements in the last exercise. Let  $x = c$  be a fixed point of a polynomial  $f(x)$ :

- (i) If  $(x - c)^2$  is a factor of  $f^2(x) - x$ , show that  $x = c$  is a hyperbolic fixed point of  $f$ .
- (ii) If in addition,  $(x - c)^3$  is not a factor of  $f^2(x) - x$ , show that  $f'(c) = 1$ .
- (iii) If  $(x - c)^3$  is a factor of  $f^2(x) - x$  and  $f''(c) \neq 0$ , show that  $f'(c) = -1$ .

11. (a) Use the Intermediate Value Theorem to show that  $f(x) = \cos(x)$  has a fixed point  $c$  in the interval  $[0, \pi/2]$ . We can show experimentally that this fixed point is approximately  $c = .739085\dots$ , for example by iterating any  $x_0 \in \mathbb{R}$ .
- (b) Show that the basin of attraction of  $c$  is all of  $\mathbb{R}$ . (Hint: You may assume that  $x \in [-1, 1]$  - why? Now use the Mean Value Theorem to show that  $|f(x)-c| < \lambda|x-c|$  for some  $0 < \lambda < 1$ ).
- (c) Does  $f(x)$  have any eventually fixed points?
- (d) Can  $f(x)$  have any points  $p$  with  $f^2(p) = p$  other than  $c$ ?
12. (a) Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a function having a continuous first derivative. If  $x_0 < x_1$  are two fixed points with the properties  $f'(x_0) < 1$  and  $f'(x_1) < 1$ , show that  $f$  must have another fixed point lying between  $x_0$  and  $x_1$ . Note that a similar result will hold if both  $f'(x_0) > 1$  and  $f'(x_1) > 1$ .
- (b) If this fixed point  $x = c$  is unique, show that  $f'(c) \geq 1$ . Can we have  $f'(c) = 1$ ?
- (c) Give an example of a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  with fixed points at  $x_0 = 0$  and  $x_1 = 1$  illustrating (a) above.

## CHAPTER 2

### Bifurcations and the Logistic Family.

#### 2.1 The Basin of Attraction.

In 1976, the population biologist Robert M. May wrote a review article on simple mathematical models with very complicated dynamics [87]. His article gave a systematic account of what was then known about the dynamics of the logistic maps and their corresponding difference equations, and it also posed some open questions. His introduction ended with the following paragraph: “The review ends with an evangelical plea for the introduction of these difference equations into elementary mathematics courses, so that students’ intuition may be enriched by seeing the wild things that simple nonlinear equations can do.”

In this chapter we look in detail at many of the properties of the logistic map that arose in his article. We start with an examination of the basins of attraction of fixed points of these maps, i.e., for a given  $\mu$  and fixed point  $x = c$ , we look for the set of those  $x \in [0, 1]$  which converge to  $c$  under iteration by  $L_\mu(x) = \mu x(1 - x)$ . These maps are used to illustrate many of the ideas that we met in Chapter 1, and are used throughout the remainder of the text to motivate and test new concepts. The study of how the nature of the dynamics of such maps changes, as a parameter changes, is called *bifurcation theory*. something into two branches or parts, originating from the medieval Latin word *bifurcus*, meaning two-pronged”. Of interest throughout this text, is how the dynamical behavior of families of maps such as  $L_\mu$  changes, as  $\mu$  increases through a range of values.

**Definition 2.1.1** Let  $f : I \rightarrow I$  be a function defined on an interval  $I$ . The *basin of attraction*  $B_f(c)$ , of a fixed point  $c$  of  $f(x)$ , is the set of all  $x \in I$  for which the sequence  $x_n = f^n(x)$  converges to  $c$ :

$$B_f(c) = \{x \in I : f^n(x) \rightarrow c, \text{ as } n \rightarrow \infty\}.$$

The *immediate basin of attraction* of  $f$  is the largest interval containing  $c$ , contained in the basin of attraction of  $c$ . We first show that this is an interval which is open as a subset of  $I$ , when  $c$  is an attracting fixed point. If  $I$  is a closed or half-open interval

having end points  $a$  and  $b$ ,  $a < b$ , we regard a subinterval  $J \subseteq I$  as being *open as a subset of  $I$*  if it is either  $I$  itself, an open interval, or an interval of the form  $[a, d)$  or  $(d, b]$  for some  $a < d < b$ .

**Proposition 2.1.2** *Let  $f : I \rightarrow I$  be a continuous function defined on a subinterval of  $\mathbb{R}$  having an attracting fixed point  $c$ . The immediate basin of attraction of  $c$  is an interval which is open as a subset of  $I$ .*

**Proof.** First suppose that  $I$  is an open interval. Since  $c$  is an attracting fixed point there is an  $\epsilon > 0$  such that for all  $x \in I_\epsilon = (c - \epsilon, c + \epsilon)$ ,  $f^n(x) \rightarrow c$  as  $n \rightarrow \infty$ . Denote by  $J$  the largest interval containing  $c$  for which  $f^n(x) \rightarrow c$  for  $x \in J$ , as  $n \rightarrow \infty$ .

Suppose that  $J = [a, b]$ , a closed interval, then there exists  $r \in \mathbb{Z}^+$  with  $f^r(a) \in I_\epsilon$ . Now  $f^r$  is also a continuous function, so points close to  $a$  will also get mapped into  $I_\epsilon$ , leading to a contradiction.

More precisely, there exists  $\delta > 0$  such that if  $|x - a| < \delta$ , then  $|f^r(x) - f^r(a)| < \eta$ , where  $\eta = \min\{|f^r(a) - (c - \epsilon)|, |(c + \epsilon) - f^r(a)|\}$ . Thus there are points  $x < a$  close to  $a$  for which  $f^r(x) \in I_\epsilon$ , so  $f^{rn}(x) \rightarrow c$  as  $n \rightarrow \infty$ , a contradiction. We conclude that  $a \notin J$  (and similarly for  $b$ ), so  $J$  is an open interval.

If  $I = [a, b]$ ,  $a < b$  and  $c \in (a, b)$ , then the above argument is still valid, but we cannot preclude the possibility that  $a$  or  $b$  (or both) belong to the immediate basin of attraction.

When the fixed point is  $c = a$ , the immediate basin of attraction is the form  $[a, d)$  ( $a < d < b$ ), or  $I = [a, b]$ . Similar considerations apply when  $c = b$  and also when  $I$  is a half open interval.

□

The basin of attraction also has the property of being an invariant set for  $f$ :

**Definition 2.1.3** The set  $J \subset I$  is an *invariant set* for the dynamical system  $(f, I)$  if  $f(J) \subseteq J$ .

**Proposition 2.1.4** *Let  $f : I \rightarrow I$  be a continuous function on the interval  $I$ , having a fixed point  $c \in I$ .*

(a) *The basin of attraction  $B_f(c)$  is invariant under  $f$ .*

(b) *If  $c$  is an attracting fixed point of  $f$ , with immediate basin of attraction equal to the interval  $J$ , then  $f(J) \subset J$ .*

**Proof.** (a) Let  $x \in B_f(c)$ , then  $f^n(x) \rightarrow c$  as  $n \rightarrow \infty$ . By the continuity of  $f$ ,  $f(f^n(x)) \rightarrow f(c)$  as  $n \rightarrow \infty$ . In other words,  $f^n(f(x)) \rightarrow c$  as  $n \rightarrow \infty$ , so  $f(x) \in B_f(c)$ .

(b) See Exercises 2.2. □

**Examples 2.1.5** 1. If  $f(x) = x^2$ , the fixed points are  $c = 0$  and  $c = 1$ , both hyperbolic, the first being attracting and the second repelling. Clearly  $B_f(0) = (-1, 1)$  and  $B_f(1) = \{-1, 1\}$ . We often regard  $c = \infty$  as an attracting fixed point of  $f$ , so that  $B_f(\infty) = [-\infty, -1] \cup (1, \infty]$  (see the Chapter 14 on complex dynamics).

2. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^2 + 1/4$ . We have seen that  $x = 1/2$  is the only fixed point, with stability on the left, but not on the right (so  $x = 1/2$  is not an attracting fixed point and Proposition 2.1.2 is not applicable). We use ad hoc methods to find the basin of attraction of the fixed point.

If  $x \neq 1/2$ ,  $f(x) > x$  (this is equivalent to  $(2x - 1)^2 > 0$ ), and we see inductively that  $f^n(x)$  is an increasing sequence. Furthermore, if  $x < 1/2$ , then  $f(x) < 1/2$ , and by induction,  $f^n(x) < 1/2$  for all  $n \geq 1$ . It follows that if  $x < 1/2$ ,  $f^n(x) \rightarrow 1/2$  as  $n \rightarrow \infty$  (since  $1/2$  is the only fixed point). Since  $f$  is repelling on the right of the fixed point, we conclude that  $B_f(1/2) = (-\infty, 1/2]$ .

3. Suppose that  $L_\mu : [0, 1] \rightarrow [0, 1]$ ,  $L_\mu(x) = \mu x(1 - x)$ , then we saw that when  $\mu = 1$ ,  $x = 0$  is a fixed point that is stable on the right, unstable on the left. Proposition 2.1.2 is still applicable in this situation. It tells us that the immediate basin of attraction of  $c = 0$  is an interval of the form  $[0, d)$  for some  $0 < d < 1$ , or it is equal to  $[0, 1]$ . In the next section we show that it is equal to  $[0, 1]$ .

## 2.2 The Logistic Family.

The logistic maps  $L_\mu(x) = \mu x(1 - x)$  are functions of two real variables  $\mu$  and  $x$ . We usually restrict  $x$  to the interval  $[0, 1]$ , and in this chapter we restrict  $\mu \in (0, 4]$  so that  $L_\mu$  is a dynamical system of  $[0, 1]$ .

$\mu$  is a *parameter* which we allow to vary, but then we study the function  $L_\mu$  for specific fixed values of this parameter. As the parameter  $\mu$  is varied, we shall see a corresponding change in the nature of the function  $L_\mu$ . This is called *bifurcation*. For example, if  $0 < \mu \leq 1$ ,  $L_\mu$  has exactly one fixed point in  $[0, 1]$ ,  $c = 0$ , which is attracting. As  $\mu$  increases beyond 1, a new fixed point  $c = 1 - 1/\mu$ , is created in  $[0, 1]$ , so now  $L_\mu$  has two fixed points. The fixed point  $c = 0$  is now repelling and

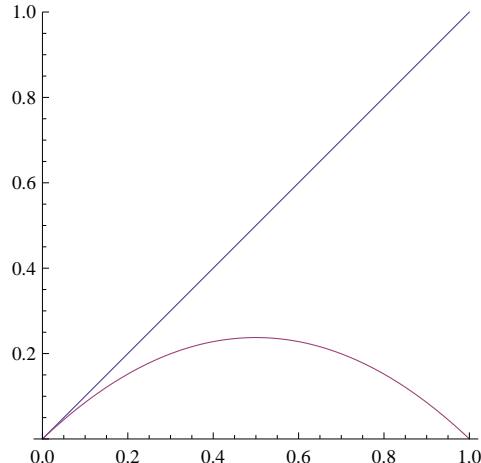
$c = 1 - 1/\mu$  is attracting (for  $1 < \mu \leq 3$ ). At  $\mu = 3$ , the nature of the fixed points changes. In this section we determine the basin of attraction of the fixed points as  $\mu$  increases from 0 to 3. We notice that the “dynamics” (long term behavior) of iterates of  $L_\mu(x)$  is quite well behaved for this range of values of  $\mu$ .

The function  $L_\mu(x) = \mu x(1 - x)$ ,  $0 \leq x \leq 1$ , has a maximum value of  $\mu/4$  when  $x = 1/2$ . Consequently, for  $0 < \mu \leq 4$ ,  $L_\mu$  maps the unit interval  $[0, 1]$  into itself. We shall consider later what happens when  $\mu > 4$ . We start by showing that the basin of attraction of  $L_\mu$ , for  $0 < \mu \leq 1$  is all of the domain of  $L_\mu$ , namely  $[0, 1]$ . We say that 0 is a *global attractor* in this case.

**Theorem 2.2.1** *Let  $L_\mu(x) = \mu x(1 - x)$ ,  $0 \leq x \leq 1$  be the logistic map. For  $0 < \mu \leq 1$ ,  $B_{L_\mu}(0) = [0, 1]$ , and for  $1 < \mu \leq 3$ ,  $B_{L_\mu}(1 - 1/\mu) = (0, 1)$ .*

We split the proof into a number of different cases:

**Case 2.2.2  $0 < \mu \leq 1$ .**



For  $0 < \mu < 1$ , the only fixed point is 0.

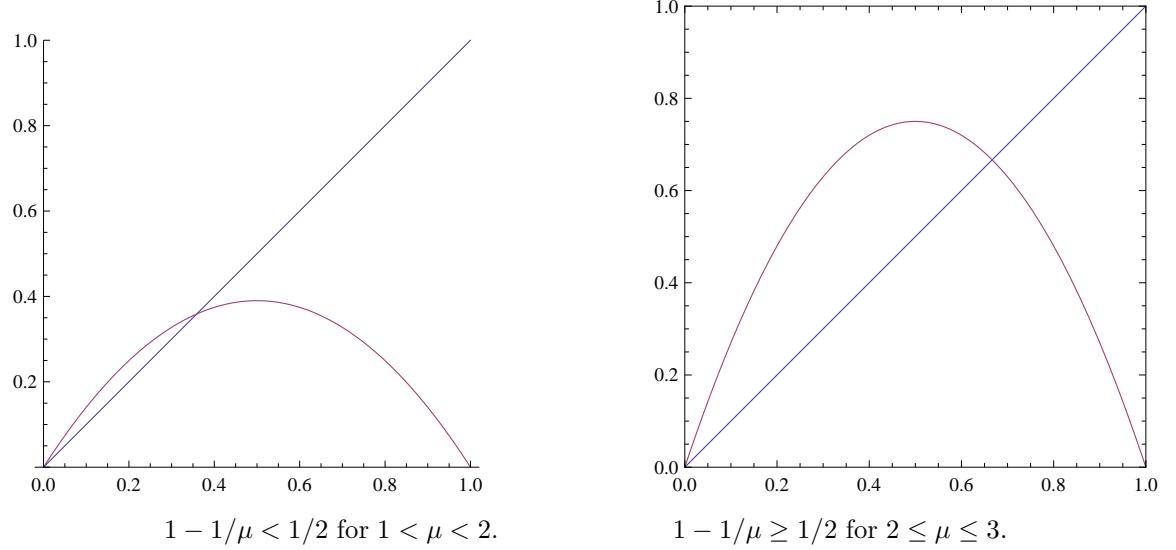
We have seen that for  $\mu \in (0, 1)$ ,  $L_\mu$  has only the one fixed point  $x = 0$  in  $[0, 1]$  (the other fixed point is  $1 - 1/\mu \leq 0$ ). For  $\mu < 1$  this fixed point is asymptotically stable, (strictly it has only stability on the right at  $x = 0$ , but it is asymptotically stable when we restrict  $L_\mu$  to  $[0, 1]$ ). In any case

$$0 < \mu \leq 1, \quad 0 < 1 - x < 1 \Rightarrow 0 < \mu(1 - x) < 1$$

$$\Rightarrow 0 < L_\mu(x) = \mu x(1 - x) < x, \quad x \in (0, 1].$$

In a similar way,  $L_\mu^2(x) < L_\mu(x)$ , and we see that the sequence  $(L_\mu^n(x))$  is decreasing, bounded below by 0, and hence must converge to the only fixed point, namely 0. It follows that the basin of attraction is  $B_{L_\mu}(0) = [0, 1]$ , ( $L_\mu(1) = 0$ ).

**Case 2.2.3  $1 < \mu \leq 3$ .**



We have seen that for  $\mu > 1$  the fixed point 0 is repelling, but a new fixed point  $c = 1 - 1/\mu$  has been created, which is attracting (for  $1 < \mu \leq 3$ ). By Proposition 2.1.2 the immediate basin of attraction is an open interval  $I = (a, b)$ , containing the fixed point with  $L_\mu(I) \subseteq I$  (from Proposition 2.1.4 (b)). If the basin of attraction of  $c$  is  $B_\mu(c)$ , then  $0, 1 \notin B_\mu(c)$  because  $L_\mu(0) = 0$  and  $L_\mu(1) = 0$ , so  $B_\mu(c) \neq [0, 1]$ . Furthermore, clearly  $a, b \notin B_\mu(c)$ .

Let  $x_n$  be a sequence in  $(a, b)$  with  $\lim_{n \rightarrow \infty} x_n = a$ . By the continuity of  $L_\mu$ ,  $\lim_{n \rightarrow \infty} L_\mu(x_n) = L_\mu(a)$ . Since  $x_n \in (a, b)$  for every  $n \in \mathbb{Z}^+$ , we have  $L_\mu(x_n) \in (a, b)$  for all  $n \in \mathbb{Z}^+$ . Since  $L_\mu(a) \notin (a, b)$ , this can only happen if  $L_\mu(a) = a$  or  $L_\mu(a) = b$ , and similarly for  $b$ . The only possibilities for  $a$  and  $b$  are:

- (i)  $a$  and  $b$  are fixed points:  $f(a) = a$ ,  $f(b) = b$ ,
- (ii)  $L_\mu(a) = b$ ,  $L_\mu(b) = a$  (we call  $\{a, b\}$  a *2-cycle* in this case - see Section 2.3), or
- (iii)  $a$  and  $b$  are eventual fixed points:  $f(a) = a$ ,  $f(b) = a$  or  $f(a) = b$ ,  $f(b) = b$ .

Clearly (i) does not hold, and we will show that (ii) cannot happen, so (iii) must hold. This leads to the conclusion that  $a = 0$  and  $b = 1$ , since there are no other fixed

or eventual fixed points in  $[0, 1]$  that can satisfy these conditions. Consequently, we must have  $B_\mu(1 - 1/\mu) = (0, 1)$ .

Suppose that (ii) holds for  $a$  and  $b$ . We can check that

$$L_\mu^2(x) - x = x[\mu^2x^2 - \mu(\mu + 1)x + \mu + 1](\mu x - \mu + 1),$$

and  $a$  and  $b$  must satisfy this equation (why? - see Exercises 2.2). We can disregard the linear factors as they give the two fixed points. The discriminant of the quadratic factor is

$$\mu^2(\mu + 1)^2 - 4\mu^2(\mu + 1) = \mu^2(\mu + 1)(\mu - 3) < 0$$

for  $1 < \mu < 3$ , so there is no solution  $\{a, b\}$  when  $1 < \mu < 3$ . When  $\mu = 3$ , the discriminant is zero, the fixed point is  $c = 2/3$  and the quadratic factor gives rise to no additional roots, so again there is no 2-cycle.

□

**Remarks 2.2.4** Population biologists have proposed the logistic map as a model for the growth of various types of populations. For  $L_\mu$  with  $0 < \mu \leq 1$ , the results of this section tell us that if we have an initial population  $x_0 \in (0, 1)$ , then  $L_\mu^n(x_0) \rightarrow 0$  as  $n \rightarrow \infty$ , so the population will rapidly die off. On the other hand, if  $1 < \mu \leq 3$ , the population will reach an equilibrium point  $c = 1 - 1/\mu$ , where  $c$  is the solution to the equation  $L_\mu(x) = x$ , the fixed point of  $L_\mu$ . In Section 2.4 we examine what happens to the population when  $\mu > 3$ .

## Exercises 2.2

1. Find the basins of attraction of the fixed points of the following functions:

(i)  $f(x) = x^3$ , (ii)  $f(x) = x(x^2 - 3)$ , (iii)  $f(x) = \frac{x}{2} + \frac{1}{2x}$ .

2. Show that when  $\mu = 2$ ,  $x = 1/2$  is a super-attracting fixed point of  $L_\mu$ .

3. In the proof of Theorem 2.2.1, show that if  $L_\mu(a) = b$  and  $L_\mu(b) = a$ , then both  $a$  and  $b$  satisfy the equation

$$L_\mu^2(x) - x = x[\mu^2x^2 - \mu(\mu + 1)x + \mu + 1](\mu x - \mu + 1) = 0.$$

Why can we disregard the linear factors?

4. If  $L_\mu(x) = \mu x(1 - x)$  is the logistic map, show that  $x = 1/2$  is the only critical point of  $L^2(x)$  for  $0 < \mu \leq 2$ , but when  $\mu > 2$ , two new critical points are created. Use this to show that for  $2 < \mu < 3$ , the interval  $[1/\mu, 1 - 1/\mu]$  is mapped by  $L_\mu^2$  onto the interval  $[1/2, 1 - 1/\mu]$ .
5. Let  $f : I \rightarrow I$  be a continuous function on the interval  $I$ . If  $J \subset I$  is the immediate basin of attraction of an attracting fixed point  $c$ , show that  $f(J) \subset J$ , i.e.,  $J$  is an invariant set. (Hint: Use the fact that the image of an interval under a continuous function is an interval).
6. If  $L_\mu(x) = \mu x(1 - x)$  is the logistic map with  $1 < \mu \leq 2$ , show directly that the basin of attraction of the fixed point  $c = 1 - 1/\mu$  is  $(0, 1)$  using the following steps:
- Consider the graph of  $L_\mu$ , and note that  $1 - 1/\mu < 1/2$ , and  $1/\mu$  is an eventual fixed point of  $L_\mu$ .
  - If  $0 < x < 1 - 1/\mu$ , show that  $L_\mu^n(x)$  is an increasing sequence bounded above by  $1 - 1/\mu$ , so must converge to  $1 - 1/\mu$ .
  - If  $1 - 1/\mu < x \leq 1/2$ , show that  $L_\mu^n(x)$  is a decreasing sequence bounded below by  $1 - 1/\mu$ , so must converge to  $1 - 1/\mu$ .
  - Now consider what happens if  $1/2 < x \leq 1/\mu$  and  $1/\mu < x < 1$ . (Hint: It is helpful to look at the graph of  $L_\mu$ ).
7. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be continuous with two attracting fixed points  $\alpha$  and  $\beta$  with basins of attraction  $B_f(\alpha) = (-\infty, p)$  and  $B_f(\beta) = (p, \infty)$ . What can be said about the point  $x = p$ ? Give an example of a map with this property.
8. Use an argument similar to that in Case 2.2.3 for  $L_\mu$ , to show that if  $0 < \mu < 1$ , then the basin of attraction of  $c = 0$  is  $[0, 1]$ .
9. Let  $f : [0, 1] \rightarrow [0, 1]$  be a continuous function with  $f(0) = 0$ ,  $f(1) = 0$  and  $f(x) > 0$  for  $x \in (0, 1)$ . If  $x = 0$  is a repelling fixed point and there is a unique fixed point  $c \in (0, 1)$ , which is attracting. Use an argument similar to that in Case 2.2.3 to

show that if  $f$  has no period 2-points, then the basin of attraction of the fixed point  $c$  is  $(0, 1)$ .

10\*. Use the method of question 6 to show that for  $2 < \mu \leq 3$ , the basin of attraction of the fixed point  $c = 1 - 1/\mu$  is  $(0, 1)$ .

### 2.3 Periodic Points.

Points having finite orbits are important in the study of dynamical systems. We have seen that the fixed points of a dynamical system may determine the long term behavior of the dynamical system. If a continuous map has a finite number of fixed points and no points of greater period, it seems likely that an iterate will converge to one of the fixed points, and the long term behavior will be uncomplicated (we often regard  $\infty$  as a fixed point). The same may be true for period 2-points, or points of greater period. Usually chaotic behavior or unpredictability in the long-term, occurs when there are infinitely many points having different periods.

**Definition 2.3.1** Let  $f : I \rightarrow I$  be a dynamical system with  $c \in I$ .

(i)  $c$  is a *periodic point* of  $f(x)$  with *period*  $r \in \mathbb{Z}^+$ , if  $f^r(c) = c$  and  $f^k(c) \neq c$  for  $0 < k < r$  (in particular,  $c$  is a fixed point of  $f^r$ ). The set

$$O(c) = \{c, f(c), f^2(c), \dots, f^{r-1}(c)\},$$

is called an *r-cycle* for  $f$ . We write

$$\text{Per}_r(f) = \{x \in I : f^r(x) = x\},$$

so that  $\text{Fix}(f) \subseteq \text{Per}_n(f)$ ,  $n = 1, 2, \dots$ , since the points in  $\text{Per}_n(f)$  may not be of period  $n$ , but of some lesser period.

(ii)  $c$  is *eventually periodic* for  $f$  if there exists  $m \in \mathbb{Z}^+$  such that  $f^m(c)$  is a periodic point of  $f$  (we assume that  $c$  is not a periodic point).

(iii) If  $c$  is of period  $r$ , then  $c$  is *stable* if it is a stable fixed point of  $f^r$  (respectively *asymptotically stable*, *unstable* etc.).

The following criteria for stability now follows immediately from Theorem 1.3.3:

**Theorem 2.3.2** Suppose that  $c$  is a point of period  $r$  for  $f$ , and  $f'(x)$  is continuous at  $x = c$ . If  $c_i = f^i(c)$ ,  $i = 0, 1, \dots, r - 1$  then:

(i)  $c$  is asymptotically stable if

$$|f'(c_0) \cdot f'(c_1) \cdot f'(c_2) \cdots f'(c_{r-1})| < 1.$$

(ii)  $c$  is unstable if

$$|f'(c_0) \cdot f'(c_1) \cdot f'(c_2) \cdots f'(c_{r-1})| > 1.$$

**Proof.** Let us look at the case where  $r = 3$  as this is typical:

$$O(c) = \{c, f(c), f^2(c)\} = \{c_0, c_1, c_2\}.$$

Using the chain rule on the third iterate gives

$$\begin{aligned} \frac{d}{dx}(f^3(x)) &= \frac{d}{dx}(f(f^2(x))) = f'(f^2(x))(f^2(x))' = f'(f^2(x))f'(f(x))f'(x) \\ &= f'(c_2)f'(c_1)f'(c_0), \quad \text{when } x = c. \end{aligned}$$

The result now follows from Theorem 1.4.4. □

**Example 2.3.3** Consider the quadratic function  $f(x) = x^2 - 2$ . To find the fixed points we solve  $f(x) = x$ , or  $x^2 - 2 = x$ ,  $x^2 - x - 2 = (x - 2)(x + 1) = 0$ , so  $x = 2$  or  $x = -1$ .

To find the period 2-points we solve  $f^2(x) = x$  or  $f^2(x) - x = 0$ . This is simplified when we realize that the fixed points must be solutions of this equation, so that  $(x - 2)(x + 1)$  is a factor. We can then check that

$$f^2(x) - x = x^4 - 4x^2 - x + 2 = (x - 2)(x + 1)(x^2 + x - 1).$$

Solving the quadratic gives

$$x = \frac{-1 \pm \sqrt{5}}{2},$$

so that  $\{\frac{-1 + \sqrt{5}}{2}, \frac{-1 - \sqrt{5}}{2}\}$  is a 2-cycle. In general, finding periodic points of quadratics can be complicated. If  $f(x)$  is a quadratic,  $f^n(x) - x$  is a polynomial of degree  $2^n$ .

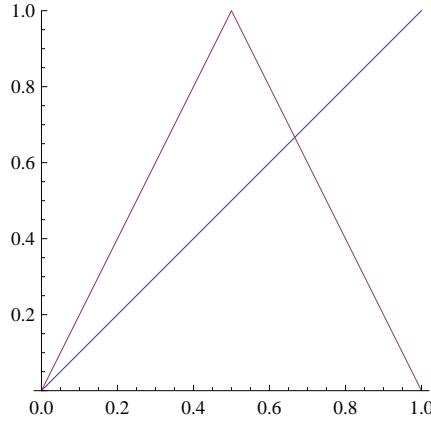
To check the stability, we calculate

$$|f'((-1 - \sqrt{5})/2)f'((-1 + \sqrt{5})/2)| = |(-1 - \sqrt{5})(-1 + \sqrt{5})| = |1 - 5| = 4 > 1,$$

giving an unstable 2-cycle.

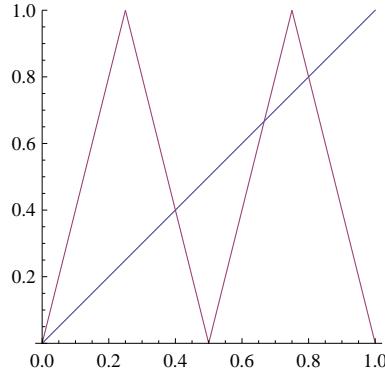
**Example 2.3.4** If we look at the graph of the tent map  $T$ , we see it has two fixed points,  $c = 0$  and  $c = 2/3$ . If we graph  $T^3$ , it has eight fixed points, arising from two 3-cycles:  $\{2/7, 4/7, 6/7\}$  and  $\{2/9, 4/9, 8/9\}$ , together with the two fixed points of  $T$ , so that

$$\text{Per}_3(T) = \{0, 2/3, 2/7, 4/7, 6/7, 2/9, 4/9, 8/9\}.$$



The tent map has fixed points at  $c = 0$  and  $c = 2/3$ .

The graph of  $T^2$  shows that  $T^2$  has four fixed points coming from a 2-cycle  $\{2/5, 4/5\}$  and the two fixed points of  $T$ . Since  $|T'(x)| = 2$  and  $|(T^2)'(x)| = |T'(x)||T'(Tx)| = 4$ , etc. (except at points of non-differentiability), all periodic points will be unstable.

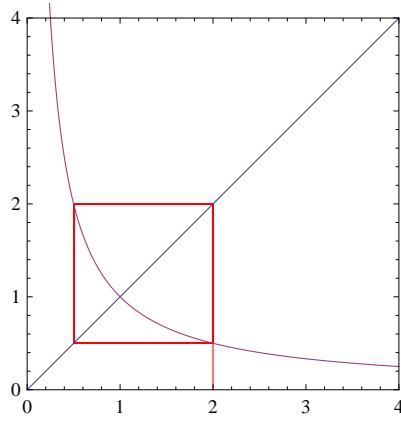


$\{2/5, 4/5\}$  is a 2-cycle for the tent map.

**Example 2.3.5** Let  $f(x) = 1/x$ ,  $x \neq 0, x \neq \pm 1$ . Note that  $f^2(x) = x$  and  $f(x) \neq x$  for all such  $x$ , giving rise to the 2-cycle  $\{x, 1/x\}$ . In this case

$$|f'(x)f'(1/x)| = |-1/x^2(-x^2)| = 1,$$

so the theorem is inconclusive. However, we see that the periodic points are stable but are neither attracting nor repelling. These are examples of non-hyperbolic 2-cycles  $\{c_0, c_1\}$ : either  $(f^2)'(c_0) = f'(c_0)f'(c_1) = 1$ , or  $(f^2)'(c_0) = f'(c_0)f'(c_1) = -1$ .



Every  $x \neq 0, \pm 1$  gives rise to a 2-cycle.

**Definition 2.3.6** Let  $\{c_0, c_1, \dots, c_{n-1}\}$  be an  $n$ -cycle for a differentiable function  $f$ . This  $n$ -cycle is *hyperbolic* if  $|f'(c_0)f'(c_1)\dots f'(c_{n-1})| \neq 1$ . This is equivalent to saying that any member of the set  $\{c_0, c_1, \dots, c_{n-1}\}$  is a hyperbolic fixed point of  $f^n$  (see Exercises 2.3). If the  $n$ -cycle is not hyperbolic, we say it is *non-hyperbolic*.

**Example 2.3.7** Let  $f(x) = -x^3/2 - x/2 + 1$ , then  $f(0) = 1$  and  $f(1) = 0$ , so  $\{0, 1\}$  is a 2-cycle for  $f$ . We can check that  $f'(0)f'(1) = 1$ , so this 2-cycle is non-hyperbolic. Note that if

$$g(x) = f^2(x) - x = x^2(x-1)^2(x^2+2x+3)(x^3+3x-2)/16,$$

then  $g(x)$  has repeated roots at  $x = 0$  and  $x = 1$ . This will always happen when we have a hyperbolic 2-cycle of the above type (see Exercises 2.3).

**Remark 2.3.8** 1. Periodic points can be stable but not attracting (as above with  $f(x) = 1/x$  at  $x \neq 1$ ). They can also be attracting but not stable as in Example 1.4.7).

2. Functions such as  $f(x) = \sin x$  can have no period 2 points or points of a higher period, since this would contradict the basin of attraction of  $x = 0$  being all of  $\mathbb{R}$ . Similarly, the logistic map  $L_\mu$ ,  $0 < \mu \leq 3$  cannot have period  $n$ -points for  $n > 1$ .

### Exercises 2.3

1. For each of the following functions,  $c = 0$  lies on a periodic cycle. Classify this cycle as attracting, repelling or neutral (non-hyperbolic). State if it is super-attracting:

$$(i) \quad f(x) = \frac{\pi}{2} \cos x, \quad (ii) \quad g(x) = -\frac{1}{2}x^3 - \frac{3}{2}x^2 + 1.$$

2. Let  $f_c(x) = x^2 + c$ . Show that for  $c < -3/4$ ,  $f_c$  has a 2-cycle, and find it explicitly. For what values of  $c$  is the 2-cycle attracting?

3. Let  $a, b, c \in \mathbb{R}$ . Investigate the existence of 2-cycles for the following maps:

$$(a) \quad f(x) = ax + b, \quad a \neq 0.$$

$$(b) \quad f(x) = ax^2 - x + c, \quad a, c > 0.$$

$$(c) \quad f(x) = a - \frac{b}{x}, \quad a \neq 0, \quad b \neq 0.$$

$$(d) \quad f(x) = \frac{ax + b}{cx - a}, \quad a^2 + bc \neq 0.$$

4. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be continuous.

(a) If  $f$  has a 2-cycle  $\{x_1, x_2\}$ , show that  $f$  has a fixed point.

(b) If  $f$  has a 3-cycle  $\{a, b, c\}$ ,  $a < b < c$  with  $f(a) = b$ ,  $f(b) = c$  and  $f(c) = a$ , show that there is a fixed point  $x_0$  with  $b < x_0 < c$ , and a point  $x_1$ :  $a < x_1 < b$  with  $f^2(x_1) = x_1$ .

5. Suppose that  $\{0, 1\}$  is a non-hyperbolic 2-cycle for  $f(x) = ax^2 + bx + c$ , where  $a \neq 0$ .

- (a) Find the possible values of  $a$ ,  $b$  and  $c$  and the corresponding function  $f$ .
- (b) Graph  $f(x)$  and  $f^2(x)$  in each case using a computer algebra system (note that the slopes of  $f^2(x)$  at  $x = 0$  and  $x = 1$  are the same - why is that?).
- (c) Use the computer algebra system to find  $Sf(0)$  and  $Sf(1)$  (the Schwarzian derivative), and deduce the stability of the 2-cycle.
6. (a) If  $f(x) = x^3 - 6x^2 + 7x + 2$ , show that  $\{1, 3\}$  is a 2-cycle for the Newton function  $N_f$ .
- (b) Determine the stability of the 2-cycle of  $N_f$ . (Hint: It is helpful to recall that  $N'_f(x) = f(x)f''(x)/[f'(x)]^2$ ).
7. Let  $f(x) = ax^3 + bx + 1$ ,  $a \neq 0$ . If  $\{0, 1\}$  is a 2-cycle for  $f(x)$ , find  $a$  and  $b$  so that the 2-cycle is non-hyperbolic, and determine the stability.
8. (a) Show that  $C_\mu(x) = \mu \cos(x)$  has a super-attracting 3-cycle  $\{0, \lambda, \pi/2\}$ , where  $\mu = \lambda$  and  $\lambda$  satisfies the equation  $\lambda \cos(\lambda) = \pi/2$ .
- (b) Give similar conditions for  $S_\mu(x) = \mu \sin(x)$  to have (i) a super-attracting 2-cycle, (ii) a super-attracting 3-cycle.
- (c) Show that  $x = \pi/2$  is a super-attracting fixed point of  $S_{\pi/2}$ , whose immediate basin of attraction (an open interval) is strictly contained in its basin of attraction. Use a computer algebra system to show that a 2-cycle is created when  $\mu = 2.6182\dots$  approximately, and deduce that the immediate basin of attraction of the fixed point  $x = \pi/2$  is  $(0, \pi)$ .
- (d) Do a similar analysis for  $C_\mu$ , for a suitable value of  $\mu$  and corresponding fixed point.
9. Explain why, for families of maps, say  $F_\mu$ , one member of a super-attracting  $n$ -cycle is a super-attracting fixed point (for a different value of  $\mu$ ).

10. Let  $f(x) = xe^{p-x}$ , for  $x > 0$  and  $p > 0$ .
- Find the fixed points and their stability.
  - Using the results of Exercises 1.5 # 12, applied to  $f^2(x)$ , show that for  $p > 2$ ,  $f(x)$  has a 2-cycle.
11. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a function which is differentiable everywhere. If  $\{x_0, x_1, \dots, x_{n-1}\}$  is an  $n$ -cycle for  $f$ , show that the derivative of  $f^n$  is the same at each  $x_i$ ,  $i = 0, 1, \dots, n-1$ .
12. Show that if  $x_0$  is a periodic point of  $f$  with period  $n$ , and  $f^m(x_0) = x_0$   $m \in \mathbb{Z}^+$ , then  $m = kn$  for some  $k \in \mathbb{Z}^+$ . (Hint: Write  $m = qn + r$  for some  $r$ ,  $0 \leq r < n$  and show that  $r = 0$ ).
13. Show that if  $f^p(x) = x$  and  $f^q(x) = x$ , and  $n$  is the highest common factor of  $p$  and  $q$ , then  $f^n(x) = x$ . (Hint: Use the previous exercise and the fact that every common factor of  $p$  and  $q$  is a factor of  $n$ ).
14. (a) Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be an *odd function* ( $f(-x) = -f(x)$  for all  $x \in \mathbb{R}$ ). Show that  $x = 0$  is a fixed point of  $f$ , and that the intersection of the graph of  $f$  with the line  $y = -x$  gives rise to points  $c$  of period 2 (when  $c \neq 0$ ).
- (b) Use part (a) to find the 2-cycles of  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^3 - 3x/2$  and determine their stability.
- (c) Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be an *even function* ( $f(-x) = f(x)$  for all  $x \in \mathbb{R}$ ). Show that the intersection of the graph of  $f$  with the line  $y = -x$  gives rise to eventually fixed points  $c$  (when  $c \neq 0$ ). What are the eventually fixed points of  $f(x) = \cos(x)$ ?
15. (a) Let  $f(x) = \begin{cases} x \sin(1/x); & x \neq 0 \\ 0; & x = 0 \end{cases}$ . In Exercises 1.5 # 8, we found the fixed points of  $f$ , and showed that the non-zero fixed points are non-hyperbolic. Find the eventual fixed points of  $f$ . (Hint: Note that  $f(x)$  is an even function).

(b) Discuss the stability of the fixed points of  $f(x)$ , their basins of attractions and the existence of period 2-points. Note that  $f(x) \leq x$  for all  $x > 0$ .

(c) Now set  $g(x) = \begin{cases} x \cos(1/x); & x \neq 0 \\ 0; & x = 0 \end{cases}$ . Find the fixed points of  $g$ , and show that in this case there are period 2-points. (Hint: Note that  $g(x)$  is an odd function).

16. (a) Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a continuously differentiable function having exactly two distinct fixed points  $x_0, x_1$  with  $|f'(x_i)| > 1$ , for  $i = 0, 1$ . Show that  $f$  has a 2-cycle.

(b) Use (a) to show that the logistic map  $L_\mu(x) = \mu x(1 - x)$  has a 2-cycle, for  $\mu > 3$ .

17. Suppose that  $f(x) = ax^2 + bx + c$ ,  $a \neq 0$ , has a 2-cycle  $\{x_0, x_1\}$ . Show that the 2-cycle cannot be non-hyperbolic of the type where  $f'(x_0)f'(x_1) = 1$ .

18. Let  $f(x)$  be a polynomial for which  $g(x) = f^2(x) - x$  has a repeated root at  $x_0$  (where  $f(x_0) = x_1 \neq x_0$ ). Show that  $\{x_0, x_1\}$  is a non-hyperbolic 2-cycle for  $f$  of the type where  $f'(x_0)f'(x_1) = 1$ . Does the converse holds?

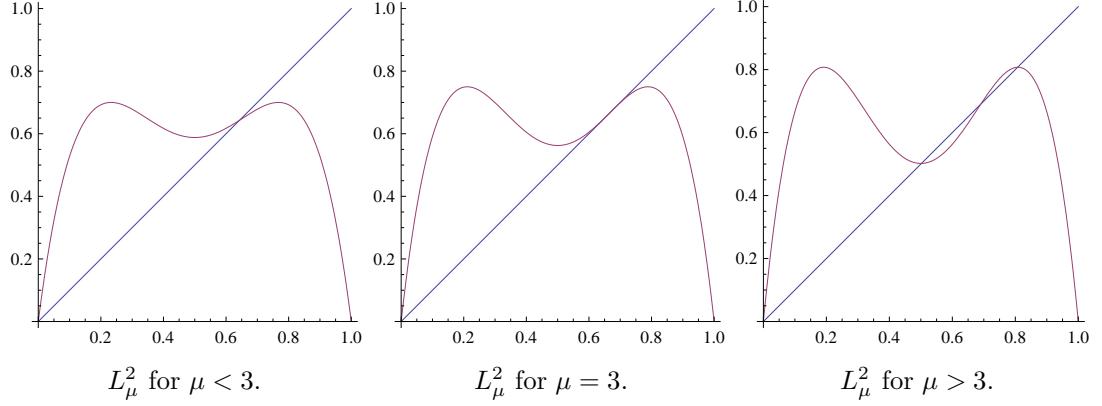
## 2.4 Periodic Points of the Logistic Map.

If we look at the graph of  $L_\mu^2$  for various values of  $\mu$  close to 3, we can see how the points of period 2 are created. For  $\mu$  close to, but less than 3, the graph of  $L_\mu^2$  intersects  $y = x$  in exactly one place  $x = c$  (besides  $x = 0$ ). This is the fixed point  $c = 1 - 1/\mu$  of  $L_\mu$ , and the slope here is  $L'_\mu(c) = 2 - \mu$ , having absolute value less than 1. Clearly there are no period-2 points.

When  $\mu = 3$ , the graph of  $L_\mu^2$  is tangential to  $y = x$  at the fixed point  $c$ ,  $L'_\mu(c) = -1$  and again there are no period-2 points. As  $\mu$  increases beyond 3, the graph of  $L_\mu^2$  “snakes around”  $c$ , creating two new fixed points of  $L_\mu^2$ , say  $\{c_1, c_2\}$ , which must be a 2-cycle for  $L_\mu$ . Notice that at the fixed point  $c$ ,

$$(L_\mu^2)'(c) = L'_\mu(c)L'_\mu(L_\mu(c)) = (2 - \mu)^2 > 1,$$

so the slope at  $c$  (for  $L_\mu^2$ ) has increased from being less than one, to being greater than one, and this has forced the creation of the 2-cycle. This is an example of a *period doubling bifurcation*.



To find the 2-cycles of the logistic map  $L_\mu(x) = \mu x(1 - x)$ ,  $0 \leq x \leq 1$ , we solve the equation

$$L_\mu^2(x) = x,$$

or

$$\mu x(1 - x)[1 - \mu x(1 - x)] - x = 0,$$

or

$$-\mu^3 x^4 + 2\mu^3 x^3 - (\mu^3 + \mu^2)x^2 + \mu^2 x - x = 0.$$

Clearly,  $x$  is a factor (since  $c = 0$  is a fixed point of  $L_\mu(x)$ ), and similarly  $x - (1 - 1/\mu)$  must be a factor. Thus

$$L_\mu^2(x) - x = -x(\mu x - \mu + 1)(\mu^2 x^2 - \mu(\mu + 1)x + \mu + 1),$$

giving a quadratic equation which we will see has no roots if  $\mu < 3$ ,

$$\mu^2 x^2 - \mu(\mu + 1)x + \mu + 1 = 0.$$

Solving, using the quadratic formula gives

$$\begin{aligned} c &= \frac{\mu(\mu + 1) \pm \sqrt{\mu^2(\mu + 1)^2 - 4\mu^2(\mu + 1)}}{2\mu^2} \\ &= \frac{(1 + \mu) \pm \sqrt{(\mu - 3)(\mu + 1)}}{2\mu}. \end{aligned}$$

These are real only for  $\mu \geq 3$  (called the “birth of period two”). Let us call these roots  $c_1$  and  $c_2$  (dependent on  $\mu$ ).

The 2-cycle  $\{c_1, c_2\}$  is asymptotically stable if

$$|(L_\mu^2)'(c_1)| = |L'_\mu(c_1)L'_\mu(c_2)| < 1,$$

or

$$\begin{aligned} -1 &< \mu^2(1 - 2c_1)(1 - 2c_2) < 1, \\ -1 &< (-1 - \sqrt{(\mu^2 - 2\mu - 3)})(-1 + \sqrt{(\mu^2 - 2\mu - 3)}) < 1, \\ -1 &< 1 - (\mu^2 - 2\mu - 3) < 1. \end{aligned}$$

This gives rise to the two inequalities

$$\mu^2 - 2\mu - 3 > 0 \quad \text{and} \quad \mu^2 - 2\mu - 5 < 0,$$

and solving

$$3 < \mu < 1 + \sqrt{6},$$

giving the condition for asymptotic stability of the 2-cycle  $\{c_1, c_2\}$ .

For  $\mu = 1 + \sqrt{6}$ , it can be seen that

$$L'_\mu(c_1)L'_\mu(c_2) = -1, \quad \text{and} \quad SL_\mu^2(c_1) < 0,$$

so Theorem 1.5.7 (i) shows that the 2-cycle is asymptotically stable. Also, the 2-cycle is unstable for  $\mu > 1 + \sqrt{6}$ . In summary,

**Theorem 2.4.1** *For  $3 < \mu \leq 1 + \sqrt{6}$ , the logistic map  $L_\mu(x) = \mu x(1 - x)$  has an asymptotically stable 2-cycle. This 2-cycle is unstable for  $\mu > 1 + \sqrt{6}$ .*

The above shows that a bifurcation occurs when  $\mu = 3$ , a 2-cycle is created which was not previously present. There is another bifurcation at  $\mu = 1 + \sqrt{6}$ . This means that for  $3 < \mu \leq 1 + \sqrt{6}$ , when we use graphical iteration of points close to  $c_1$  and  $c_2$ , they will approach the period 2 orbit, and not the fixed point (which is now unstable). In fact, it can be shown that for this range of values of  $\mu$ , the basin of attraction of the 2-cycle consists of all of  $(0, 1)$ , (except for the fixed point  $1 - 1/\mu$  and eventual fixed points such as  $1/\mu$ ). When  $\mu$  exceeds  $1 + \sqrt{6} = 3.449499\dots$ , the period 2-points become unstable. Instead, when  $\mu = 1 + \sqrt{6}$ , we have another bifurcation, with the birth of an attracting period 4-cycle. This is called *period doubling*.

## Exercises 2.4

1. Let  $f_c(x) = x^2 + c$ ,  $c \in \mathbb{R}$ .

(i) For what values of  $c$  does  $f_c$  have a super-attracting fixed point, and what is the fixed point?

- (ii) For what values of  $c$  does  $f_c$  have a super-attracting 2-cycle, and find the 2-cycle?
- (iii) Show that if  $f_c$  has a super-attracting 3-cycle, then  $c$  satisfies the equation  $c^3 + 2c^2 + c + 1 = 0$ , and the 3-cycle is  $\{0, c, c^2 + c\}$  for that value of  $c$ .

## 2.5 The Period Doubling Route to Chaos.

We summarize what we have determined so far, and discuss what happens as  $\mu$  increases from zero to around 3.57.

For  $0 < \mu \leq b_1$  (where  $b_1 = 1$ ),  $c = 0$  is the only fixed point and it is attracting for these values of  $\mu$ . There is a bifurcation at  $b_1 = 1$ , where a non-zero fixed point  $c = 1 - 1/\mu$  is created. This fixed point is attracting for  $1 < \mu \leq 3$  (and  $c = 0$  is no longer attracting), and super attracting when  $\mu = s_1 = 2$ . The second bifurcation occurs when  $\mu = b_2 = 3$ . The fixed point  $c = 1 - 1/\mu$  becomes unstable, and an attracting 2-cycle is created for  $3 < \mu < 1 + \sqrt{6} = b_3$ .

### 2.5.1 A Super-Attracting 2-Cycle

We saw that when  $\mu = 2$ ,  $c = 1/2$  is a super-attracting fixed point for  $L_\mu$ . We now look for a super-attracting period 2-cycle for  $L_\mu$  when  $3 < \mu < 1 + \sqrt{6}$ , as it illustrates an important general method that can be used for finding where period three is born.

Suppose that  $\{x_1, x_2\}$  is a 2-cycle for the logistic map  $L_\mu$ , which is super-attracting, then

$$x_1 = \mu x_2(1 - x_2), \quad \text{and} \quad x_2 = \mu x_1(1 - x_1),$$

so multiplying these equations together and canceling  $x_1 x_2$  gives the equation

$$\mu^2(1 - x_1)(1 - x_2) = 1.$$

In addition, we must have

$$(L_\mu^2)'(x_1) = L_\mu'(x_1)L_\mu'(x_2) = 0,$$

so that

$$\mu^2(1 - 2x_1)(1 - 2x_2) = 0.$$

Thus either  $x_1 = 1/2$ , or  $x_2 = 1/2$ , so suppose the former holds, then  $x_2 = \mu/4$ . Substituting into the first equation gives

$$\mu^2(1 - \mu/4)(1 - 1/2) = 1,$$

or

$$\mu^3 - 4\mu^2 + 8 = 0.$$

$\mu - 2$  must be a factor of this cubic, so we have

$$(\mu - 2)(\mu^2 - 2\mu - 4) = 0,$$

giving  $\mu = 1 + \sqrt{5}$ . We have shown:

**Proposition 2.5.2** *When  $\mu = s_2 = 1 + \sqrt{5}$ ,  $\{1/2, \frac{1+\sqrt{5}}{4}\}$  is a super-attracting 2-cycle for  $L_\mu$ .*

When  $\mu$  exceeds  $b_3 = 1 + \sqrt{6}$ , the 2-cycle ceases to be attracting and becomes repelling. In addition, a 4-cycle is created which is attracting until  $\mu$  exceeds a value  $b_4$ , when it becomes repelling, and an attracting 8-cycle is created. This type of period doubling continues so that when  $\mu$  exceeds  $b_n$ , an attracting  $2^{n-1}$ -cycle is created until  $\mu$  reaches  $b_{n+1}$ . These cycles become super attracting at some  $s_n$  ( $b_n < s_n < b_{n+1}$ ). This behavior continues with  $2^n$ -cycles for all  $n \in \mathbb{Z}^+$  being created, until  $\mu$  reaches a value  $b_\infty$ , approximately 3.57. In other words, for  $b_n < \mu < b_{n+1}$ ,  $L_\mu$  has a stable  $2^n$ -cycle. It can be shown that  $b_\infty = \lim_{n \rightarrow \infty} b_n = 3.570$  approximately.

Although  $b_{n+1} - b_n \rightarrow 0$  as  $n \rightarrow \infty$ ,

$$\delta = \lim_{n \rightarrow \infty} \frac{b_n - b_{n-1}}{b_{n+1} - b_n} = 4.6692016 \dots,$$

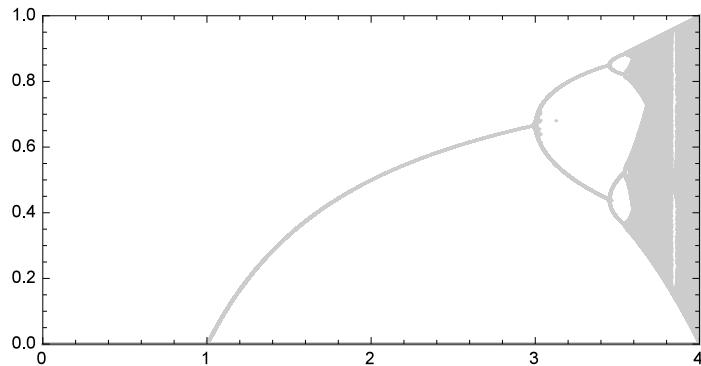
exists, and is called Feigenbaum's number. Feigenbaum showed that you get the same constant  $\delta$  in this way, for any family  $f_\mu$  of unimodal maps, ( $f_\mu(x)$  is *unimodal* if  $f_\mu(0) = 0$ ,  $f_\mu(1) = 0$ ,  $f_\mu$  is continuous on  $[0, 1]$  with a single critical point between 0 and 1).

## 2.6 The Bifurcation Diagram and 3-Cycles of the Logistic Map.

The behavior described above can be illustrated graphically using a *bifurcation diagram*. To create a bifurcation diagram we plot  $\mu$ ,  $0 \leq \mu \leq 4$ , along the  $x$ -axis, and values of  $L_\mu^n(x)$  along the  $y$ -axis. For each value of  $\mu$ , we calculate the first 500 iterates (say), of some arbitrarily chosen point  $x_0$  in  $(0, 1)$ . We ignore the first 450 iterates and plot the next 50. So for example, if  $1 < \mu < 3$ , since the fixed point is attracting, the iterates will approach the fixed point  $1 - 1/\mu$ , and for  $n$  large, what we see plotted will be (very close to) the value  $1 - 1/\mu$ . On the bifurcation diagram we will see the curve  $y = 1 - 1/\mu$ ,  $1 \leq \mu \leq 3$ . For  $3 < \mu < 1 + \sqrt{6}$  the fixed point

becomes repelling, so this no longer shows up, but the 2-cycle has becomes attracting, so we see plotted the 2 points of the 2-cycle. This continues with the 4-cycle, 8-cycle etc. This is called the *period doubling route to chaos*.

We can create a bifurcation diagram for  $L_\mu$  using a computer algebra system such as Mathematica or Maple.

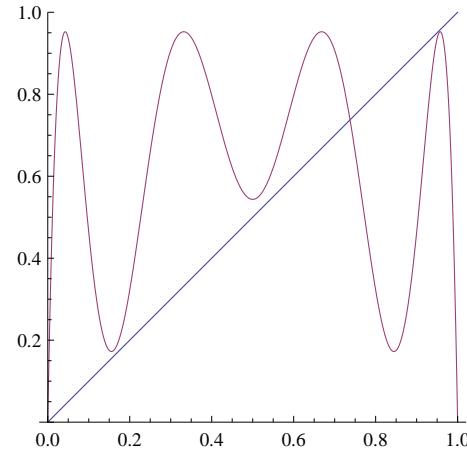


The Bifurcation Diagram for the Logistic Map.

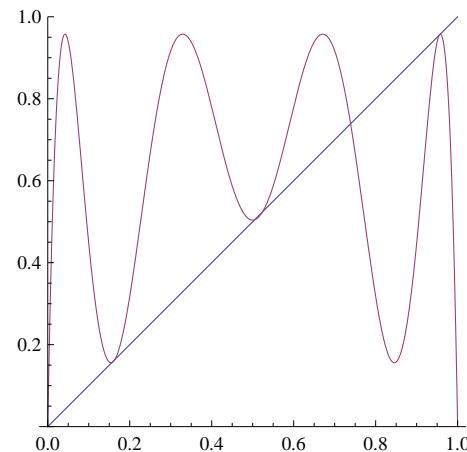
### 2.6.1 Where Does Period Three Occur for the Logistic Map?

Looking at the graph of  $L_\mu^3$  for values of  $\mu$  close to 3.8, we get some idea of the values of  $\mu$  for which there is a 3-cycle.

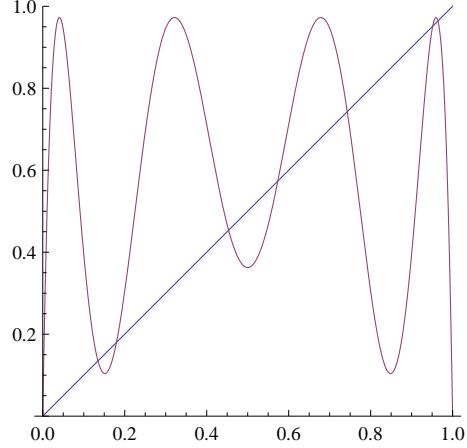
We aim to show that a 3-cycle first occurs when  $\mu = 1 + \sqrt{8}$ . For values of  $\mu$  slightly smaller than  $1 + \sqrt{8}$ , the graph of  $L_\mu^3$  shows only two fixed points - the fixed points of  $L_\mu$ . For  $\mu = 1 + \sqrt{8}$  we see that the graph of  $L_\mu^3$  touches the line  $y = x$  tangentially in three different places - these are the period 3-points, and they constitute a 3-cycle. For larger values of  $\mu$ , we see two 3-cycles. For a small range of values of  $\mu \geq 1 + \sqrt{8}$ , the 3-cycle is attracting.



The Graph of  $L_\mu^3(x)$  for  $\mu < 1 + \sqrt{8}$ .



The Graph of  $L_\mu^3(x)$  for  $\mu = 1 + \sqrt{8}$ .



The Graph of  $L_\mu^3(x)$  for  $\mu > 1 + \sqrt{8}$ .

To show that the bifurcation occurs when  $\mu = 1 + \sqrt{8}$ , we follow the argument of Feng [44] (see also [12], [63] and [112]).

Period 3-points of  $L_\mu$  occur where  $L_\mu^3(x) = x$ , so we look at where

$$L_\mu^3(x) - x = 0.$$

To disregard the fixed points of  $L_\mu$  we set

$$g_\mu(x) = \frac{L_\mu^3(x) - x}{L_\mu(x) - x}.$$

We can check that this simplifies to the following (this is most easily done using a computer algebra system):

$$\begin{aligned} g_\mu(x) = & \mu^6 x^6 - (\mu^5 + 3\mu^6)x^5 + (\mu^4 + 4\mu^5 + 3\mu^6)x^4 - (\mu^3 + 3\mu^4 + 5\mu^5 + \mu^6)x^3 \\ & + (\mu^2 + 3\mu^3 + 3\mu^4 + 2\mu^5)x^2 - (\mu + 2\mu^2 + 2\mu^3 + \mu^4)x + 1 + \mu + \mu^2. \end{aligned}$$

Set  $\lambda = 7 + 2\mu - \mu^2$  and let

$$h_\mu(z) = g_\mu(-z/\mu),$$

then

$$\begin{aligned} h_\mu(z) = & z^6 + (3\mu + 1)z^5 + (3\mu^2 + 4\mu + 1)z^4 + (\mu^3 + 5\mu^2 + 3\mu + 1)z^3 \\ & + (2\mu^3 + 3\mu^2 + 3\mu + 1)z^2 + (\mu^3 + 2\mu^2 + 2\mu + 1)z + \mu^2 + \mu + 1. \end{aligned}$$

Then if

$$k_\mu(z) = \left\{ z^3 + z^2 \frac{(3\mu + 1)}{2} + z \left( 2\mu + 3 - \frac{\lambda}{2} \right) + \frac{(\mu + 5)}{2} - \frac{\lambda}{2} \right\}^2 + \frac{\lambda}{4}(z + 1)^2(z + \mu)^2,$$

we can again check that  $k_\mu(z) = h_\mu(z)$  for all  $z$ , using a computer algebra system.

Note that

$$\lambda > 0 \quad \text{for } \mu < 1 + \sqrt{8}, \quad \lambda = 0 \quad \text{for } \mu = 1 + \sqrt{8}, \quad \text{and} \quad \lambda < 0 \quad \text{for } \mu > 1 + \sqrt{8}.$$

This means that for  $\mu < 1 + \sqrt{8}$ ,  $h_\mu$  is positive definite ( $h_\mu(z) > 0$  for all  $z$ ), so cannot have any roots, i.e.,  $g_\mu(x) = 0$  has no solution, so  $L_\mu$  cannot have any 3-cycles. We summarize this as follows:

**Theorem 2.6.2 [44]** (i) *If  $0 < \mu < 1 + \sqrt{8}$ , then  $h_\mu(z)$  is positive definite and the equation  $h_\mu(z) = 0$  does not have any real roots. Consequently, the logistic map  $L_\mu(x)$  does not have a 3-cycle.*

(ii) *If  $\mu = 1 + \sqrt{8}$ , then  $h_\mu(z) = 0$  has three distinct roots, each of multiplicity two. These three roots constitute a 3-cycle for  $L_\mu(x)$ .*

(iii) *If  $\mu > 1 + \sqrt{8}$  (with  $\mu - (1 + \sqrt{8})$  sufficiently small), the equation  $h_\mu(z) = 0$  has six simple roots which give rise to two 3-cycles for  $L_\mu(x)$ .*

**Proof.** (i) If  $\mu < 1 + \sqrt{8}$ , then since  $h_\mu(z)$  is positive definite ( $h_\mu(z) > 0$  for all  $z$ ), the result follows.

(ii) If  $\mu = 1 + \sqrt{8}$ , then  $\lambda = 0$  and the equation becomes

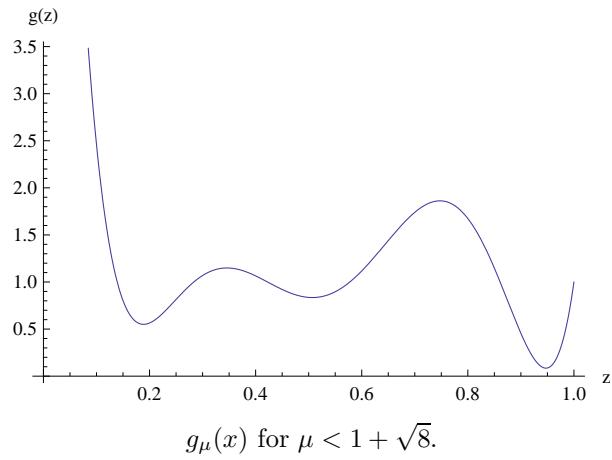
$$h_\mu(z) = \left( z^3 + (2 + 3\sqrt{2})z^2 + (5 + 4\sqrt{2})z + 3 + \sqrt{2} \right)^2.$$

The resulting cubic can be solved using the cubic formula to give three real solutions,  $z_1, z_2, z_3$ , and these can be used to give the three solutions to  $g_\mu(x) = 0$ , corresponding to the 3-cycle of  $L_\mu(x)$ :

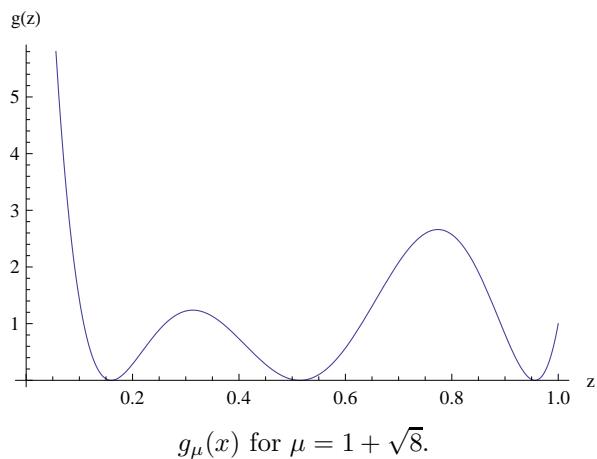
$$z_k = \frac{2\sqrt{7}}{3} \cos \left( \frac{1}{3} \arccos \left( -\frac{1}{2\sqrt{7}} + \frac{2k\pi}{3} \right) \right) - \frac{2 + 3\sqrt{2}}{3}, \quad k = 0, 1, 2,$$

(see the graph of  $g_\mu(x)$  below).

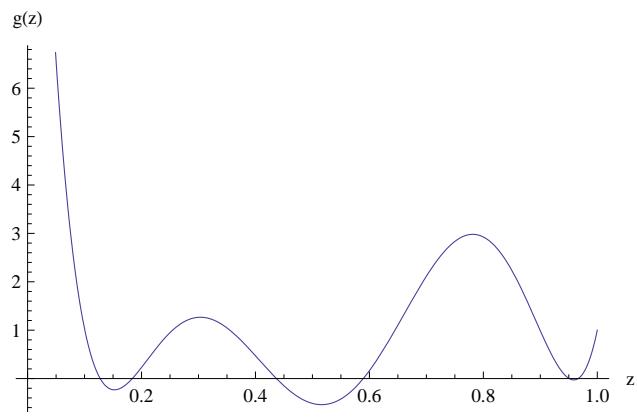
(iii) For  $\lambda < 0$  we can factor  $h_\mu(z) = h_1(z)h_2(z)$  using the difference of two squares, and then use the Intermediate Value Theorem on each of  $h_1(z)$  and  $h_2(z)$ , to see that they each have three different roots corresponding to two 3-cycles. These can be shown to be distinct (see [44] for details). □



$g_\mu(x)$  for  $\mu < 1 + \sqrt{8}$ .



$g_\mu(x)$  for  $\mu = 1 + \sqrt{8}$ .



$g_\mu(x)$  for  $\mu > 1 + \sqrt{8}$ .

### 2.6.3 A Super-Attracting 3-Cycle for the Logistic Map.

Recall that super-attracting periodic points occur where the derivative is zero. For a super-attracting 3-cycle  $\{c_1, c_2, c_3\}$ , we require

$$(L_\mu^3)'(c_1) = L'_\mu(c_1)L'_\mu(c_2)L'_\mu(c_3) = 0,$$

i.e.,

$$(1 - 2c_1)(1 - 2c_2)(1 - 2c_3) = 0,$$

so we may assume  $c_1 = 1/2$ . This means that  $x = 1/2$  is a solution of the equation  $L_\mu^3(x) = x$ , or  $L_\mu^3(1/2) = 1/2$ . But if  $\mu$  satisfies the equation  $L_\mu(1/2) = 1/2$ , then it will also satisfy the equation involving the third iterate. Consequently, we solve for  $\mu$ :  $g_\mu(1/2) = 1/2$ , where  $g_\mu(x) = (L_\mu^3(x) - x)/(L_\mu(x) - x)$  as defined in the last section, (this eliminates the root  $\mu = 2$  which gave rise to the super-attracting fixed point at  $x = 1/2$ ). We obtain

$$\frac{1}{64}(64 + 32\mu + 16\mu^2 - 24\mu^3 - 4\mu^4 + 6\mu^5 - \mu^6) = 0.$$

Set

$$p(a) = a^6 - 6a^5 + 4a^4 + 24a^3 - 16a^2 - 32a - 64,$$

then a computer algebra system indicates that there is a single real root  $\mu_0$  larger than  $1 + \sqrt{8}$  with exact value

$$\mu_0 = \frac{1}{6} \left\{ 6 + 2\sqrt{3 \left( 11 + 2 \cdot 2^{2/3} (25 - 3\sqrt{69})^{1/3} + 2 \cdot 2^{2/3} (25 + 3\sqrt{69})^{1/3} \right)} \right\}.$$

A computer algebra system is able to solve this equation exactly because it can be reduced to a cubic (and then the cubic formula may be used).

Following Lee [82], replace  $a$  by  $a + 1$  and check that

$$p(a + 1) = a^6 - 11a^4 + 35a^2 - 89 = b^3 - 11b^2 + 35b - 89.$$

$p(a + 1)$  is a cubic in  $b = a^2$  which can now be solved exactly for  $b$ , and then for  $a$ , from which the original equation can be solved. It is seen that

$$\mu_0 = 3.8318740552\dots$$

The other period 3-points may now be found since  $c_1 = 1/2$ ,  $c_2 = L_{\mu_0}(1/2) = \mu_0/4 = 0.95796\dots$ , and  $c_3 = L_{\mu_0}^2(1/2) = \mu_0^2/4(1 - \mu_0/4) = 0.15248\dots$

### 2.6.4: The 3-Cycle when $\mu = 4$

When  $\mu = 4$ ,  $L_4(x) = 4x(1 - x)$  maps  $[0, 1]$  onto all of  $[0, 1]$ . We saw in Chapter 1 that when  $\mu = 4$ , it is possible to get a closed formula for the iterates  $L_4^n$ , and for related reasons we can get quite explicit expressions for the  $n$ -cycles of this map.

In this case

$$g_4(x) = 4096x^6 - 13312x^5 + 16640x^4 - 10048x^3 + 3024x^2 - 420x + 21,$$

and we can check that (see also Lee [83])

$$\begin{aligned} g_4(x/4) &= x^6 - 13x^5 + 65x^4 - 157x^3 + 189x^2 - 105x + 21 \\ &= (x^3 - 7x^2 + 14x - 7)(x^3 - 6x^2 + 9x - 3), \end{aligned}$$

the product of two cubics. The solutions give a pair of 3-cycles which may be determined using the cubic formula. We use a different method to show they are given by the following concise formulas:

**Theorem 2.6.5** *For the logistic map  $L_4(x) = 4x(1 - x)$ :*

- (i) *The 2-cycle is  $\{\sin^2(\pi/5), \sin^2(2\pi/5)\}$ .*
  - (ii) *The 3-cycles are*
- $\{\sin^2(\pi/7), \sin^2(2\pi/7), \sin^2(3\pi/7)\}$  and  $\{\sin^2(\pi/9), \sin^2(2\pi/9), \sin^2(4\pi/9)\}$ .

**Proof.** Recall from Exercises 1.1 # 3, the difference equation  $x_{n+1} = 4x_n(1 - x_n)$ ,  $n = 0, 1, 2, \dots$ , has the solution

$$x_n = \sin^2(2^n \arcsin \sqrt{x_0}).$$

This was obtained by setting  $x_n = \sin^2(\theta_n)$  for some  $\theta_n \in (0, \pi/2]$ , ( $n = 1, 2, \dots$ ), so that

$$\sin^2(\theta_{n+1}) = 4 \sin^2(\theta_n)(1 - \sin^2(\theta_n)) = 4 \sin^2(\theta_n) \cos^2(\theta_n) = \sin^2(2\theta_n).$$

Using this formula we can now show that  $\sin^2(\theta_{n+1}) = \sin^2(4\theta_{n-1})$ , so in general

$$x_n = \sin^2(\theta_n) = \sin^2(2^n \theta_0), \quad \text{where } \theta_0 = \arcsin(\sqrt{x_0}).$$

In particular, we have

$$\theta_1 = \arcsin\left(\sqrt{\sin^2 2\theta_0}\right) = \begin{cases} 2\theta_0; & 0 \leq \theta_0 \leq \pi/4 \\ \pi - 2\theta_0; & \pi/4 \leq \theta_0 \leq \pi/2. \end{cases}$$

In the situation where we have a 2-cycle  $\{c_0, c_1\}$ , if  $c_i = \sin^2(\theta_i)$ , we get  $\theta_0$  is equal to  $4\theta_0, \pi - 2\theta_0, 2\pi - 4\theta_0$  or  $\pi - 4\theta_0$ . This gives  $\theta = 0$  or  $\theta = \pi/3$  (giving rise to the two fixed points) or  $\theta_0 = \pi/5$ , or  $2\pi/5$  from which the result follows. A similar analysis gives the 3-cycles (see below).

□

**Remark 2.6.6** Of course, it is easy to check directly that the above 2-cycle and 3-cycles are as we claim, but our theorem shows where these results come from.

To find the period  $n$ -points of  $L_4$ , we can use the above ideas by solving the equation  $L_4^n(x) = x$ . When  $x = \sin^2(\theta)$ , we have

$$\sin^2(\theta) = \sin^2(2^n\theta).$$

This gives rise to the two equations

$$\pm\theta = 2^n\theta + 2k\pi, \quad \text{or} \quad \pm\theta = (2k+1)\pi - 2^n\theta, \quad \text{for some } k \in \mathbb{Z},$$

and these can be summarized as a single equation:

$$\pm\theta = 2^n\theta + k\pi \Rightarrow \theta = \frac{k\pi}{2^n \pm 1}, \quad n = 1, 2, 3, \dots, k \in \mathbb{Z}$$

so that

$$\text{Per}_n(L_4) = \left\{ \sin^2\left(\frac{k\pi}{2^n - 1}\right) : 0 \leq k < 2^{n-1} \right\} \cup \left\{ \sin^2\left(\frac{k\pi}{2^n + 1}\right) : 0 < k \leq 2^{n-1} \right\}.$$

It follows that  $L_4$  has points of all possible periods. We shall see that the set of all periodic points of  $L_4$  constitutes a “dense” subset of  $[0, 1]$ , and each periodic point is unstable.

## Exercises 2.6

1. Show, by direct substitution, that  $\{\sin^2(\pi/5), \sin^2(2\pi/5)\}$  is a 2-cycle for the logistic map  $L_4(x) = 4x(1-x)$ . Similarly, show that  $\{\sin^2(\pi/7), \sin^2(2\pi/7), \sin^2(3\pi/7)\}$  and  $\{\sin^2(\pi/9), \sin^2(2\pi/9), \sin^2(4\pi/9)\}$  are 3-cycles for this map.
2. Use Remark 2.6.6 to write down the 4-cycles and 5-cycles of  $L_4$ .
3. Recall the family of maps defined by  $S_\mu(x) = \mu \sin(x)$  for  $x \in [0, \pi]$  and  $\mu \in [0, \pi]$ . Use a computer algebra system to estimate the values of  $\mu$  where periods two and three are created.

4. Modify the bifurcation diagram of the logistic family to give a bifurcation diagram for the family  $S_\mu$ ,  $0 \leq \mu \leq \pi$ .
5. Do the same as in exercise 3 for the family  $C_\mu(x) = \mu \cos(x)$ ,  $x \in [-\pi, \pi]$  and  $\mu \in [0, \pi]$ .
6. Use the method of exercise 3 to estimate a value of  $\mu$  for which  $S_\mu$  has a super-attracting 2-cycle, or 3-cycle?
7. Use a computer algebra system to show that the 2-cycle and 3-cycles from Theorem 2.6.5 are unstable. Can you use this information to conjecture a formula for  $|(L_4^n)'(x_n)|$ , where  $x_n$  is a point of period  $n$ ?
8. Let  $g_\mu(x) = \mu x \frac{(1-x)}{(1+x)}$ ,  $\mu > 0$ .
- (a) Show that  $g_\mu$  has a maximum at  $x = \sqrt{2} - 1$  and the maximum value is  $\mu(3 - 2\sqrt{2})$ .
  - (b) Deduce that  $g_\mu$  is a dynamical system on  $[0, 1]$  for  $0 \leq \mu \leq 3 + 2\sqrt{2}$  (i.e.,  $g_\mu([0, 1]) \subseteq [0, 1]$ ).
  - (c) Find the fixed points of  $g_\mu$  for  $\mu \geq 1$ .
  - (d) Give conditions on  $\mu$  for the fixed points of  $g_\mu$  to be attracting.
  - (e) Use a computer algebra system to graph  $g_\mu^2$  and  $g_\mu^3$ , and estimate when a period 2-point is created.
  - (f) Use a computer algebra system to give a bifurcation diagram for  $g_\mu$ , for  $0 \leq \mu \leq 3 + 2\sqrt{2}$ .
9. (a) Use the results of Section 2.6 to show that the logistic map  $L_4(x) = 4x(1-x)$  cannot have a super-attracting cycle.
- (b) Find a point  $x_0 \in (0, 1)$  which is not a periodic point for  $L_4$ .

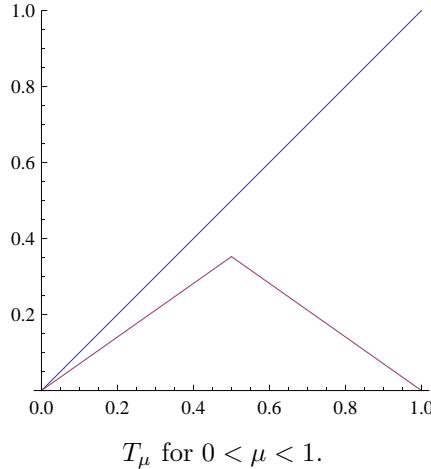
## 2.7 The Tent Family $T_\mu$ .

The *tent family*  $T_\mu$  is a parameterized family related to the family of logistic maps, and is defined in a piecewise linear manner:  $T_\mu : [0, 1] \rightarrow [0, 1]$ ,  $0 < \mu \leq 2$ ,

$$T_\mu(x) = \begin{cases} \mu x; & 0 \leq x \leq 1/2 \\ \mu(1 - x); & 1/2 < x \leq 1. \end{cases}$$

When  $\mu = 2$  we get the familiar tent map  $T = T_2$  seen earlier. We now look at the parameter values  $0 < \mu \leq 2$ , (and later we shall examine the situation where  $\mu > 2$ , and  $T_\mu : \mathbb{R} \rightarrow \mathbb{R}$ ).

If  $0 < \mu < 1$ , we see that the only fixed point is  $c = 0$ . If  $\mu = 1$ , then all  $c \in [0, 1/2]$  are fixed points, and if  $1 < \mu \leq 2$ , then there are 2 fixed points. We look at each of these cases separately.



$T_\mu$  for  $0 < \mu < 1$ .

**Case 2.7.1** If  $0 < \mu < 1$ , we see that 0 is the only fixed point of  $T_\mu$ . For  $0 < x \leq 1/2$ ,  $0 \leq T_\mu(x) = \mu x < x$ , and if  $1/2 < x \leq 1$ , then

$$0 \leq T_\mu(x) = \mu(1 - x) < 1 - x \leq \frac{1}{2} < x,$$

and continuing in this way, we see the sequence  $(T_\mu^n(x))$  is decreasing and bounded below by 0, so must converge to the fixed point 0. Thus  $B_{T_\mu}(0) = [0, 1]$ .

**Case 2.7.2** If  $\mu = 1$  and  $0 < x \leq 1/2$ , then  $T_1(x) = x$ , so  $x$  is a fixed point. If  $1/2 < x \leq 1$ , then  $x$  is an eventual fixed point since

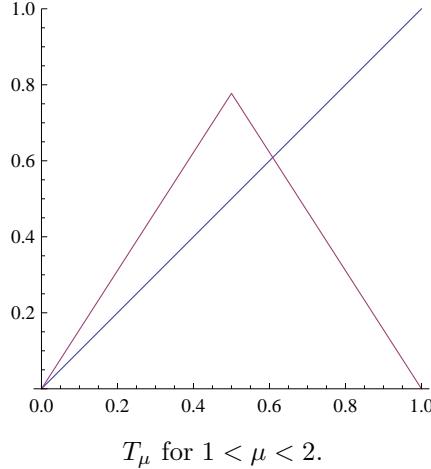
$$T_1^2(x) = T_1(1 - x) = x.$$

The fixed points are stable but not attracting.

**Case 2.7.3** If  $1 < \mu < 2$ , there is second fixed point  $c$  which is found by solving  $T_\mu(x) = x$ :

$$T_\mu(x) = \mu(1 - x) = x, \quad \text{so that} \quad c = \frac{\mu}{1 + \mu}.$$

Since  $|T'_\mu(x)| = \mu > 1$ , the fixed point is repelling.



$T_\mu$  for  $1 < \mu < 2$ .

**Case 2.7.4** If  $\mu = 2$ , we have  $T_2 = T$ , the familiar tent map. The fixed points 0 and  $2/3$  are repelling. The range of  $T$  is all of  $[0, 1]$ .  $T$  has the effect of mapping the interval  $[0, 1/2]$  onto all of  $[0, 1]$ , and then folding the interval  $[1/2, 1]$  back over the interval  $[0, 1]$ . It is this stretching and folding that gives rise to the chaotic nature of  $T$  that we will examine later. Complicated dynamics for many transformations typically arises in this way. We will examine the periodic points of  $T$  in the next section.

## 2.8 The 2-Cycles and 3-Cycles of the Tent Family.

At some stage for  $\mu \geq 1$  a 2-cycle is created. It is interesting to use a computer algebra system, with a dynamic iteration of  $T_\mu$  to see how this happens. It can be checked that for  $\mu > 1$ ,  $T_\mu^2$  is given by the formula

$$T_\mu^2(x) = \begin{cases} \mu^2 x; & 0 \leq x \leq \frac{1}{2\mu}, \\ \mu(1 - \mu x); & \frac{1}{2\mu} < x \leq \frac{1}{2}, \\ \mu(1 - \mu + \mu x); & \frac{1}{2} < x \leq 1 - \frac{1}{2\mu}, \\ \mu^2(1 - x); & 1 - \frac{1}{2\mu} < x \leq 1. \end{cases}$$

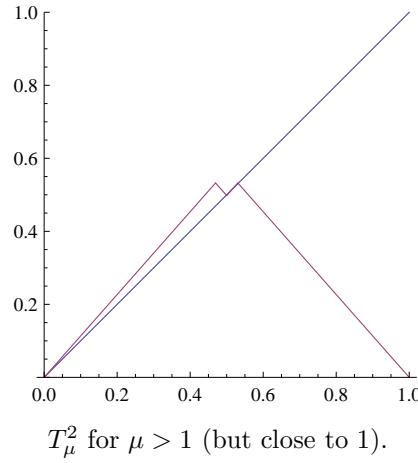
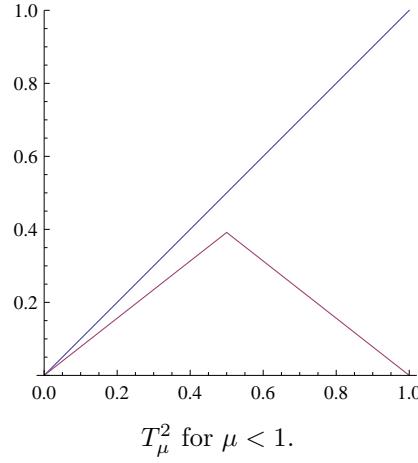
The 2-cycle is created when  $T_\mu^2(1/2) = 1/2$ , i.e., when

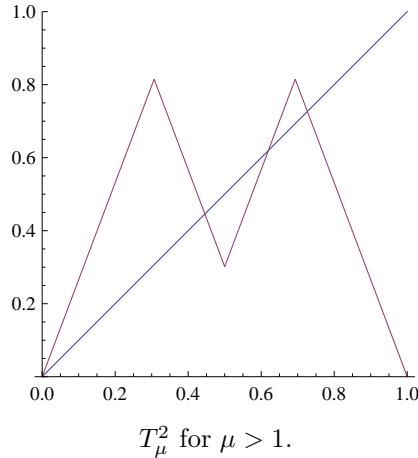
$$\mu(1 - \mu/2) = 1/2, \quad \text{or} \quad (\mu - 1)^2 = 0, \quad \mu = 1,$$

so the 2-cycle appears when  $\mu > 1$ . For  $\mu \leq 1$ , there are no period 2 points. The 2-cycle  $\{c_1, c_2\}$  say, is unstable since

$$|(T_\mu^2)'(c_1)| = |T'_\mu(c_1)T'_\mu(c_2)| = \mu^2 > 1.$$

For example, if we solve  $\mu(1 - \mu x) = x$ , we get  $c_1 = \frac{\mu}{1 + \mu^2}$ , and solving  $\mu^2(1 - x) = x$  gives  $c_2 = \frac{\mu^2}{1 + \mu^2}$ , as the period 2-points. Solving  $\mu(1 - \mu + \mu x) = x$  gives  $c = \frac{\mu}{1 + \mu}$  as the (non-zero) fixed point.

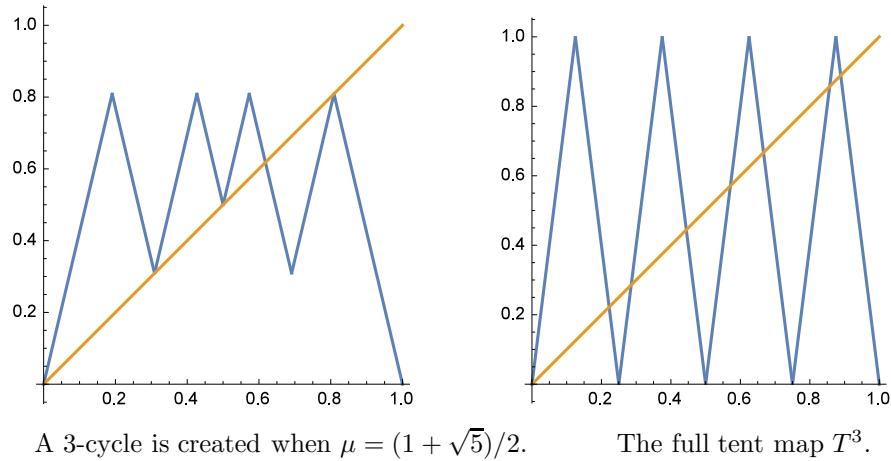




We now look for the smallest value of  $\mu$  for which there exists a 3-cycle. In a similar way to the situation for the 2-cycle, we look for where  $T_\mu^3(1/2) = 1/2$ , and we compute the value of the largest root of this equation. In general, (see Heidel [67]) the smallest value of  $\mu$  for which there exists a periodic orbit of period  $k$  is precisely the value of  $\mu$  for which  $1/2$  has period  $k$ . Using the formula for  $T_\mu^2(1/2) = \mu(1 - \mu/2)$  above, since  $\mu(1 - \mu/2) \leq 1/2$  for all  $\mu$ , we get

$$T_\mu^3(1/2) = \mu^2(1 - \mu/2) = 1/2 \quad \text{when} \quad \mu^3 - 2\mu^2 + 1 = 0,$$

or  $(\mu - 1)(\mu^2 - \mu - 1) = 0$ . Disregarding  $\mu = 1$  and solving the quadratic gives  $\mu = (1 + \sqrt{5})/2$  as the value of  $\mu$  where period 3 first occurs.



In general, it can be shown that for  $k > 3$  odd, period  $k$  first occurs when  $\mu$  is equal to the largest real root of the equation

$$\begin{aligned} \mu^k - 2\mu^{k-1} + 2\mu^{k-3} - 2\mu^{k-4} + \dots - 2\mu + 1 \\ = (\mu - 1)(\mu^{k-1} - \mu^{k-2} - \mu^{k-3} + \mu^{k-4} - \mu^{k-5} + \mu^{k-6} \dots + \mu - 1) = 0, \end{aligned}$$

which has  $\mu - 1$  as a factor (see [67]). We will return to look at the tent family in Section 6.4.

### Exercises 2.8

1. For  $\mu > 1$ , show that  $T_\mu$  has no attracting periodic points.
2. Find  $\mu \in [0, 2]$  such that  $c = 1/2$  is a point of period 3 under the tent map  $T_\mu$  (so that  $\{1/2, \mu/2, \mu(1 - \mu/2)\}$  is a 3-cycle).
3. Show that
$$\left\{ \frac{\mu}{1 + \mu^3}, \frac{\mu^2}{1 + \mu^3}, \frac{\mu}{1 + \mu^3} \right\},$$
is a 3-cycle for  $T_\mu$ , when  $\mu \geq (1 + \sqrt{5})/2$ .
4. Show that
$$\left\{ \frac{\mu}{1 + \mu + \mu^2 + \mu^3}, \frac{\mu^2}{1 + \mu + \mu^2 + \mu^3}, \frac{\mu^3}{1 + \mu + \mu^2 + \mu^3}, \frac{\mu + \mu^2 + \mu^3}{1 + \mu + \mu^2 + \mu^3} \right\},$$
is a 4-cycle for  $T_\mu$ , when  $\mu$  satisfies  $\mu^3 - \mu^2 - \mu - 1 \geq 0$  ( $\mu \geq 1.83929\dots$ , approximately).
5. If  $L_\mu(x) = \mu x(1 - x)$  is such that  $c = 1/2$  is of period  $n$ , for some  $n \in \mathbb{Z}^+$ , prove that  $1/2$  is an attracting periodic point. Is it necessarily a super-attracting periodic point?
6. Modify your bifurcation diagram of the logistic family to give a bifurcation diagram for the tent family  $T_\mu$ ,  $x \in [0, 1]$  and  $\mu \in [0, 2]$ .



## CHAPTER 3

### Sharkovsky's Theorem.

Sharkovsky's Theorem is one of the most remarkable theorems of the 20th century. One hundred years ago, mathematicians were inclined to think that most of what can be shown concerning continuous functions on intervals was known. They were wrong! In this chapter, we shall state Sharkovsky's Theorem and prove a special case, sometimes called the Li-Yorke Theorem. The proof of Sharkovsky's Theorem is given in Chapter 12, which also looks at a precursor of Sharkovsky's Theorem due to Coppel. We have seen in Chapters 1 and 2, various situations where maps have fixed points, 2-cycles and points of higher period. In this chapter, we will see how having points of period three implies a considerable degree of complexity in the map as a dynamical system. Maps having only 2-cycles generally do not have such complexity.

#### 3.1 Period Three Implies Chaos.

In 1975 in a paper entitled “Period three implies chaos”, Tien-Yien Li and James Yorke proved a surprising theorem ([84]):

**Theorem 3.1.1** *Let  $f : I \rightarrow I$  be a continuous function defined on an interval  $I \subseteq \mathbb{R}$ . If  $f(x)$  has a point of period three, then for any  $k = 1, 2, 3, \dots$ , there is a point having period  $k$ .*

This paper stirred considerable interest in the mathematical community. Shortly after it appeared, it was pointed out by P. Štefan in [121], that a Ukrainian mathematician by the name of Sharkovsky had, in 1964, published a much more general theorem (in Russian), in a Ukrainian journal. His theorem was unknown in the west until the appearance of the Li-Yorke Theorem. To state his theorem we need to define a new ordering of the positive integers  $\mathbb{Z}^+$ . In the “Sharkovsky ordering”, 3 is the largest number, followed by 5 then 7 (all of the odd integers), then  $2 \cdot 3$ ,  $2 \cdot 5$ , (2 times the odd integers), then  $2^2$  times the odd integers etc., finishing off with powers of 2 in descending order:

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

Sharkovsky's Theorem says that if a continuous map has a point of period  $k$ , then it has points of all periods less than  $k$  in the Sharkovsky ordering. The converse, which is also due to Sharkovsky, is true in the sense that, for each  $k \in \mathbb{Z}^+$  there is a continuous map having points of period  $k$ , but no points of period larger than  $k$  in the Sharkovsky ordering.

**Theorem 3.1.2 (Sharkovsky's Theorem, 1964.)** *Let  $f : I \rightarrow I$  be a continuous map on an interval  $I$  (where  $I$  may be any bounded or unbounded subinterval of  $\mathbb{R}$ ). If  $f$  has a point of period  $k$ , then it has points of period  $r$  for all  $r \in \mathbb{Z}^+$  with  $k \triangleright r$ .*

For example, this theorem tells us that if  $f$  has a 4-cycle, then it also has a 2-cycle and a 1-cycle (fixed point). If  $f$  has a 3-cycle, then  $f$  has all other possible cycles. If  $f$  has a 6-cycle, then since  $6 = 2 \cdot 3$ ,  $f$  will have  $2 \cdot 5$ ,  $2 \cdot 7$ ,  $\dots$   $2^2 \cdot 3$ ,  $2^2 \cdot 5$ ,  $\dots$   $2^2$ , 2, 1-cycles.

In this chapter we shall prove the theorem for  $k = 3$  (often known as the Li-Yorke Theorem because of their independent proof of this result in 1975). The proof of the general case is presented in Chapter 12. Although the proof is elementary in that it uses little more than the Intermediate Value Theorem, it is still a very difficult theorem, and Sharkovsky's original proof is very technical and hard to follow. A number of new proofs have appeared over the years, the first due to P. Štefan [121], who died at an early age in a climbing accident. We shall give a proof due to Bau-Sen Du ([38], [39] and [40]). Another recent proof appears in [25].

James Yorke is a professor at the University of Maryland, College Park and Li was his graduate student. A few years ago, Yorke and Benoit Mandelbrot were awarded the Japan Prize (the Japanese equivalent of the Nobel Prize) for their work in Dynamical Systems and Fractals. Mandelbrot is regarded as the “father” of fractals, and we will review some of his work later in this book. In order to prove the Li-Yorke Theorem, we shall need some preliminary lemmas. The first of these was proved in Chapter 1 (Theorem 1.2.9).

**Lemma 3.1.3** *Let  $f : I \rightarrow \mathbb{R}$  be a continuous map, where  $I$  an interval with  $J = f(I) \supseteq I$ . Then  $f(x)$  has a fixed point in  $I$ .*

**Lemma 3.1.4** *Let  $f : I \rightarrow \mathbb{R}$  be a continuous map. If  $J \subseteq f(I)$  is a closed bounded interval, then there exists a closed bounded interval  $K \subseteq I$  with  $f(K) = J$ .*

**Proof.** Write  $J = [a, b]$  for some  $a, b \in \mathbb{R}$ ,  $a < b$ . There exist  $p, q \in I$  with  $f(p) = a$  and  $f(q) = b$ . The idea of the proof is to choose  $p$  and  $q$  as close together as possible (with no points  $c$  between  $p$  and  $q$  with  $f(c) = a$ , or  $f(c) = b$ ), and then conclude that  $f([p, q]) = J$  or  $f([q, p]) = J$ .

Suppose that  $p < q$  and let  $\alpha$  be the point of  $I$  that is closest to  $q$  with the property that  $f(\alpha) = a$  (i.e.,  $\alpha = \max\{x : p \leq x \leq q, f(x) = a\}$ ). Similarly, take  $\beta$  to be that point between  $\alpha$  and  $q$  with the property that  $\beta$  is the closest to  $\alpha$  with  $f(\beta) = b$  (i.e.,  $\beta = \min\{x : \alpha \leq x \leq q, f(x) = b\}$ ).

On the other hand, if  $p > q$ , set  $\alpha = \max\{x : q \leq x \leq p, f(x) = b\}$  and  $\beta = \min\{x : \alpha \leq x \leq p, f(x) = a\}$ . In both cases, set  $K = [\alpha, \beta]$ .

Using the fact that the continuous image of an interval of the form  $K = [\alpha, \beta]$  is again an interval of this form, we will show that  $f(K) = J$ .

Suppose this is not the case, then since  $a, b \in f(K)$ , we have  $[a, b] \subset f(K)$ . Let  $w \in f(K) \setminus J$ , then  $w = f(z)$  for some  $z \in K$ . If  $w > b$ , then by the Intermediate Value Theorem there exists  $c$  between  $\alpha$  and  $\beta$  with  $f(c) = b$ , contradicting our choice of  $\beta$ . Similarly for the case where  $w < a$ .

□

### 3.1.5 Proof of Sharkovsky's Theorem for $k = 3$ .

We are assuming that  $f$  has a point of period 3, so there is a 3-cycle  $\{a, b, c\}$ ,  $a < b < c$ . Note that 3-cycles comes in two versions that are mirror images. These are  $f(a) = b$ ,  $f(b) = c$  and  $f(c) = a$ , and  $f(a) = c$ ,  $f(c) = b$  and  $f(b) = a$ . Consequently, we need only treat the first version.

We give the idea of the proof by showing why there must be points of period one, two and four. Let

$$[a, b] = L_0 \quad \text{and} \quad [b, c] = L_1.$$

Observe that

$$f(L_0) \supseteq L_1 \quad \text{and} \quad f(L_1) \supseteq L_0 \cup L_1.$$

#### Case 1. $f$ has a fixed point.

Since

$$f(L_1) \supseteq L_0 \cup L_1 \supseteq L_1,$$

Lemma 3.1.3 implies that  $f$  has a fixed point in  $L_1$ .

#### Case 2. $f$ has a point of period 2.

This time we use

$$f(L_1) \supseteq L_0 \cup L_1 \supseteq L_0,$$

and by Lemma 3.1.4 there is an interval  $B \subseteq L_1$  such that  $f(B) = L_0$ . We then have

$$f^2(B) = f(L_0) \supseteq L_1 \supseteq B,$$

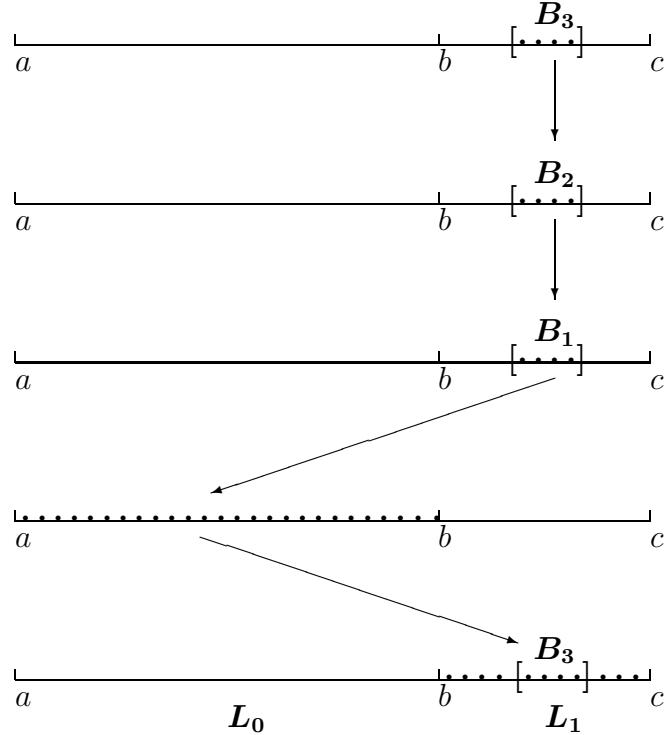
and by Lemma 3.1.3,  $B$  contains a fixed point  $c$  of  $f^2$ .  $c$  is a period 2-point of  $f$  (and not a fixed point of  $f$ ) because

$$f(c) \in L_0 \quad \text{and} \quad c \in L_1, \quad \text{so} \quad f(c) \neq c.$$

### **Case 3. $f$ has a point of period 4.**

The above two constructions do not illustrate the general method, but the following construction is easily generalized to show that there exists fixed points of any period greater than 3. Our aim is to show that there is an interval  $B$  contained in  $L_1$ , which is mapped first by  $f$  into  $L_1$ , then into  $L_1$  again, then onto  $L_0$  and then onto  $L_1$ , so that  $f^4(B) \supseteq B$ . Thus  $f^4$  has a fixed point  $c$  in  $B$ , which cannot be a point of lesser period because  $f(c) \in L_1$ ,  $f^2(c) \in L_1$ ,  $f^3(c) \in L_0$  and  $f^4(c) \in B$  (so cannot have  $f(c) = c$ ,  $f^2(c) = c$  or  $f^3(c) = c$ ).

It is useful to think of 5 copies of  $L_0 \cup L_1$  with  $f$  mapping the first to the second etc. as shown:



Five copies of  $L_0 \cup L_1$ , where  $f(a) = b$ ,  $f(b) = c$  and  $f(c) = a$ .

We find intervals  $B_1$ ,  $B_2$  and  $B_3$  as follows:

$$f(L_0) \supseteq L_1, \quad f(L_1) \supseteq L_0 \cup L_1,$$

so there exists  $B_1 \subseteq L_1$  such that  $f(B_1) = L_0$ .

There exists  $B_2 \subseteq L_1$  such that  $f(B_2) = B_1$ , and there exists  $B_3 \subseteq L_1$  such that  $f(B_3) = B_2$ . Set  $B = B_3$ , then

$$f^2(B_3) = f(B_2) = B_1, \quad \text{and so} \quad f^3(B) = L_0, \quad f^4(B) \supseteq L_1 \supseteq B_3.$$

In other words  $f^4(B) \supseteq B$ , so there exists  $c \in B$ , a fixed point of  $f^4$ , which is not a point of period 3 or less, so must be a point of period 4. □

In general, if a function has points of period 4, the most we can deduce is that there are points of period 2 and fixed points. However, the following is true:

**Proposition 3.1.6** *If  $f : I \rightarrow I$  is continuous on an interval  $I$  with*

$$f(a) = b, \quad f(b) = c, \quad f(c) = d, \quad f(d) = a, \quad a < b < c < d,$$

then  $f(x)$  has a point of period 3, so  $f$  also has points of all other periods.

**Proof.** We may assume that

$$f[a, b] = [b, c], \quad f[b, c] = [c, d], \quad f[c, d] = [a, d].$$

In particular, there exists an interval  $B_1 \subseteq [c, d]$  with  $f(B_1) = [c, d]$ , and an interval  $B_2 \subseteq [c, d]$  with  $f(B_2) = [b, c]$ .

Again, we can find an interval  $K_1 \subseteq B_1$  with  $f(K_1) = B_2$ , so that

$$f^3(K_1) = f^2(B_2) = f[b, c] = [c, d] \supseteq K_1,$$

and  $f^3$  has a fixed point in  $K_1$  which is not a fixed point of  $f(x)$ . □

The above proofs can be summarized with the following type of result:

**Proposition 3.1.7** *Let  $I$  be an interval and  $f : I \rightarrow I$  be a continuous map. Let  $I_1$  and  $I_2$  be closed subintervals of  $I$  with at most one point in common. If  $f(I_1) \supset I_2$  and  $f(I_2) \supset I_1 \cup I_2$ , then  $f$  has a 3-cycle.*

**Proof.** See Exercises 3.2. □

### 3.2 Converse of Sharkovsky's Theorem.

As we mentioned, for each  $m \in \mathbb{Z}^+$  in the Sharkovsky ordering of  $\mathbb{Z}^+$ , Sharkovsky showed that there is a continuous map  $f : I \rightarrow I$  ( $I$  an interval), such that  $f(x)$  has a point of period  $m$ , but no point of period  $k$  for  $k \triangleright m$ . The following theorems were proved by Sharkovsky ( $I$  is either the real line or a subinterval):

**Theorem 3.2.1** *For every  $k \in \mathbb{Z}^+$ , there exists a continuous map  $f : I \rightarrow I$  that has a  $k$ -cycle, but has no cycles of period  $n$  for any  $n$  appearing before  $k$  in the Sharkovsky ordering.*

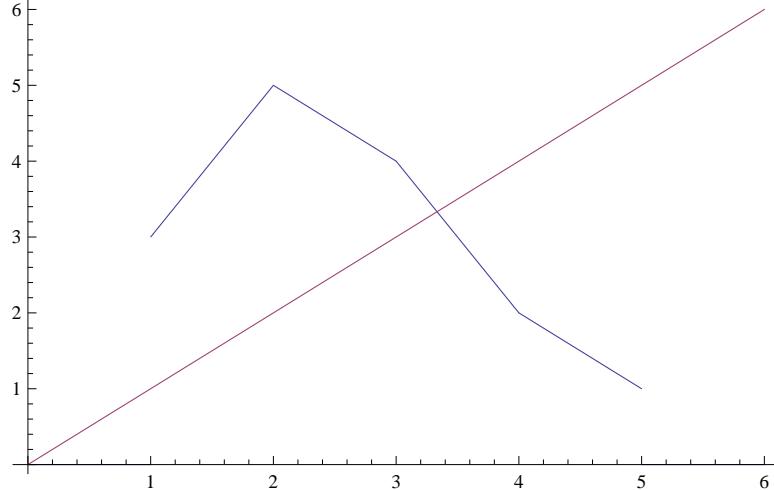
**Theorem 3.2.2** *There exists a continuous map  $f : I \rightarrow I$  that has a  $2^n$ -cycle, for every  $n \in \mathbb{Z}^+$ , and has no other cycles of any other period.*

Strictly speaking, Sharkovsky's Theorem is the combination of Theorem 3.1.2, Theorem 3.2.1 and Theorem 3.2.2 (see [92]). Sometimes the latter two theorems are referred to as the converse of Sharkovsky's Theorem (see [42]). We look at some particular cases of this:

**Examples 3.2.3** 1. Define a function  $f : [1, 5] \rightarrow [1, 5]$  as shown by the graph below (so

$$f(1) = 3, \quad f(2) = 5, \quad f(3) = 4, \quad f(4) = 2, \quad f(5) = 1,$$

with  $f(x)$  piecewise linear between these points). Then  $f$  has a point of period 5, but no points of period 3.



A function having points of period 5, but no points of period 3.

**Proof.** Clearly no member of the set  $\{1, 2, 3, 4, 5\}$  is of period 3, but this set is a 5-cycle. Theorem 1.2.7 tells us that  $f(x)$  has a fixed point  $c$ , so  $c$  is a fixed point for  $f^3$ . We shall show that  $f^3$  has no other fixed points. Suppose to the contrary that  $f^3$  has another fixed point  $\alpha$ . Now we can check that:

$$f^3[1, 2] = [2, 5], \quad f^3[2, 3] = [3, 5], \quad f^3[4, 5] = [1, 4],$$

so  $f^3$  cannot have a fixed point in the intervals  $[1, 2]$ ,  $[2, 3]$  or  $[4, 5]$ , so  $\alpha$  must lie in the interval  $[3, 4]$ . In fact,  $f^3[3, 4] = [1, 5] \supseteq [3, 4]$ , and we show that  $f^3$  cannot have another fixed point in  $[3, 4]$ .

If  $\alpha \in [3, 4]$ , then  $f(\alpha) \in [2, 4]$ , so either  $f(\alpha) \in [2, 3]$  or  $f(\alpha) \in [3, 4]$ . If the former holds, then  $f^2(\alpha) \in [4, 5]$  and  $f^3(\alpha) \in [1, 2]$ , which is impossible as we have to have  $f^3(\alpha) = \alpha \in [3, 4]$ .

Thus, we must have  $f(\alpha) \in [3, 4]$ , so that  $f^2(\alpha) \in [2, 4]$ . Again there are two possibilities: if  $f^2(\alpha) \in [2, 3]$ , then  $f^3(\alpha) \in [4, 5]$ , another contradiction, so that  $f^2(\alpha) \in [3, 4]$ .

We have shown that the orbit of  $\alpha$ :  $\{\alpha, f(\alpha), f^2(\alpha)\}$  is contained in the interval  $[3, 4]$ . On the interval  $[3, 4]$ , we can check that  $f(x)$  is given by the straight line

formula

$$f(x) = 10 - 2x, \quad \text{and} \quad f(10/3) = 10/3,$$

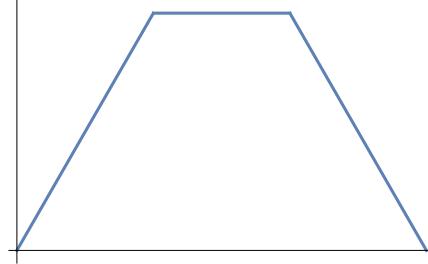
so  $c = 10/3$  is the unique fixed point of  $f$ . Also

$$f^2(x) = -10 + 4x, \quad f^3(x) = 30 - 8x,$$

also with the unique fixed point  $x = 10/3$ . It follows that  $f$  cannot have any points of period 3.

□

2. The converse of Sharkovsky's Theorem can also be demonstrated using a nice idea from [4]: Let  $T$  be the tent map and  $h \in [0, 1]$ . The *truncated tent map*  $T_h : [0, 1] \rightarrow [0, 1]$  is defined by  $T_h(x) = \min\{h, T(x)\}$ . In Exercises 3.2 # 14, examples are given with  $T_h$  having a fixed point, but no points of any other period, and  $T_h$  having a 2-cycle, but no 4-cycle. In Section 12.3 it is shown how to prove the converse of Sharkovsky's Theorem using these truncated tent maps.



A truncated tent map.

**Remark 3.2.4** Example 3.2.3 can be generalized in a natural way to give a map having points of period  $2n + 1$ , but no points of period  $2n - 1$ ,  $n = 2, 3, \dots$ .

### Exercises 3.2

1. By considering the function  $f(x) = -1 - 1/x$ , show that the condition of continuity is essential in Sharkovsky's Theorem.
  
  
  
2. Order the integers 25 to 35 inclusive, using Sharkovsky's ordering. Explain what your ordering means.

3. Assuming Sharkovsky's Theorem, show that

$$m \triangleright n \quad \text{if and only if} \quad 2m \triangleright 2n.$$

4. Use the ideas of Section 3.1 to show that if  $f : \mathbb{R} \rightarrow \mathbb{R}$  is continuous and has a 2-cycle  $\{a, b\}$ , then  $f$  has fixed point.
5. Show that the map  $f(x) = (x - 1/x)/2$ ,  $x \neq 0$ , has no fixed points but it has period 2-points. Find the 2-cycle, and by looking at the graph of  $f^3(x)$ , check to see whether or not it has a 3-cycle. Why does this not contradict Sharkovsky's Theorem?
6. A map  $f : [1, 7] \rightarrow [1, 7]$  is defined so that  $f(1) = 4, f(2) = 7, f(3) = 6, f(4) = 5, f(5) = 3, f(6) = 2, f(7) = 1$ , and the corresponding points are joined so the map is continuous and piecewise linear. Show that  $f$  has a 7-cycle but no 5-cycle.
7. Show that a continuous increasing function  $f : \mathbb{R} \rightarrow \mathbb{R}$  cannot have a 3-cycle. Can it have a 2-cycle? Answer the same questions when  $f$  is decreasing.
8. Write down the details of the proof of Theorem 3.1.7. (Hint: Use the ideas of the proof of Theorem 3.1.1).
9. Suppose  $f : \mathbb{R} \rightarrow \mathbb{R}$  has a 5-cycle  $\{a_1, a_2, a_3, a_4, a_5\}$ , where where  $f(a_i) = a_{i+1}$ ,  $i = 1, 2, 3, 4$  and  $f(a_5) = a_1$ .
- (a) Show that  $f$  always has a 7-cycle.
- (b) If  $a_1 < a_2 < a_3 < a_4 < a_5$ , show that  $f$  has a 3-cycle.

10. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Write down all the possibilities for a 4-cycle  $\{a, b, c, d\}$ ,  $a < b < c < d$  for  $f$ , (e.g.,  $f(a) = c$ ,  $f(c) = d$ ,  $f(d) = b$ , and  $f(b) = a$ )? Indicate which are mirror images, and which give rise to a 3-cycle.

11. Use Sharkovsky's Theorem to prove that if  $f : [a, b] \rightarrow [a, b]$  is a continuous function and  $\lim_{n \rightarrow \infty} f^n(x)$  exists for every  $x \in [a, b]$ , then  $f$  can have no points of period  $n > 1$ .

12\*. Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous map.

(a) If there exists  $c \in [a, b]$  with

$$f^2(c) < c < f(c),$$

show that  $f$  has a fixed point  $z > c$ , and a period-2 point  $y < c$ . (Note: The existence of  $y < c$  with  $f^2(y) = y$  is straightforward to establish. It must be shown that  $y$  can be chosen so as to be of period 2).

(b) If there exists  $c \in [a, b]$  with

$$f^3(c) < c < f(c),$$

show that  $f$  has a period-2 point.

13. For the function  $f : [1, 5] \rightarrow [1, 5]$  of Example 3.2.3, find an interval with end points consecutive integers, containing a period-2 point, and deduce the value of the period-2 point

14. Let  $T$  be the tent map and  $h \in [0, 1]$ . The *truncated tent map*  $T_h : [0, 1] \rightarrow [0, 1]$  is defined to be  $T_h(x) = \min\{h, T(x)\}$ .

(a) Show that if  $h = 2/3$ , then  $T_h$  has fixed points, but no period-2 points,

(b) Show that if  $h = 4/5$ , then  $T_h$  has a 2-cycle, but no 4-cycle.

## CHAPTER 4

### Dynamics on Metric Spaces.

In dynamical systems, a variety of examples acting on different sets, have similar properties. So far, we have seen examples where the set in question is a subinterval of  $\mathbb{R}$ . For other examples, the underlying set may be a subset of  $\mathbb{R}^2$ , or the unit circle in the complex plane, or a type of sequence space.

It is convenient to give some general results about dynamical systems which apply to every one of these situations, so avoiding unnecessary work when we meet a new example. In Chapter 6, we give a definition of chaos that will apply to examples acting on different types of metric space.

We try to keep the pathology of the underlying sets to a minimum. For this reason we restrict our attention to metric spaces, as our emphasis is on the dynamics rather than the set on which the action takes place. Metric spaces are in some sense the least abstract type of topological space. The examples that we give in this chapter are usually dynamical systems acting on *compact* metric spaces or at least *complete* metric spaces. The tools developed in this chapter are important throughout the remainder of this text, and will be built upon as we proceed through the following chapters. For example, we will define the important notions of completeness and compactness in metric spaces in Chapters 10 and 17, respectively.

#### 4.1 Basic Properties of Metric Spaces.

In what follows we introduce the idea of a *metric space* (see [117] for example). This is simply a pair  $(X, d)$  where  $X$  is a set and  $d$  is a *distance* defined on the set, called a *metric*.  $d$  must satisfy certain natural properties that one would expect of a distance function.

**Definition 4.1.1** Let  $X$  be a set with  $x, y, z \in X$ . A *metric* on  $X$  is a function  $d : X \times X \rightarrow [0, \infty)$  satisfying:

1.  $d(x, y) \geq 0$  for all  $x, y \in X$ ,
2.  $d(x, y) = 0$  if and only if  $x = y$ ,

3.  $d(x, y) = d(y, x)$ ,
4.  $d(x, y) \leq d(x, z) + d(z, y)$  (the *triangle inequality*).

**Definition 4.1.2** A pair  $(X, d)$  satisfying the conditions of Definition 4.1.1 is called a *metric space*, i.e., a metric space is just a set  $X$  with a distance function  $d$  satisfying certain natural conditions.

**Examples 4.1.3** 1.  $X = \mathbb{R}$  with  $d(x, y) = |x - y|$  for  $x, y \in \mathbb{R}$  is a metric space. Here the triangle inequality is the statement:

$$|x - y| \leq |x - z| + |z - y|, \quad x, y, z \in \mathbb{R}.$$

In a similar way,  $X = [0, 1]$  with the same distance function is a metric space.

2.  $X = \mathbb{R}^2$  with the usual distance in the plane defined by

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},$$

for  $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ , is a metric space. More generally,  $\mathbb{R}^n$  with its usual distance function and  $\mathbb{C}$ , the set of all complex numbers with  $d(z, w) = |z - w|$ ,  $z, w \in \mathbb{C}$ , are metric spaces.

3. Let  $X = \mathbb{S}^1$ , where  $\mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$  is the unit circle in the complex plane. The natural metric on  $\mathbb{S}^1$  is given by the (shortest) distance around the circle between the two points on the circle.

4. If  $X$  is any set and  $d(x, y) = \begin{cases} 0; & x = y \\ 1; & x \neq y \end{cases}$ , for  $x, y \in X$ , then  $d$  is a metric on  $X$ , called the *discrete metric*. We can use this to define a metric on the set  $X = \{0, 1\}$ , or on  $X = \mathbb{R}$ , different to the standard metric of Example 1.

5. Let  $\mathcal{A} = \{0, 1\}$  and  $n \in \mathbb{Z}^+$ . We write  $\mathcal{A}^n$  for the set

$$\mathcal{A}^n = \{(a_0, a_1, a_2, \dots, a_{n-1}) : a_i \in \{0, 1\}, 0 \leq i < n\},$$

the  $n$ -fold direct product of  $\mathcal{A}$  with itself. For example,  $\mathcal{A}^2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ .

This leads to the notation

$$\mathcal{A}^{\mathbb{N}} = \{(a_0, a_1, a_2, a_3, \dots) : a_i \in \{0, 1\}, i \in \mathbb{N}\},$$

so that the members of  $\mathcal{A}^{\mathbb{N}}$  are one-sided infinite sequences of zeros and ones (more formally,  $\mathcal{A}^{\mathcal{B}}$  is defined as the set of all functions  $f : \mathcal{B} \rightarrow \mathcal{A}$ ).

We define a distance  $d$  on  $\mathcal{A}^{\mathbb{N}}$  in the following way: if

$$\omega_1 = (s_0, s_1, s_2, s_3, \dots), \quad \omega_2 = (t_0, t_1, t_2, t_3, \dots),$$

are members of  $\mathcal{A}^{\mathbb{N}}$ , then

$$d(\omega_1, \omega_2) = \sum_{n=0}^{\infty} \frac{|s_n - t_n|}{2^n}.$$

It is now straightforward to check that  $d$  is a metric on  $\mathcal{A}^{\mathbb{N}}$ , so  $(\mathcal{A}^{\mathbb{N}}, d)$  is a metric space (see Exercises 4.2).

For example, suppose that  $\omega_1 = (1, 1, 1, 1, \dots)$  and  $\omega_2 = (1, 0, 1, 0, 1, \dots)$ , then

$$d(\omega_1, \omega_2) = \sum_{n=1}^{\infty} \frac{1}{2^{2n-1}} = \frac{2}{3}.$$

With this metric, points with coordinates which differ for small  $n$  are further apart than those with coordinates that differ for large values of  $n$ . The distance between  $(1, 0, 1, 1, 1, \dots)$  and  $(1, 1, 1, 1, \dots)$  is  $1/2$ , whereas the distance between  $(1, 1, 1, 0, 1, 1, 1, 1, \dots)$  and  $(1, 1, 1, 1, 1, 1, \dots)$  is  $1/8$ .

Some other sequence spaces that arise are  $\mathcal{A}^{\mathbb{Z}^+} = \{(a_1, a_2, a_3, \dots) : a_i \in \{0, 1\}, i \in \mathbb{Z}^+\}$  and  $\mathcal{A}^{\mathbb{Z}} = \{(\dots, a_{-1}, a_0, a_1, a_2, \dots) : a_i \in \{0, 1\}, i \in \mathbb{Z}\}$ , the latter space being the set of all two-sided sequences of zeros and ones with metric defined by

$$d(\omega_1, \omega_2) = \sum_{n=-\infty}^{\infty} \frac{|s_n - t_n|}{2^{|n|}},$$

when  $\omega_1 = (\dots, s_{-1}, s_0, s_1, s_2, \dots)$ ,  $\omega_2 = (\dots, t_{-1}, t_0, t_1, t_2, \dots)$ . In this case, points are close when their components are equal near the 0th coordinate.

**Definition 4.1.4** If  $(X, d)$  is a metric space and  $a \in X$ ,  $\epsilon > 0$ , then

$$B_\epsilon(a) = \{x \in X : d(a, x) < \epsilon\},$$

is called the *open ball* centered on  $a$  with radius  $\epsilon$ .

**Examples 4.1.5** If  $X = \mathbb{R}$  as in Example 4.1.3, then the open balls are intervals of the form  $(a - \epsilon, a + \epsilon)$ . If  $X = \mathbb{R}^2$ , and  $a = (a_1, a_2)$  is a point of  $\mathbb{R}^2$ , the points of  $B_\epsilon(a)$  are those interior to the circle centered on  $(a_1, a_2)$  and of radius  $\epsilon$  (the boundary of the circle is not included). If  $X = [0, 1]$  with the same metric, the open balls are intervals of the form  $(c, d)$  for  $0 < c < d < 1$ , together with the half-open intervals of the form  $[0, c)$  and  $(d, 1]$ .

**Definition 4.1.6** (i) A set  $A \subseteq X$  is said to be *open* if for each  $a \in A$  there exists  $\epsilon > 0$  satisfying  $B_\epsilon(a) \subseteq A$  (i.e., every point of  $A$  can be surrounded by an open ball entirely contained in  $A$ ).

- (ii) The set  $A \subseteq X$  is *closed* if its *complement*  $X \setminus A = \{x \in X : x \notin A\}$ , (sometimes written  $A^c$ ), is open.
- (iii) A point  $a \in X$  is a *limit point* of a set  $A \subseteq X$  if every ball  $B_\epsilon(a)$  contains a point of  $A$  other than  $a$ . In other words

$$A \cap B'_\epsilon(a) \neq \emptyset, \quad \text{for all } \epsilon > 0,$$

where  $B'_\epsilon(a) = B_\epsilon(a) \setminus \{a\}$  is the open ball centered on  $a$  with  $a$  missing (sometimes call a *punctured ball*).

**Examples 4.1.7** 1. The open intervals  $(a, b)$  in  $\mathbb{R}$  are open sets and any union of open intervals is open. The closed intervals such as  $[a, b], [a, \infty)$  are closed sets (see Exercises 4.2). Sets such as  $\mathbb{Q}$  and  $\mathbb{R} \setminus \mathbb{Q}$  are neither open nor closed.

2. Any open ball  $B_\epsilon(a)$  in a metric space  $X$  is an open set. The *closed ball*

$$\overline{B}_\epsilon(a) = \{x \in X : d(a, x) \leq \epsilon\},$$

will be a closed set. The empty set  $\emptyset$  and the whole space  $X$  are both open and closed.

3. In any metric space  $X$ , the union of open sets is open and the intersection of closed sets is closed.

**Proof.** Let  $A = \bigcup_{\alpha \in \Lambda} A_\alpha$  be a union of open sets, with  $a \in A$ . By definition,  $a \in A_\alpha$  for some  $\alpha \in \Lambda$ . Since  $A_\alpha$  is open, there exists  $\delta > 0$  with  $B_\delta(a) \subseteq A_\alpha$ . It follows that  $B_\delta(a) \subseteq A$ , so  $A$  is open.

If each  $C_\alpha$  is closed,  $\bigcap_{\alpha \in \Lambda} C_\alpha$  is an intersection of closed sets, and each  $X \setminus C_\alpha$  is open, so  $\bigcup_{\alpha \in \Lambda} (X \setminus C_\alpha)$  is open. But *DeMorgan's Laws* tell us that

$$\bigcup_{\alpha \in \Lambda} (X \setminus C_\alpha) = X \setminus \bigcap_{\alpha \in \Lambda} C_\alpha,$$

so that  $\bigcap_{\alpha \in \Lambda} C_\alpha$  is closed. □

4. The interval  $(0, 1) \subset \mathbb{R}$ , and the set  $\{1, 1/2, 1/3, \dots, 1/n, \dots\} \subset \mathbb{R}$  both have 0 as a limit point. This is because every open ball centered on 0 contains points of the set in question, other than 0.

5. If  $X$  is given the discrete metric and  $0 < \epsilon < 1$ , then

$$B_\epsilon(a) = \{a\}, \quad \text{for all } a \in X.$$

In particular, each *singleton set* (i.e., each set containing a single point) is an open set. It follows that every set is open since it is the union of open sets.

6. If  $(X, d)$  is a metric space and  $A$  is a subset of  $X$ , then we may regard  $(A, d)$  as a metric space (called a *subspace* of  $X$ ). For example if  $A = [0, 1]$ , the subset of  $\mathbb{R}$  with the usual metric, then in  $A$ ,  $B_\epsilon(0) = [0, \epsilon]$  is the open ball centered on 0.

## 4.2 Dense Sets.

In order to define the notion of chaos for one-dimensional maps, we need various topological notions such as denseness and transitivity.

**Definition 4.2.1** The *closure* of a set  $A$  in a metric space  $X$ , is defined to be

$$\overline{A} = A \cup \{ \text{the limit points of } A \}.$$

**Proposition 4.2.2**  $\overline{A}$  is the smallest closed set containing  $A$ , (i.e., if  $B$  is another closed set containing  $A$  then  $\overline{A} \subseteq B$ ).

**Proof.** Clearly  $A \subset \overline{A}$ . To see that  $\overline{A}$  is a closed set, let  $a \in X \setminus \overline{A}$ . Then  $a$  is not a limit point of  $A$  so there exists  $\delta > 0$  with  $B_\delta(a) \cap A = \emptyset$ .

We claim  $B_\delta(a) \cap \overline{A} = \emptyset$ , for if  $x$  belongs to this set, then  $x \notin A$  but  $x$  is a limit point of  $A$ . Also  $x \in B_\delta(a)$ , an open set, so there exists  $\epsilon > 0$  with  $B_\epsilon(x) \subseteq B_\delta(a)$ . Hence

$$B_\epsilon(x) \cap A \subseteq B_\delta(a) \cap A = \emptyset,$$

contradicting the fact that  $x$  is a limit point of  $A$ .

It now suffices to show that if  $B$  is any other closed set containing  $A$ , then  $\overline{A} \subseteq B$ . Let  $x \in X \setminus B$ , an open set. Then

$$B_\delta(x) \subset X \setminus B \subseteq X \setminus A \quad \text{for some } \delta > 0,$$

and in particular  $B_\delta(x) \cap A = \emptyset$ . This says that  $x$  is not a limit point of  $A$ , so  $x \in X \setminus \overline{A}$ , and hence  $X \setminus B \subseteq X \setminus \overline{A}$ , or  $\overline{A} \subseteq B$ .

□

**Definition 4.2.3** The set  $A \subseteq X$  is *dense* in  $X$  if  $\overline{A} = X$ .

**Examples 4.2.4** 1. Let  $I \subseteq \mathbb{R}$  be an interval.  $A \subseteq I$  is *dense* in  $I$  if for any open interval  $U$  contained in  $I$ , we have  $U \cap A \neq \emptyset$ . This is because every  $x \in I$  is a limit point of  $A$ : if  $x \in I$ ,  $B_\delta(x) \cap A \neq \emptyset$  for every  $\delta > 0$ .

Equivalently,  $A$  is dense in  $I$  if for any  $x \in I$  and any  $\delta > 0$ , the interval  $(x-\delta, x+\delta)$  contains a point of  $A$ .

Intuitively, the points of  $A$  are spread uniformly over the interval  $I$  in such a way that every subinterval of  $I$  (no matter how small), contains some points of  $A$ .

2. The set  $\mathbb{Q}$  of all rational numbers is dense in  $\mathbb{R}$ .  $\mathbb{Q} \cap [0, 1]$  is dense in  $[0, 1]$ .

**Proof.** We show that the set  $\mathbb{Q} \cap [0, 1]$  is dense in  $[0, 1]$ . Let  $x \in (0, 1)$ . It suffices to find a rational number  $y$  arbitrarily close to  $x$ , i.e., satisfying  $|x - y| < \delta$ , for some arbitrary number  $\delta > 0$ . Suppose that  $x$  has a decimal expansion

$$x = \sum_{n=1}^{\infty} \frac{d_n}{10^n} = \cdot d_1 d_2 d_3 \dots, \quad \text{where } d_n \in \{0, 1, 2, \dots, 9\}.$$

Choose  $m \in \mathbb{Z}^+$  so large that  $10^{-m} < \delta$  and set  $y = \sum_{n=1}^m \frac{d_n}{10^n} \in \mathbb{Q}$ , then

$$|x - y| = |\cdot d_{m+1} d_{m+2} \dots| = \sum_{n=m+1}^{\infty} \frac{d_n}{10^n} \leq \sum_{n=m+1}^{\infty} \frac{9}{10^n} = \frac{1}{10^m} < \delta.$$

□

3. The set  $\mathbb{R} \setminus \mathbb{Q}$  of all irrational numbers is dense in  $\mathbb{R}$ .

4. For intervals in  $\mathbb{R}$ ,  $\overline{(a, b)} = [a, b]$ . If  $A = \{1, 1/2, 1/3, \dots\}$ , then  $\overline{A} = A \cup \{0\}$ .

5. A proper subset of a metric space  $X$  can be both open and dense. For example, in  $\mathbb{R}$ , the set  $E = \mathbb{R} \setminus \{\sqrt{2}\}$  is both open and dense, with the property that  $\mathbb{Q} \subset E \neq \mathbb{R}$ .

**Definition 4.2.5** If  $(X, d)$  is a metric space containing a sequence  $(x_n)$ , then we say  $\lim_{n \rightarrow \infty} x_n = a \in X$ , if for all  $\epsilon > 0$  there exists  $N \in \mathbb{Z}^+$  such that  $d(a, x_n) < \epsilon$  for all  $n > N$ .

**Theorem 4.2.6** *The following are equivalent for a metric space  $(X, d)$ :*

- (i) *The set  $A$  is dense in  $X$ .*
- (ii) *Let  $\epsilon > 0$ , then for all  $x \in X$  there exists  $a \in A$  such that  $d(a, x) < \epsilon$ .*
- (iii) *For all  $x \in X$ , there is a sequence  $(a_n)$  in  $A$  with  $\lim_{n \rightarrow \infty} a_n = x$ .*

**Proof.** (i)  $\Rightarrow$  (ii). Suppose that  $A$  is dense in  $X$  and let  $x \in X$ , then either  $x \in A$  (so (ii) holds) or  $x$  is a limit point of  $A$ :

i.e., any ball  $B_\epsilon(x)$  contains points of  $A$ , so there exists  $a \in A$  with  $d(x, a) < \epsilon$ .

(ii)  $\Rightarrow$  (iii). Let  $x \in X$  and  $n \in \mathbb{Z}^+$ , then there exists  $a_n \in A$  with  $d(x, a_n) < 1/n$  ( $n = 1, 2, \dots$ ), so that as  $n \rightarrow \infty$ ,  $d(x, a_n) \rightarrow 0$ .

i.e., given any  $\epsilon > 0$  there exists  $N \in \mathbb{Z}^+$  such that

$$n > N \Rightarrow d(x, a_n) < \epsilon, \quad \text{or} \quad \lim_{n \rightarrow \infty} a_n = x.$$

(iii)  $\Rightarrow$  (i). We must show that  $\overline{A} = X$ . Let  $x \in X$ . Then there exists a sequence  $(a_n)$  in  $A$  with  $\lim_{n \rightarrow \infty} a_n = x$ , i.e.,  $x$  is a limit point of  $A$  since  $B_\delta(x) \cap A \neq \emptyset$  for each  $\delta > 0$ , so  $\overline{A} = X$ .

□

**Example 4.2.7** 1. For a continuous function  $f : \mathbb{R} \rightarrow \mathbb{R}$  the set  $\text{Per}_1(f) = \text{Fix}(f)$  of all fixed points of  $f$  is a closed subset of  $\mathbb{R}$ . It follows that a continuous function on  $\mathbb{R}$ , with a dense set of fixed points satisfies  $f(x) = x$  for all  $x \in \mathbb{R}$ . We will see later that this type of result holds for continuous functions on metric spaces.

**Proof.** Let  $x \in X$  and suppose that  $(x_n)$  is a sequence of points in  $\text{Fix}(f)$  with  $x_n \rightarrow x$ . The continuity of  $f$  implies  $\lim_{n \rightarrow \infty} f(x_n) = f(x)$ , or

$$x = \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} f(x_n) = f(x),$$

so  $x$  is also fixed point and  $\text{Fix}(f)$  is closed.

2. If  $f(x) = \begin{cases} x \sin(1/x); & x \neq 0 \\ 0; & x = 0 \end{cases}$ , then  $x = 0$  and  $x_k = \frac{2}{\pi(1+4k)}$ ,  $k \in \mathbb{Z}$  are fixed points. Note that  $\lim_{k \rightarrow \infty} x_k = 0$ . Also, each  $x_k$  is a non-hyperbolic fixed point, and  $f'(0)$  is not defined. The set  $\text{Fix}(f) = \{x_k : k \in \mathbb{Z}\} \cup \{0\}$  is clearly a closed set. This function has many eventually fixed points which are easy to find. The fixed point  $x = 0$  is stable but not attracting since there are other fixed points arbitrarily close to it (see Exercises 2.3 # 15).

## Exercises 4.2

1. Prove that every open ball  $B_\epsilon(a)$  in a metric space  $(X, d)$  is an open set and that every finite subset of  $X$  is a closed set.

2 Show that the closed ball  $\{x \in X : d(a, x) \leq \epsilon\}$  in a metric space is a closed set, but it need not be equal to the closure of the open ball  $B_\epsilon(a)$  (Hint: Consider the two point space  $\mathcal{A} = \{0, 1\}$  with metric  $d(0, 1) = 1$ ).

3. Distances  $d_1, d_2, d_3, d_4, d_5$  are defined on  $\mathbb{R}$  by:

$$\begin{aligned} d_1(x, y) &= (x - y)^2, \quad d_2(x, y) = \sqrt{x - y}, \quad d_3(x, y) = |x^2 - y^2|, \\ d_4(x, y) &= |x - 2y|, \quad d_5(x, y) = \frac{|x - y|}{1 + |x - y|}, \quad x, y \in \mathbb{R}. \end{aligned}$$

Which of these (if any) defines a metric on  $\mathbb{R}$ ? (Hint: For  $d_5$ , consider the properties of the map  $f(x) = x/(1 + x)$ ).

4. Let  $x = (x_1, x_2), y = (y_1, y_2)$  be points in  $\mathbb{R}^2$ . Show that the following define metrics on  $\mathbb{R}^2$ :

$$d'(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}, \quad d''(x, y) = |x_1 - y_1| + |x_2 - y_2|.$$

What do the open balls  $B_1(a)$ , where  $a = (a_1, a_2)$ , look like in each case?

5. Show that the intersection of a finite number of open sets  $A_1, A_2, \dots, A_n$ , in a metric space  $(X, d)$  is an open set. Show, by considering the intervals  $(-1/n, 1/n)$  ( $n \in \mathbb{Z}^+$ ), in  $\mathbb{R}$  that the intersection of infinitely many open sets need not be open.

6. If  $\mathcal{A} = \{0, 1\}$ ,  $\mathcal{A}^\mathbb{N}$  denotes the metric space of all sequences of 0's and 1's:

$$\mathcal{A}^\mathbb{N} = \{\omega = (a_0, a_1, a_2, \dots) : a_i = 0 \text{ or } 1\},$$

with metric

$$d(\omega_1, \omega_2) = \sum_{k=0}^{\infty} \frac{|s_k - t_k|}{2^k},$$

when  $\omega_1 = (s_0, s_1, s_2, \dots)$  and  $\omega_2 = (t_0, t_1, t_2, \dots)$ . Show that  $(\mathcal{A}^\mathbb{N}, d)$  is a metric space.

Find  $d(\omega_1, \omega_2)$  if:

- (i)  $\omega_1 = (0, 1, 1, 1, 1, \dots)$  and  $\omega_2 = (1, 0, 1, 1, 1, \dots)$ ,
- (ii)  $\omega_1 = (0, 1, 0, 1, 0, \dots)$  and  $\omega_2 = (1, 0, 1, 0, 1, \dots)$ .

7. Let  $f : I \rightarrow I$  be a continuous function defined on an interval  $I$ .

- (a) What can you say about the graph of  $f$ , if  $f$  has a dense set of points with  $f^2(x) = x$ ?
- (b) Show that the inverse of  $f$  must exist and that  $f$  must have at least one fixed point.
- (c) Deduce that if there exist  $x \in I$  with  $f(x) \neq x$ , then  $f$  must be strictly decreasing.
- (d) If  $f'(x)$  exists for all  $x \in I$ , show that the 2-cycles are non-hyperbolic, and any fixed point  $x_0$  is non-hyperbolic of the type  $f'(x_0) = -1$  (when  $f(x)$  is not the function  $y = x$ ).
- (e) Give an example of a function of the type appearing in (d).

8. (a) Show that the function  $f(x) = -1/(x+1)$  has the property that  $f^3(x) = x$  for all  $x \neq -1, x \neq 0$ .

(b) Show that if  $f^3(x) = x$  on some set  $I$ , and if  $g(x) = f^2(x)$ , then  $g^3(x) = x$  for all  $x \in I$ . Use this fact to give another function distinct from  $f$  in (a) with the property that its third iterate is the identity map on its domain.

(c) In general, what can you say about a function  $f$  with the property that  $f^3(x) = x$  for all  $x$  in its domain? Can it be continuous? Can it have points of period 2? Need it have a fixed point? If  $f'(x)$  exists in the domain of  $f$ , show that the 3-cycles are non-hyperbolic, and any fixed point  $x_0$  must be non-hyperbolic with  $f'(x_0) = 1$ .

9. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a continuous function. Show that the set  $\text{Per}_n(f) = \{x \in \mathbb{R} : f^n(x) = x\}$  is a closed set. Later we shall see that the set of all periodic points of a continuous map need not be a closed set, and may in fact be dense. This is the case for the logistic map  $L_4 : [0, 1] \rightarrow [0, 1]$ .

10. Let  $f : I \rightarrow I$  be a continuous function. The set of points of period exactly  $n > 1$  (sometimes called *prime period n*), need not be a closed set. Give an example to illustrate this. (Hint: Consider the map  $f(x) = 1/x$ ).

11. Denote by  $C[a, b]$  the set of all continuous functions  $f : [a, b] \rightarrow \mathbb{R}$ . If the distance between two such continuous functions is given by:

$$d_1(f, g) = \max_{x \in [a, b]} |f(x) - g(x)| \quad \text{or} \quad d_2(f, g) = \int_a^b |f(x) - g(x)| dx,$$

show that both  $d_1$  and  $d_2$  define metrics on  $C[a, b]$ .

12\*. Show that the periodic points of the tent map  $T : [0, 1] \rightarrow [0, 1]$  are dense in  $[0, 1]$ .

### 4.3 Functions Between Metric Spaces.

We now consider a dynamical system to be a pair  $(X, f)$ , where  $f : X \rightarrow X$  is a continuous function on a metric space  $X$ . Sometimes, we drop the requirement of continuity. In this section we look at continuity for functions between two metric spaces  $(X, d_1)$  and  $(Y, d_2)$ , in order to apply it to the study of dynamical systems.

**Definition 4.3.1** A function  $f : X \rightarrow Y$  between the metric spaces  $(X, d_1)$  and  $(Y, d_2)$  is said to be *continuous* at  $a \in X$  if given any  $\epsilon > 0$ , there exists  $\delta > 0$  such that if  $x \in X$ , then

$$d_1(x, a) < \delta \Rightarrow d_2(f(x), f(a)) < \epsilon.$$

**Proposition 4.3.2** *The following are equivalent for  $f : X \rightarrow Y$  between metric spaces  $(X, d_1)$  and  $(Y, d_2)$ :*

- (i)  $f$  is continuous at  $x = a$ .
- (ii) Given any  $\epsilon$ -ball  $B_\epsilon(f(a))$  centered on  $f(a)$ , there exists a  $\delta$ -ball  $B_\delta(a)$  centered on  $a$  such that

$$f(B_\delta(a)) \subset B_\epsilon(f(a)).$$

(iii) Given any open set  $V$  containing  $f(a)$ , there exists an open set  $U$  containing  $a$  such that

$$f(U) \subseteq V.$$

**Proof.** Since the equivalence of (i) and (ii) is a fairly straightforward restatement of the definition of continuity, we will only show the equivalence of (i) and (iii).

(i)  $\Rightarrow$  (iii) Let  $V \subset Y$  be open with  $f(a) \in V$ . Then there exists  $\epsilon > 0$  with  $B_\epsilon(f(a)) \subseteq V$  (since  $V$  is open). Since  $f$  is continuous at  $a$ , there exists  $\delta > 0$  with

$$f(B_\delta(a)) \subseteq B_\epsilon(f(a)) \subseteq V,$$

so set  $U = B_\delta(a)$ . Then  $f(U) \subseteq V$ .

(iii)  $\Rightarrow$  (i)  $B_\epsilon(f(a))$  is open in  $Y$  and contains  $f(a)$ , so by hypothesis there exists  $U$  open in  $X$ , containing  $a$  such that

$$f(U) \subseteq B_\epsilon(f(a)).$$

Since  $U$  is open, we can find  $\delta > 0$  such that  $B_\delta(a) \subseteq U$ . Thus

$$f(B_\delta(a)) \subseteq f(U) \subseteq B_\epsilon(f(a)),$$

so that  $f$  is continuous at  $a$ . □

**Definition 4.3.3** A function  $f : X \rightarrow Y$  is *continuous* if it is continuous at every  $a \in X$ .

We now give a “global” criteria for continuity of  $f$ . Recall that if  $V$  is a subset of  $Y$ , then the *inverse image* of  $V$  under  $f$  is the set  $f^{-1}(V) = \{x \in X : f(x) \in V\}$ .  $f^{-1}(V)$  is defined even if  $f$  is not invertible.

**Theorem 4.3.4** *The following are equivalent for a function  $f : X \rightarrow Y$  between metric spaces:*

- (i)  $f : X \rightarrow Y$  is continuous.
- (ii)  $f^{-1}(V)$  is open in  $X$  whenever  $V$  is open in  $Y$ .

**Proof.** (i)  $\Rightarrow$  (ii) Let  $V$  be open in  $Y$ . Then we may assume  $f^{-1}(V) \neq \emptyset$ . Let  $a \in f^{-1}(V)$ , then  $f(a) \in V$ . Since  $V$  is open there exists  $\epsilon > 0$  with  $B_\epsilon(f(a)) \subset V$ . Now  $f$  is continuous at  $a$ , so there exists  $\delta > 0$  such that

$$f(B_\delta(a)) \subseteq B_\epsilon(f(a)), \quad \text{so} \quad B_\delta(a) \subseteq f^{-1}B_\epsilon(f(a)) \subseteq f^{-1}(V),$$

and  $f^{-1}(V)$  is open in  $X$ .

(ii)  $\Rightarrow$  (i) If  $f^{-1}(V)$  is open in  $X$  whenever  $V$  is open in  $Y$ , let  $a \in X$  and  $\epsilon > 0$ , then  $f(a) \in B_\epsilon(f(a))$ , is an open set in  $Y$ .

By hypothesis,  $f^{-1}(B_\epsilon(f(a)))$  is open in  $X$ , and since  $a \in f^{-1}(B_\epsilon(f(a)))$ , there exists  $\delta > 0$  such that

$$B_\delta(a) \subseteq f^{-1}(B_\epsilon(f(a))) \quad \text{or} \quad f(B_\delta(a)) \subseteq B_\epsilon(f(a)),$$

so that  $f$  is continuous at  $a$ . □

As an application of this theorem, we show that the basin of attraction  $B_f(p)$ , of a fixed point  $p$  for a continuous map  $f : X \rightarrow X$ ,  $X$  a metric space, is an open set. This generalizes Proposition 2.1.2 concerning the immediate basin of attraction of an attracting fixed point. In the context of metric spaces we define attracting fixed point as follows:

**Definition 4.3.5** Let  $f : X \rightarrow X$  be map on a metric space  $(X, d)$  with fixed point  $f(p) = p$ . Then

- (i)  $p$  is a *stable fixed point* if for any open ball  $B_\epsilon(c)$ , there is an open ball  $B_\delta(c)$  such that  $f^n(x) \in B_\epsilon(c)$  for any  $x \in B_\delta(c)$  and  $n \in \mathbb{Z}^+$ .
- (ii)  $p$  is an *attracting fixed point* if there exists  $\epsilon > 0$  such that for all  $x \in B_\epsilon(p)$ ,  $f^n(x) \rightarrow p$  as  $n \rightarrow \infty$ .
- (iii)  $p$  is *asymptotically stable* if it is both stable and attracting.

As in the case of 1-dimensional dynamical systems, if we set  $n = 1$  in (i) above, we see that  $f$  has to be continuous at  $x = p$ . When dealing with metric spaces, our results mostly concern attracting fixed points and attracting periodic points (rather than stable or asymptotically stable fixed points).

Recall that the basin of attraction of  $p$  is the set

$$B_f(p) = \{x \in X : f^n(x) \rightarrow p \text{ as } n \rightarrow \infty\}.$$

We saw that for  $0 < \mu < 1$ , 0 is an attracting fixed point of the logistic map  $L_\mu(x) = \mu x(1 - x)$  with basin of attraction  $[0, 1]$ . In this case, the metric space is  $X = [0, 1]$ , so the basin of attraction is an open set, namely the whole space (the fixed point 0 is globally attracting). For continuous maps  $f : \mathbb{R} \rightarrow \mathbb{R}$  having an asymptotically stable fixed point  $p$ , the basin of attraction is an open set, which in this case is just

the union of open intervals. The largest such open interval to which  $p$  belongs is the *immediate basin of attraction of  $p$*  under  $f$ . It is unclear how to define the immediate basin of attraction for a map on a general metric space, consequently we will not be examining this issue.

Recall that  $A \subseteq X$  is *invariant* under  $f$  if  $f(x) \in A$  for all  $x \in A$ .

**Theorem 4.3.6** *If  $f : X \rightarrow X$  is a continuous map of a metric space  $X$  and  $p$  is an attracting fixed point of  $f$ , then  $B_f(p)$ , the basin of attraction of  $f$ , is an invariant open set.*

**Proof.** If  $x \in B_f(p)$ , then  $f^n(x) \rightarrow p$  as  $n \rightarrow \infty$ , and  $f^n(f(x)) \rightarrow f(p) = p$  as  $n \rightarrow \infty$ , so  $f(x) \in B_f(p)$  and  $B_f(p)$  is an invariant set.

Since  $p$  is an attracting fixed point, there exists  $\epsilon > 0$  such that if  $x \in B_\epsilon(p)$ ,  $f^n(x) \rightarrow p$  as  $n \rightarrow \infty$ . Since  $f$  is a continuous function, the set  $f^{-1}(B_\epsilon(p))$  is open (from Theorem 4.3.4). We claim that

$$B_f(p) = \bigcup_{n=1}^{\infty} f^{-n}(B_\epsilon(p)),$$

from which the result follows, since the union of open sets is open.

Let  $x \in \bigcup_{n=1}^{\infty} f^{-n}(B_\epsilon(p))$ , then  $x \in f^{-n}(B_\epsilon(p))$  for some  $n \in \mathbb{Z}^+$ , or  $f^n(x) \in B_\epsilon(p)$ , so clearly  $x \in B_f(p)$ .

On the other hand, suppose that  $x \in B_f(p)$ . Then there exists  $n \in \mathbb{Z}^+$  with  $f^n(x) \in B_\epsilon(p)$  (since  $p$  is attracting), so that  $x \in f^{-n}(B_\epsilon(p))$  and  $x \in \bigcup_{n=1}^{\infty} f^{-n}(B_\epsilon(p))$ .  $\square$

**Remarks 4.3.7** 1. If  $f : X \rightarrow X$  is continuous with  $\Lambda = \text{Fix}(f)$ , then  $\bigcup_{\alpha \in \Lambda} B_f(\alpha)$  is an open set. Thus  $C = X \setminus \bigcup_{\alpha \in \Lambda} B_f(\alpha)$  is a closed invariant set, which may be quite complicated.

2. For metric spaces, the idea of “sameness” is given by the notion of “homeomorphism”. Two metric spaces are regarded as “topologically” the same if there is a homeomorphism between them.

**Definition 4.3.8** A function  $h : X \rightarrow Y$  between metric spaces is a *homeomorphism* if:

- (i)  $h$  is one-to-one, so that if  $h(x) = h(y)$ , then  $x = y$ .
- (ii)  $h$  is onto, so that for every  $y \in Y$  there exists  $x \in X$  with  $h(x) = y$ .

- (iii)  $h$  is continuous.
- (iv) The inverse mapping  $h^{-1} : Y \rightarrow X$  is continuous.

The spaces  $X$  and  $Y$  are said to be *homeomorphic*, when there is a homeomorphism between them.

**Examples 4.3.8** 1. If  $f : [0, 1] \rightarrow [0, 1]$ ,  $f(x) = x^2$ , then  $f$  is one-to-one, onto, and is continuous with a continuous inverse, so  $f$  is a homeomorphism. In fact, any strictly increasing ( $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$ ), continuous function  $f : [0, 1] \rightarrow [0, 1]$  with  $f(0) = 0$  and  $f(1) = 1$  is a homeomorphism. Also, any strictly decreasing continuous function  $f : [0, 1] \rightarrow [0, 1]$  with  $f(0) = 1$  and  $f(1) = 0$  is a homeomorphism. It can be shown that any homeomorphism of  $[0, 1]$  to itself is of one of the above two types. The function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^3$  is a homeomorphism. The inverse function is  $f^{-1}(x) = x^{1/3}$ .

2. The function  $f : \mathbb{R} \rightarrow (-\pi/2, \pi/2)$ ,  $f(x) = \arctan x$  is a homeomorphism of the respective metric spaces. The logistic map  $L_\mu(x) = \mu x(1 - x)$ ,  $0 < \mu \leq 4$ , is not a homeomorphism of  $[0, 1]$  as it is not one-to-one (it is only onto when  $\mu = 4$ ).
3. Consider the spaces  $X = [0, 1]$  and  $\{0, 1\}^{\mathbb{N}} = \{(a_0, a_1, a_2, a_4, \dots) : a_i \in \{0, 1\}, i \in \mathbb{N}\}$  with the metrics given previously. We shall show later that these two metric spaces are not homeomorphic.
4. One might ask whether the metric spaces  $[0, 1]$  and  $[0, 1] \times [0, 1]$  are homeomorphic? It is known that there is a continuous onto map  $h : [0, 1] \rightarrow [0, 1] \times [0, 1]$ , but such a map cannot be one-to-one. This is related to the question of the existence of *space-filling* curves. In a similar way, the metric spaces  $\mathbb{R}$  and  $\mathbb{R}^2$  are not homeomorphic and there is no homeomorphism between the interval  $[0, 1]$  and the unit circle  $\mathbb{S}^1$ .

### Exercises 4.3

1. Let  $f : \mathbb{R} \rightarrow (-1, 1)$  be defined by:

$$f(x) = \frac{x}{1 + |x|}.$$

- (a) Show that  $f$  is a homeomorphism, and find the inverse map.

- (b) Extend  $f$  to a function  $f : [-\infty, \infty] \rightarrow [-1, 1]$  by setting  $f(-\infty) = -1$  and  $f(\infty) = 1$ . Define a metric  $d$  on  $[-\infty, \infty]$  by

$$d(x, y) = |f(x) - f(y)| \quad \text{for all } x, y \in [-\infty, \infty].$$

Show that  $d$  defines a metric on this space whose restriction to  $\mathbb{R}$  is different to the usual metric defined on  $\mathbb{R}$ .  $[-\infty, \infty]$  with this metric is called the *extended real line*.

2. Let  $f : X \rightarrow X$  be a map defined on the metric space  $(X, d)$  and let  $\alpha \in (0, \infty)$  be fixed, with the property:

$$d(f(x), f(y)) \leq \alpha d(x, y) \quad \text{for all } x, y \in X.$$

Show that  $f$  is continuous on  $X$ .

3. A map  $f : X \rightarrow X$  defined on the metric space  $(X, d)$  and which satisfies

$$d(f(x), f(y)) = d(x, y) \quad \text{for all } x, y \in X,$$

is called an *isometry*. Show that  $f$  is continuous, one-to-one, and hence a homeomorphism onto its range. What are the isometries  $f : \mathbb{R} \rightarrow \mathbb{R}$  ( $\mathbb{R}$  with the usual metric)?

4. Show that if  $f : [a, b] \rightarrow [a, b]$  is a homeomorphism, then either  $a$  and  $b$  are fixed points or  $\{a, b\}$  is a 2-cycle.

5. Let  $f : X \rightarrow X$  be a continuous function on a metric space  $(X, d)$ . Use the definition of continuity, limits and limit point, in this context to prove:

- (a) If  $\lim_{n \rightarrow \infty} x_n = a$ , then  $\lim_{n \rightarrow \infty} f(x_n) = f(a)$ .
- (b) If  $\lim_{n \rightarrow \infty} f^n(x_0) = p$ , then  $p$  is a fixed point of  $f$ .
- (c) If there is exactly one limit point of the set  $O(x_0) = \{x_0, f(x_0), f^2(x_0), \dots\}$ , then it is a fixed point of  $f$ .

6. Prove the following properties of a function  $f : X \rightarrow Y$  when  $U \subset X$ ,  $V \subset Y$ :

- (a)  $f(U) \subset V$  if and only if  $U \subset f^{-1}(V)$ .
- (b)  $f^{-1}(Y \setminus V) = X \setminus f^{-1}(V)$ .
- (c) Deduce that  $f : X \rightarrow Y$  is continuous if and only if for each closed set  $C$  in  $Y$ ,  $f^{-1}(C)$  is a closed set in  $X$ .

7. (a) A homeomorphism  $f : \mathbb{R} \rightarrow \mathbb{R}$  is an *involution* if  $f^2(x) = x$  for all  $x \in \mathbb{R}$ . Prove that  $f$  is an involution if and only if its graph is symmetric about the line  $y = x$ .

(b) Let  $a, b, c, d \in \mathbb{R}$ . A function of the form  $f(x) = \frac{ax + b}{cx + d}$  is called a *Möbius transformation*, or a *linear fractional transformation*.

Prove that  $f : \mathbb{R} \setminus \{-d/c\} \rightarrow \mathbb{R} \setminus \{a/c\}$  is a homeomorphism when  $ad - bc \neq 0$ , and find its inverse.

(c) Give conditions for  $f$  to have (i) a unique fixed point and (ii) to have two fixed points.

(d) Let  $f(x) = \frac{ax + b}{cx + a}$ , where  $a^2 - bc \neq 0$ ,  $x \neq -a/c$ . Show that  $f$  is an involution.

(e) If  $f(x) = \frac{ax + b}{cx - a}$ , where  $a^2 + bc \neq 0$ ,  $x \neq a/c$ , show that the graph of  $f$  is symmetric about the line  $y = -x$ . (Hint:  $y = f(x)$  is symmetric about  $y = -x$ , if whenever the point  $(x, y)$  lies on its graph, the points  $(-y, -x)$  also lies on its graph).

8. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a continuous map with fixed point  $c$  and basin of attraction  $B_f(c) = (a, b)$  an interval. Show that one of the following must hold:

- (a)  $a$  and  $b$  are fixed points.
- (b)  $a$  or  $b$  is fixed and the other is eventually fixed.
- (c)  $\{a, b\}$  is a 2-cycle.

9. Let  $X$  be a metric space that is not the union of two open sets ( $X$  is said to be *connected*). Let  $f : X \rightarrow X$  be a continuous map having at least 2 attracting fixed

points. If  $\Lambda$  is the set of all attracting fixed points of  $f$ , show that the complement of  $\cup_{\alpha \in \Lambda} B_f(\alpha)$  is a non-empty, closed invariant subset of  $X$ .

#### 4.4 Diffeomorphisms of $\mathbb{R}$ .

In this section, we study homeomorphisms  $f$  defined on subintervals of  $\mathbb{R}$  for which  $f$  and  $f^{-1}$  are differentiable functions. These are the *diffeomorphisms* of  $\mathbb{R}$ . If  $f$  is a homeomorphism, then the inverse function  $f^{-1}$  exists, with  $y = f(x)$  if and only if  $x = f^{-1}(y)$ . Denote by  $I$  and  $J$  open subintervals of  $\mathbb{R}$ , e.g.,  $I = (a, b)$  for some  $a, b \in \mathbb{R}$ ,  $(a < b)$ , or  $I = (a, \infty)$  etc.

**Definition 4.4.1** Let  $I$  and  $J$  be open intervals in  $\mathbb{R}$ . A function  $f : I \rightarrow J$  is said to be of *class  $C^1$*  on  $I$  if  $f'(x)$  exists and is continuous at all  $x \in I$ . Such a function is also said to be *smooth*. Functions of class  $C^2$ ,  $C^3$  etc. can be defined in a similar way, so that a function  $f$  is of class  $C^n$  on  $I$  if  $f$  is  $n$ -times differentiable at all  $x$ , and the  $n$ th derivative is continuous on  $I$ .

**Definition 4.4.2** A homeomorphism  $f : I \rightarrow J$  is called a *diffeomorphism* on  $I$  if  $f$  and  $f^{-1}$  are both  $C^1$ -functions on  $I$ . If  $f$  is a diffeomorphism on  $\mathbb{R}$  and  $I \subset \mathbb{R}$  is some closed interval, we sometimes refer to  $f$  as a diffeomorphism on  $I$ , onto the closed interval  $J$  where  $f(I) = J$ . This allows us to talk about diffeomorphisms on closed intervals without having to worry about differentiability at the end points.

If  $y = f(x)$ , then  $f^{-1}(f(x)) = x$ , so applying the chain rule gives

$$(f^{-1})'(f(x)) \cdot f'(x) = 1 \quad \text{or} \quad (f^{-1})'(y) = \frac{1}{f'(x)}.$$

A diffeomorphism  $f$  must have  $f'(x) \neq 0$  for all  $x \in I$ . Since  $f'(x)$  is a continuous function, if it were to take both positive and negative values, the Intermediate Value Theorem would imply it must take the value 0, giving a contradiction. Thus a diffeomorphism always has either  $f'(x) > 0$  or  $f'(x) < 0$  for all  $x \in I$ . This proves:

**Proposition 4.4.3** Let  $I$  and  $J$  be subintervals of  $\mathbb{R}$ . A diffeomorphism  $f : I \rightarrow J$  is either:

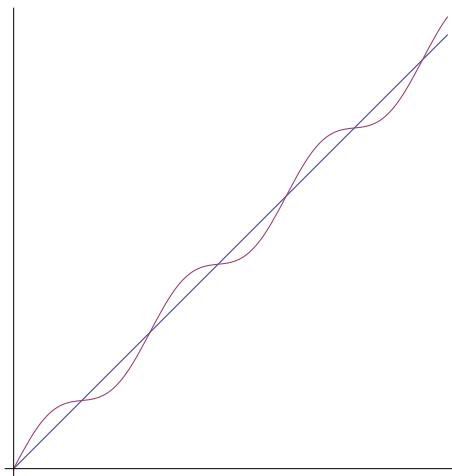
1. *Order preserving: for all  $x_1, x_2 \in I$ ,  $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$ , ( $f$  is a strictly increasing function) or*

2. *Order reversing:* for all  $x_1, x_2 \in I$ ,  $x_1 < x_2 \Rightarrow f(x_1) > f(x_2)$ , ( $f$  is a strictly decreasing function).

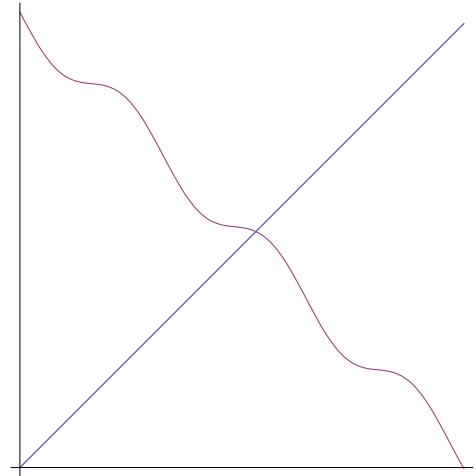
**Examples 4.4.4** 1. The function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^3$  is a  $C^1$  function but it is not a diffeomorphism since  $f'(0) = 0$ . The inverse function  $f^{-1}(x) = x^{1/3}$  is not differentiable at  $x = 0$  (vertical tangent). We see that  $f : (0, 1) \rightarrow (0, 1)$ ,  $f(x) = x^3$  is an order preserving diffeomorphism, but  $f : [0, 1] \rightarrow [0, 1]$ ,  $f(x) = x^3$  is not a diffeomorphism.

The function  $f : \mathbb{R} \rightarrow (-\pi/2, \pi/2)$ ,  $f(x) = \arctan x$  is an order preserving diffeomorphism. The inverse function is  $f^{-1} : (-\pi/2, \pi/2) \rightarrow \mathbb{R}$ ,  $f^{-1}(x) = \tan x$ .

2. Order preserving diffeomorphisms can have more than one fixed point, but we shall see that this is not true in the order reversing case. For example,  $f : [0, 1] \rightarrow [0, 1]$ ,  $f(x) = 1 - x$  is an order reversing diffeomorphism with a single fixed point and a 2-cycle  $\{0, 1\}$ .



Order preserving.



Order reversing.

**Theorem 4.4.5** Let  $I$  and  $J$  be intervals and  $f : I \rightarrow J$  an order reversing diffeomorphism with  $f(I) = J \subseteq I$ , then  $f$  has a unique fixed point in  $I$ .

**Proof.** If  $I = [a, b]$ , then Theorem 1.2.9 ensures the existence of a fixed point.

Suppose that  $I = (a, b)$ . Then if  $f(x)$  does not have a fixed point, either  $f(x) > x$  or  $f(x) < x$  for all  $x \in (a, b)$ . If the former holds, then since  $x < f(x) < b$ , as  $x \rightarrow b$ ,  $f(x) \rightarrow b$ , so  $f(x)$  becomes arbitrarily close to  $b$ . Clearly this is impossible for a continuous and strictly decreasing function. The case where  $f(x) < x$  is similar.

If  $I = \mathbb{R}$  let  $\alpha = \lim_{x \rightarrow -\infty} f(x)$  and  $\beta = \lim_{x \rightarrow \infty} f(x)$  ( $\alpha$  and  $\beta$  could be  $\pm\infty$ ). Then  $\alpha > \beta$  since  $f$  is order reversing. Let  $g(x) = f(x) - x$ , then

$$\lim_{x \rightarrow -\infty} g(x) = \infty, \quad \text{and} \quad \lim_{x \rightarrow \infty} g(x) = -\infty.$$

By the Intermediate Value Theorem, there exists  $c \in \mathbb{R}$  with  $g(c) = 0$ , so  $f(c) = c$ .

The situation for other types of intervals is similar.

Suppose now  $f$  is order reversing, and has two fixed points, say  $f(\alpha) = \alpha$  and  $f(\beta) = \beta$  with  $\alpha < \beta$ . Then  $f(\alpha) > f(\beta)$  or  $\alpha > \beta$ . Since they are fixed points, this is a contradiction.  $\square$

**Examples 4.4.6** Although order preserving diffeomorphisms can have any number of fixed points, they cannot have points of period greater than 1. On the other hand, order reversing diffeomorphisms can have points of period 2, but no points of greater period e.g., consider  $f(x) = -x$ . The bottom line is that the dynamics of one-dimensional diffeomorphisms are not complicated. This is not the case for two-dimensional diffeomorphisms (i.e., diffeomorphisms  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ ).

**Theorem 4.4.7** *Let  $f : I \rightarrow I$  be a diffeomorphism on an open interval  $I$ .*

- (i) *If  $f$  is order preserving, then  $f$  has no periodic points of period greater than 1.*
- (ii) *If  $f$  is order reversing, then  $f$  has no periodic points of period greater than 2.*

**Proof.** (i) Let  $O(x_0) = \{x_0, x_1, \dots, x_{n-1}\}$  be an  $n$ -cycle for  $f$ ,  $f(x_{n-1}) = x_0$ .

If  $x_1 > x_0$  then  $f(x_1) > f(x_0)$  or  $x_2 > x_1$ . Repeating this argument gives

$$x_0 < x_1 < x_2 < \dots < x_{n-1}, \quad \text{so that} \quad f^{n-1}(x_0) = x_{n-1} \neq x_0,$$

a contradiction.

On the other hand, if  $x_1 < x_0$ , we again get a contradiction, so we must have  $x_0 = x_1$ .

(ii) Notice that if  $f$  is an order reversing diffeomorphism, then  $f^2$  is an order preserving diffeomorphism. In fact if  $f'(x) < 0$  for all  $x$ , then

$$(f^2)'(x) = f'(f(x)) \cdot f'(x) > 0,$$

so by (i),  $f^2$  has no  $n$ -cycles with  $n > 1$ , and  $f$  has no  $2n$ -cycles with  $n > 1$ .

If  $n$  is odd, then  $f^n$  is a diffeomorphism with  $(f^n)'(x) < 0$ , so that  $f^n$  has a unique fixed point which must be the fixed point of  $f$ .  $\square$

We can improve upon some of the above results in the following way:

**Proposition 4.4.8** *Let  $f : [a, b] \rightarrow [a, b]$  ( $a < b$ ), be a continuous and one-to-one function.*

- (i) *Either  $f$  is strictly increasing or strictly decreasing on  $[a, b]$ .*
- (ii) *If  $f$  is strictly increasing, then every periodic point of  $f$  is a fixed point of  $f$ .*
- (iii) *If  $f$  is strictly decreasing, then  $f$  has exactly one fixed point and all other periodic points have period 2.*

**Proof.** (i) Suppose that  $f(a) < f(b)$  (they cannot be equal as  $f$  is one-to-one), and let  $x_1, x_2 \in (a, b)$  with  $x_1 < x_2$  and  $f(x_1) > f(x_2)$ . If  $f(a) < f(x_2) < f(x_1)$ , then by the Intermediate Value Theorem there exists  $c \in (a, x_1)$  with  $f(c) = f(x_2)$ , contradicting the one-to-oneness of  $f$ , so this is not possible. Similarly, if  $f(x_2) < f(a) < f(x_1)$  we find  $c \in (x_1, x_2)$  with  $f(c) = f(a)$ . Other possibilities are treated in a similar way.

(ii) Suppose that  $c$  is a periodic point of  $f$  having period  $p$ . If  $c < f(c)$ , then since  $f$  is strictly increasing,  $f(c) < f^2(c)$ , and we get an increasing sequence whose limit exists (say  $L$ ). Then

$$c < L = \lim_{n \rightarrow \infty} f^n(c) = \lim_{n \rightarrow \infty} f^{np}(c) = c,$$

a contradiction. Similarly  $f(c) < c$  leads to a contradiction, so  $f(c) = c$  and  $p = 1$ .

(iii) Since  $f$  is continuous on the interval  $[a, b]$  into itself,  $f$  has at least one fixed point. We must have  $f(a) > a$  since otherwise  $a = f(a) > f(b) \geq a$  which is impossible. Similarly  $f(b) < b$ .

If  $c_1$  and  $c_2$  are different fixed points of  $f$  with  $c_1 < c_2$ , then  $c_1 = f(c_1) > f(c_2) = c_2$ , which is impossible, so  $f$  has exactly one fixed point. Now notice that since  $f$  is order reversing,  $f^2$  is order preserving, so the only periodic points of  $f^2$  are fixed points, and these are period-2 points of  $f$ .

□

**Remarks 4.4.9** 1. If  $f : [a, b] \rightarrow [a, b]$  is a homeomorphism, then it is either strictly increasing or strictly decreasing. If it is increasing,  $f$  is order preserving with  $f(a) = a$ ,  $f(b) = b$  and the only other periodic points are fixed points. If it is decreasing (order reversing), then it has exactly one fixed point with all other periodic points having period 2. Also,  $f^2$  is order preserving and we must have  $f(a) = b$ ,  $f(b) = a$ . See [94] for a discussion of one-to-oneness.

### Exercises 4.4

1. Show that  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^3$  is not a diffeomorphism of  $\mathbb{R}$ . Give an example of a diffeomorphism  $f : \mathbb{R} \rightarrow \mathbb{R}$  which has (i) exactly one fixed point, (ii) exactly 2 fixed points, (iii) exactly 3 fixed points.
  
2. Find values of  $\lambda$  (if any) for which the function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is (a) a homeomorphism, (b) a diffeomorphism. If it is a homeomorphism, decide if it is order preserving or order reversing.
  - (i)  $f(x) = \lambda x + 3$ ,
  - (ii)  $f(x) = \lambda x^2$ ,
  - (iii)  $f(x) = x^3 + \lambda x$ ,
  - (iv)  $f(x) = x + \lambda \sin(x)$ ,
  - (v)  $f(x) = \lambda x + \sin(x)$ ,
  - (vi)  $f(x) = \lambda x + \arctan(x)$ .
  
3. Graph the functions in the last exercise using a computer algebra system together with a manipulate type plot, to confirm your answers.
  
- 4\*. (a) Show that  $f_n(x) = 1 + x + x^2/2! + x^3/3! + \cdots + x^{2n+1}/(2n+1)!$  is an order preserving diffeomorphism of  $\mathbb{R}$ . Deduce that for  $n \geq 1$ ,  $f_n$  has a unique fixed point and no points of any other period.
  - (b) If  $g_n(x) = 1 + x + x^2/2! + x^3/3! + \cdots + x^{2n}/(2n)!$ , use (a) to show that  $g_n(x) > 0$  for all  $x \in \mathbb{R}$ , and all  $n \geq 1$ . Deduce that  $g_n$  is concave up with a single critical point, and  $g_n(x)$  has no fixed points.
  
5. Show that for a continuous, increasing function  $f : [a, b] \rightarrow [a, b]$ , the periodic points cannot be dense in  $[a, b]$ . (It follows that a homeomorphism of  $[a, b]$  cannot be chaotic).

6. We have used the Intermediate Value Theorem throughout this text in the following form as one of our main tools: *Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function. If  $f(a) < 0$  and  $f(b) > 0$  then there exists  $c \in (a, b)$  with  $f(c) = 0$ .* Prove the Intermediate Value Theorem using the following steps:

- (a) Use the Bisection Method to obtain a sequence of nested intervals  $[a_n, b_n]$  of length  $(b - a)2^{-n}$ , where  $f(a_n) < 0$  and  $f(b_n) \geq 0$ . (Subdivide  $[a, b]$ , setting  $a_1 = (a + b)/2$  if  $f(a + b)/2) < 0$ , otherwise set  $b_1 = (a + b)/2$ , so that  $f(a_1) < 0$  and  $f(b_1) \geq 0$  and continue in this way to define sequences  $(a_n)$  and  $(b_n)$ .)
- (b) Show that  $\lim_{n \rightarrow \infty} a_n$  and  $\lim_{n \rightarrow \infty} b_n$  exist and are equal (use the Monotone Sequence Theorem from Appendix A).
- (c) Use the continuity of  $f$  to conclude that

$$f\left(\lim_{n \rightarrow \infty} a_n\right) \leq 0 \quad \text{and} \quad f\left(\lim_{n \rightarrow \infty} b_n\right) \geq 0.$$

- (d) Deduce that if  $c = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n$ , then  $f(c) = 0$ .

## CHAPTER 5

### Countability, Sets of Measure Zero and the Cantor Set.

Given an infinite set  $A$ , there are different ways of thinking of  $A$  as being large. For example, we might say that  $\mathbb{Q} \subset \mathbb{R}$  is a large set because it is dense in  $\mathbb{R}$ . However, in some sense it is the smallest type of infinite set, a *countable set*. It is also a small set in terms of *measure*. It is a *set of measure zero*. Sets of measure zero are sets that are small in the sense that they can be covered by a collection of intervals whose total length can be made as small as we please. We make this idea precise in this chapter. Many of the ideas in this section are due to George Cantor, a German mathematician of the 19th century whose studies of set theory helped to lay the foundations of modern point set topology. Of particular importance is the Cantor set, which is a metric space with many remarkable properties, and which we will see is homeomorphic to the metric space consisting of all sequences of 0's and 1's.

In this chapter we develop some of these important set theoretic ideas, and use them to help us describe the behavior of various dynamical systems such as the tent family  $T_\mu$  for  $\mu > 2$ , and the logistic family  $L_\mu$ , for  $\mu > 4$ . This will lead to an understanding of the chaotic nature of  $T_\mu$  and  $L_\mu$  for these parameter values.

#### 5.1 Countability and Sets of Measure Zero.

Before we examine the notion of countability, we look at sets of measure zero.

**Definition 5.1.1** Let  $I$  be a bounded subinterval of  $\mathbb{R}$ , having end-points  $a, b$  ( $a \leq b$ ). The *length* of  $I$  is then  $|I| = b - a$ . If  $I$  is an unbounded interval, we set  $|I| = \infty$ .

**Definition 5.1.2** We say that  $A \subseteq \mathbb{R}$  is a *set of measure zero* if we can cover  $A$  by bounded open intervals indexed by the set  $\mathbb{Z}^+$ , so that the total length of the intervals can be chosen to be arbitrarily small. More precisely, given any  $\epsilon > 0$ , there is a collection of open intervals  $\{I_n : n \in \mathbb{Z}^+\}$  with

$$A \subseteq \bigcup_{n=1}^{\infty} I_n \quad \text{and} \quad \sum_{n=1}^{\infty} |I_n| \leq \epsilon.$$

A collection of open sets  $O_\lambda$ , ( $\lambda \in \Lambda$ ), whose union contains a set  $A$  (in a metric space) is called an *open cover* of  $A$ .

The requirement that the intervals be open is not a serious one, as they may be closed or half open. For example, if the open intervals  $I_n$  are replaced by  $J_n$  with  $I_n \subset J_n$ ,  $J_n$  closed, and  $|J_n| = |I_n| + \epsilon/2^n$ , then

$$\sum_{n \in \mathbb{Z}^+} |J_n| = \sum_{n \in \mathbb{Z}^+} |I_n| + \sum_{n \in \mathbb{Z}^+} \epsilon/2^n \leq 2\epsilon.$$

If a property  $P$  holds everywhere except for a set of measure zero, we sometimes say that it holds *almost everywhere* (abbreviated a.e.). For example, if  $f : \mathbb{R} \rightarrow \mathbb{R}$  is defined by

$$f(x) = \begin{cases} 1; & x \in \mathbb{A} \\ 0; & \text{otherwise,} \end{cases}$$

where  $\mathbb{A}$  is a set of measure zero, then  $f(x) = 0$  a.e.

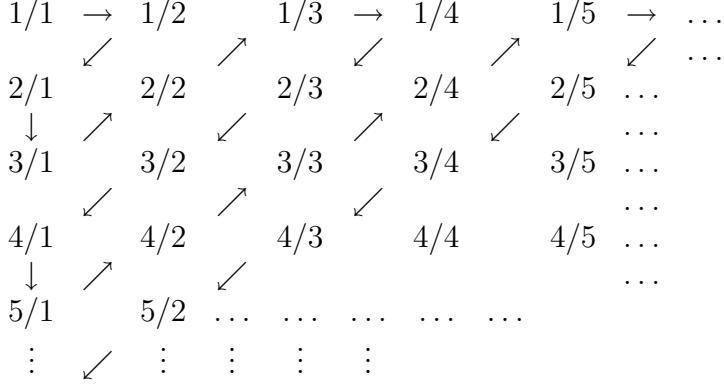
**Example 5.1.3** 1. A finite subset  $A = \{a_1, a_2, \dots, a_n\}$  of  $\mathbb{R}$  is a set of measure zero, for if  $\epsilon > 0$ , we can find open intervals, say  $I_i = (a_i - \epsilon/2n, a_i + \epsilon/2n)$ ,  $i = 1, 2, \dots, n$ , with  $A \subset \bigcup_{i=1}^n I_i$  and the total length of the intervals being  $\epsilon$ .

2. It is a remarkable fact that the set  $\mathbb{Q}$  of all rational numbers, is a set of measure zero, so can be covered by open intervals of arbitrary small total length. The proof, which appears later in this section, uses the notion of *countability*. Countably infinite sets (said to be *denumerable*) are in a sense the smallest type of infinite set.

**Definition 5.1.4** Any set  $A$  that is finite or can be put into a one-to-one correspondence with the set  $\mathbb{Z}^+$  is said to be *countable*. This is equivalent to saying that the members of  $A$  may be listed as a sequence.

**Examples 5.1.5** 1. If  $A = \{1, 4, 9, 16, \dots, n^2, \dots\}$ , then  $A$  is countable, for if we define  $f : \mathbb{Z}^+ \rightarrow A$  by  $f(n) = n^2$ , then  $f$  is a one-to-one and onto map. Any subset of a countable set can be seen to be countable.

2. The set  $\mathbb{Q}$  of rational numbers is a countable set. We list the positive rationals as a sequence (with some repetitions), by following the arrows below. The general case can be demonstrated with some minor modifications:



3. The set  $\mathbb{R}$  is not countable (it is said to be *uncountable*). In fact, we show using a *diagonal argument* due to Cantor, that the interval  $[0, 1]$  is uncountable.

Suppose that  $[0, 1]$  is a countable set. Then we can list its members as a sequence  $x_1, x_2, x_3, \dots, x_n, \dots$ . Each member of this sequence has a *binary expansion*, say

$$\begin{aligned}
 x_1 &= \cdot a_{11} a_{12} a_{13} a_{14} \dots \\
 x_2 &= \cdot a_{21} a_{22} a_{23} a_{24} \dots \\
 x_3 &= \cdot a_{31} a_{32} a_{33} a_{34} \dots \\
 &\vdots \quad \vdots \quad \vdots \quad \ddots \\
 x_n &= \cdot a_{n1} a_{n2} a_{n3} a_{n4} \dots \\
 &\vdots \quad \vdots \quad \vdots
 \end{aligned}$$

Let  $y = .b_1 b_2 b_3 b_4 \dots b_n \dots$ , where  $b_i = \begin{cases} 0; & a_{ii} = 1 \\ 1; & a_{ii} = 0 \end{cases}$ . Then  $y$  differs from  $x_1$  in the first binary digit, it differs from  $x_2$  in the second binary digit etc., differing from  $x_n$  in the  $n$ th binary digit, so  $y$  does not appear anywhere in the sequence, contradicting the fact that we have listed all members of  $[0, 1]$  as a sequence.

**Remarks 5.1.6** 1. Each  $x \in (0, 1)$  has a unique *non-terminating* binary expansion, for example:

$$\frac{1}{2} = .10000\dots, \quad \text{and} \quad \frac{1}{2} = .0111\dots,$$

so  $1/2$  has two different binary representations, but only the second is non-terminating. In the proof of the uncountability of  $[0, 1]$ , we should assume the  $x_i$  have a non-terminating expansion. This gives  $y = .b_1 b_2 b_3 b_4 \dots b_n \dots$ , but we do not know if it is terminating or not. To avoid this difficulty, we should start by listing all possible binary expansions (both terminating and non-terminating), so some numbers are listed twice, but there are only countably many having terminating expansions (these are

the binary rationals such as  $1/2$ ,  $3/4 = .1100\dots$  etc.) Now proceed as previously to see that  $y$  is not one of the enumerated elements.

Incidentally, the fact that  $\mathbb{R}$  is uncountable and  $\mathbb{Q}$  is countable shows the existence of irrational numbers.

2. It is easy to see that there is a one-to-one correspondence between the interval  $(0, 1)$  and  $\mathbb{R}$  (we say these sets have the same *cardinality*). For example, the map  $F : \mathbb{R} \rightarrow (-\pi/2, \pi/2)$ ,  $F(x) = \arctan(x)$  is both one-to-one and onto. If we take a linear (affine) map between  $(-\pi/2, \pi/2)$  and  $(0, 1)$ , the result is clear.

We can now prove the surprising fact that the rationals (and in fact any countable set), is a set of measure zero. For example, we can cover the rationals in  $[0, 1]$  with intervals whose total length is less than  $10^{-10}$ , so that all that remains uncovered are irrational numbers. We will see that an interval such as  $[0, 1]$ , is not a set of measure zero, and also that the irrational numbers in  $[0, 1]$  do not constitute a set of measure zero.

**Proposition 5.1.7** *The rational numbers form a set of measure zero.*

**Proof.** We can enumerate  $\mathbb{Q}$  as a sequence  $r_1, r_2, r_3, \dots, r_n, \dots$ , say. Let  $\epsilon > 0$ . For each  $n \in \mathbb{Z}^+$ , define an interval  $I_n$  by

$$I_n = (r_n - \epsilon/2^{n+1}, r_n + \epsilon/2^{n+1}).$$

Clearly  $r_n \in I_n$ ,  $n = 1, 2, \dots$ , so that  $\mathbb{Q} \subset \bigcup_{i=1}^{\infty} I_n$ . The total length of these intervals is

$$\sum_{i=1}^{\infty} |I_n| = \sum_{i=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon,$$

which can be made as small as we please. The set  $E = \bigcup_{n=1}^{\infty} I_n$  is another example of a set in  $\mathbb{R}$  which is both open and dense, containing  $\mathbb{Q}$  and not equal to  $\mathbb{R}$ .

□

We will show that an interval  $I$ , with end-points  $a$  and  $b$ ,  $a < b$ , cannot be a set of measure zero. This result is due to Borel (see Oxtoby [96]). We first prove a related result whose ideas are important in our study of compactness, in Chapter 17. It is known as the *Heine-Borel Theorem*:

**Lemma 5.1.8** *Let  $\{O_\lambda\}_{\lambda \in \Lambda}$  be an open cover of the interval  $[a, b]$ ,  $a \leq b$  (i.e.,  $[a, b] \subset \bigcup_{\lambda \in \Lambda} O_\lambda$ , where the sets  $O_\lambda$  are open). Then there is a finite subcollection of these sets that cover  $[a, b]$ .*

**Proof.** Let

$$\mathcal{S} = \{x \in [a, b] : \exists \text{ a finite subcollection from } \{O_\lambda\}_{\lambda \in \Lambda} \text{ that cover } [a, x]\}.$$

Clearly  $a \in \mathcal{S}$ , and the set  $\mathcal{S}$  is bounded above by  $b$ , so  $\alpha = \sup(\mathcal{S})$  exists and  $\alpha \leq b$ . Suppose that  $\alpha < b$ . Then  $\alpha$  belongs to some open set, say  $O_\mu$ , from our collection. Since  $O_\mu$  is open, there is an open interval  $(\alpha - \epsilon, \alpha + \epsilon) \subset O_\mu$ . If  $\beta \in (\alpha, \alpha + \epsilon)$ , then we can cover the interval  $[\alpha, \beta]$  by a finite subcollection of the open sets since we can always include  $O_\mu$ . This contradicts the fact that  $\alpha$  is the least upper bound of the set  $\mathcal{S}$ , so we must have  $\alpha = b$ .

□

**Proposition 5.1.9** *If a finite or infinite sequence of intervals  $(I_n)$ ,  $n = 1, 2, \dots$ , covers an interval  $I$  with end points  $a, b \in \mathbb{R}$ ,  $a < b$ , then  $\sum_n |I_n| \geq |I|$ .*

**Proof.** We give the proof for the case where the intervals  $I_n$  are open, and  $I = [a, b]$  is a closed interval. A slight modification will give the general result.

Denote by  $(a_1, b_1)$  the first interval in the sequence that contains the point  $a$ . If  $b_1 < b$ , let  $(a_2, b_2)$  be the first interval in the sequence  $I_n$  that contains the point  $b_1$ . Continue in this way so that if  $b_{n-1} < b$ ,  $(a_n, b_n)$  is the first interval in the sequence that contains the point  $b_{n-1}$ .

This procedure must terminate with some  $b_N > b$ , for if not, we would have an increasing sequence  $(b_n)$  bounded above by  $b$ , which converges to  $\alpha = \lim_{n \rightarrow \infty} b_n = \sup\{b_n : n \geq 1\} \leq b$ . Now  $\alpha$  belongs to  $I_k = (a', b')$  for some  $k$  since these intervals cover  $[a, b]$ .

Since  $\lim_{n \rightarrow \infty} b_n = \alpha$ , we must have  $b_m \in I_k$  for some  $m > 1$ . This implies  $b_{m+1} \geq b' > \alpha$ , contradicting  $\alpha$  being the supremum of the  $b_n$ 's. We have covered  $[a, b]$  by a finite subcollection of overlapping intervals  $(a_1, b_1), (a_2, b_2), \dots, (a_N, b_N)$ , with

$$b - a < b_N - a_1 = \sum_{i=2}^N (b_i - b_{i-1}) + b_1 - a_1 \leq \sum_{i=1}^N (b_i - a_i),$$

and the result follows.

□

We deduce that the interval  $[a, b]$ ,  $a < b$ , is not a set of measure zero. This also gives an alternative proof of the fact that  $[a, b]$  is an uncountable set:

**Corollary 5.1.10** *An interval  $I$  with end-points  $a, b \in \mathbb{R}$ ,  $(a < b)$ , is not of measure zero, and hence cannot be a countable set.*

## 5.2 The Cantor Set.

We will define a set  $C$ , called the *Cantor set*, or *Cantor's middle thirds set*, studied by George Cantor in 1883, but first introduced by Henry Smith in 1874 ([116]). It is the first fractal we shall meet, and it is a set having some remarkable properties.

Set  $S_0 = [0, 1]$ , the unit interval. We remove the open interval  $(1/3, 2/3)$  from  $S_0$  to give  $S_1 = [0, 1/3] \cup [2/3, 1]$ . We continue removing *open middle thirds* to give

$$S_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1],$$

and continue in this way removing open middle thirds so that

$$S_n = [0, \frac{1}{3^n}] \cup [\frac{2}{3^n}, \frac{3}{3^n}] \cup \dots \cup [\frac{3^n - 1}{3^n}, 1].$$

Denoting the total length of the subintervals making up  $S_n$  by  $|S_n|$  (so  $|S_0| = 1$ ), we have:

$S_1$  consists of 2 intervals of total length  $|S_1| = 2/3$ ,

$S_2$  consists of  $2^2$  intervals of total length  $|S_2| = 2^2/3^2$ ,

$S_3$  consists of  $2^3$  intervals of total length  $|S_3| = 2^3/3^3$ ,

and generally,

$S_n$  consists of  $2^n$  intervals of total length  $|S_n| = 2^n/3^n$ .



The first fours steps in the construction of the Cantor Set.

It may seem reasonable to define the Cantor set as the limiting process of removing these middle third sets, but we have not defined such limits. We use:

**Definition 5.2.1** The Cantor set  $C$  is defined by

$$C = \bigcap_{n=1}^{\infty} S_n.$$

Notice that the total length of the intervals removed is

$$1/3 + 2/3^2 + 2^2/3^3 + \dots = \frac{1/3}{1 - 2/3} = 1,$$

so it may seem that there will not be anything left in  $C$ . However, the end-points of the intervals  $S_n$  are never removed, and numbers such as  $3/10$  are never removed.

Our aim is to show that the Cantor set is large, in the sense that it is not countable, but small in the sense that it is a set of measure zero. We also investigate the properties of  $C$ , and show a connection between  $C$  and the ternary expansion of certain numbers in  $[0, 1]$ . The Cantor set is our first example of a *fractal* - it has the property of *self-similarity*. For example  $C \cap [0, 1/3]$  looks exactly like the Cantor set, but on a smaller scale.  $C \cap [0, 1/9]$  is a replica of  $C$  but on a smaller scale still. We can continue like this indefinitely to see  $C$  on smaller and smaller scales within itself.

**Proposition 5.2.2** *The Cantor set  $C$  is a closed, non-empty subset of  $[0, 1]$ , having measure zero.*

**Proof.**  $C$  is non-empty because it contains all of the end points of each of the intervals constituting the set  $S_n$ ,  $n \in \mathbb{Z}^+$  (for example,  $1/3 \in S_n$  for every  $n \in \mathbb{Z}^+$ ).  $C \subset S_n$  for  $n = 1, 2, \dots$  where  $|S_n| = (2/3)^n \rightarrow 0$  as  $n \rightarrow \infty$ , so that  $C$  may be covered by a collection of intervals whose total length can be made arbitrarily small. Thus,  $C$  is of measure zero.  $C$  is a closed set because each of the sets  $S_n$ ,  $n \in \mathbb{Z}^+$  is closed, and  $C$  is the intersection of closed sets.

□

## Exercises 5.2

1. A map  $f : \mathbb{N}^2 \rightarrow \mathbb{N}$  is defined by  $f(m, n) = 2^m 3^n$ . Show that  $f$  is one-to-one into a subset of  $\mathbb{N}$ . Deduce that  $\mathbb{N}^2$  is a countable set. Use this to give another proof that the set  $\mathbb{Q}^+$  of positive rationals, is a countable set.
  
2. Prove the following about sets of *measure zero*:
  - (a) A subset of a set of measure zero has measure zero.
  - (b) Any countable set has measure zero.

- (c) The countable union of sets of measure zero has measure zero.
3. Use a diagonal argument to prove that the set  $\mathcal{A}^{\mathbb{Z}^+} = \{(s_1, s_2, s_3, \dots) : s_i = 0 \text{ or } s_i = 1\}$  of all sequences of 0's and 1's, is uncountable.
4. Lemma 5.1.8 shows that if we cover an interval  $[a, b]$  with an arbitrary collection of open sets, there is a finite subcollection of these sets that also covers  $[a, b]$ . A set  $A \subseteq \mathbb{R}$  having this property (every open cover has a finite subcover), is said to be *compact*, and therefore the interval  $[a, b]$  is a compact set. Clearly every finite subset of  $\mathbb{R}$  is compact.
- (a) Show that the sets (i)  $A = [0, \infty)$ , (ii)  $B = \{1, 1/2, 1/3, \dots\}$  are not compact, by exhibiting an open cover that does not have a finite subcover. On the other hand, show that  $B \cup \{0\}$  is compact.
- (b) Show that any closed bounded subset  $K$  of  $\mathbb{R}$  is compact. (Hint: Use the fact that  $K$  is a subset of a closed interval of the form  $[a, b]$ , and extend an open cover of  $K$  to one for  $[a, b]$ . Then use the compactness of  $[a, b]$ ).
5. Show that a set of measure zero cannot contain an open interval (we say that a subset  $A \subset X$  of a metric space  $X$  has *empty interior* if it does not contain any open balls).
6. Two sets  $A$  and  $B$  have the same *cardinality* if there is a one-to-one, onto map  $f : A \rightarrow B$ .
- (a) If  $\mathcal{P}(A)$  denotes the *power set* of  $A$  (the set of all subsets of  $A$ ), show that  $A$  and  $\mathcal{P}(A)$  do not have the same cardinality. (Hint: Show that there is no onto map  $f : A \rightarrow \mathcal{P}(A)$ . Set  $B = \{x \in A : x \notin f(x)\} \in \mathcal{P}(A)$ , and show that  $B \neq f(x)$  for any  $x \in A$ ).
- (b) The above result shows that the sets  $\mathbb{N}$  and  $\mathcal{P}(\mathbb{N})$  are of different cardinality, so that  $\mathcal{P}(\mathbb{N})$  is not a countable set. By considering the power set of  $\mathcal{P}(\mathbb{N})$ , show that there are infinitely many sets having different cardinalities.

(c) Prove that the sets  $\mathbb{R}$  and  $\mathbb{R}^2$  have the same cardinality.

7. Use the Heine-Borel Theorem to show that if  $f : [a, b] \rightarrow \mathbb{R}$  is a continuous function, then  $f$  is a bounded (there exists  $K > 0$  such that  $|f(x)| < K$  for all  $x \in [a, b]$ ). (Hint: Use the continuity of  $f$  to show that for each  $x \in [a, b]$ , there exists  $\delta_x > 0$ , and an interval  $I_x = (x - \delta_x, x + \delta_x)$  on which  $f$  is bounded. Note that the set  $\{I_x : x \in [a, b]\}$  is an open cover of  $[a, b]$ ).

### 5.3 Ternary Expansions and the Cantor Set.

Each  $x \in [0, 1]$  has a *ternary expansion*, i.e., can be written as

$$x = \cdot a_1 a_2 a_3 a_4 \dots = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \frac{a_4}{3^4} + \dots, \quad \text{where } a_i = 0, 1 \text{ or } 2.$$

Note that if  $a_1 = 0$ , then  $x \in [0, 1/3]$  since  $x = \cdot 0 a_2 a_3 \dots \leq \cdot 0222 \dots = 1/3$ . Similarly, if  $a_1 = 1$ , then  $x \in [1/3, 2/3]$ , and if  $a_1 = 2$ , then  $x \in [2/3, 1]$ . Conversely, if  $x \in [0, 1/3]$ , the  $x$  has a ternary expansion with  $a_1 = 0$ , and similarly for the other two intervals.

In addition, just as in the binary expansion situation, every  $x \in (0, 1)$  has a unique *non-terminating* ternary expansion. For example

$$\frac{1}{3} = \cdot 1000 \dots = \cdot 0222 \dots, \quad \text{and} \quad \frac{2}{3} = \cdot 2000 \dots = \cdot 1222 \dots$$

We shall show that  $x \in C$  if and only if  $x$  has a ternary expansion (possibly terminating), consisting only of 0's and 2's. This means, for example, that  $1/3 \in C$ ,  $1 = \cdot 2222 \dots \in C$ ,  $2/3 \in C$  and  $3/4 = \cdot 20202 \dots \in C$ .

**Theorem 5.3.1**  $x \in C$  if and only if  $x$  has a ternary expansion

$$x = \cdot a_1 a_2 a_3 \dots, \quad \text{where } a_i = 0 \text{ or } 2.$$

**Proof.** First suppose that  $x \in C$  has a ternary expansion

$$x = \cdot a_1 a_2 a_3 \dots$$

Then  $x \in S_n$  for each  $n \in \mathbb{Z}^+$ . In particular  $x \in S_1$ , so either  $x \in [0, 1/3]$  and we may take  $a_1 = 0$ , or  $x \in [2/3, 1]$  and  $a_1 = 2$ . Also  $x \in S_2$ , so either  $x \in [0, 1/9]$  and  $a_2 = 0$ , or  $x \in [2/9, 1/3]$  and  $a_2 = 2$ , or  $x \in [2/3, 7/9]$  and  $a_2 = 0$ , or  $x \in [8/9, 1]$  and  $a_2 = 2$ . Continuing in this way we see that we may take  $a_n = 0$  or  $a_n = 2$  for each  $n \in \mathbb{Z}^+$ .

Conversely, suppose that we can choose the ternary expansion so that  $a_n = 0$  or  $a_n = 2$  for each  $n \in \mathbb{Z}^+$ . Since this holds for  $n = 1$ , we must have  $x \in [0, 1/3]$  or  $x \in [2/3, 1]$  so that  $x \in S_1$ . Similarly  $a_2 = 0$  or 2 implies that  $x \in S_2$ , and continuing in this way,  $x \in S_n$  for  $n = 1, 2, \dots$ , and we deduce that  $x \in \cap_{n=1}^{\infty} S_n = C$ .

□

We can now prove that  $C$  is not a countable set by showing that it can be put into a one-to-one correspondence with  $[0, 1]$ .

**Theorem 5.3.2** *The Cantor set  $C$  is uncountable.*

**Proof.** We define a one-to-one function  $f : [0, 1] \rightarrow C$  as follows:

Let  $x \in [0, 1]$  have a binary representation (non-terminating),

$$x = \cdot a_1 a_2 a_3 a_4 \dots \quad \text{where } a_i = 0 \text{ or } 1.$$

Define  $f(x)$  by

$$f(x) = \cdot b_1 b_2 b_3 b_4 \dots \quad \text{where } b_i = \begin{cases} 0; & a_i = 0, \\ 2; & a_i = 1. \end{cases}$$

Then  $f(x)$  is one-to-one (it is not onto because certain numbers such as  $2/3 = \cdot 1222\dots \in C$  are not in the range of  $f$ ). It follows that there is a subset of  $C$  that is uncountable, so  $C$  itself is uncountable.

□

Finally, we show that the Cantor set is *totally disconnected* and *perfect*. Our definition of being totally disconnected applies to the metric space  $\mathbb{R}$ . We will give a more general definition of total disconnectedness in Chapter 17.

**Definition 5.3.3** A subset  $A \subset \mathbb{R}$  is said to be *totally disconnected* (or has *empty interior*), if  $A$  contains no non-empty open intervals. For example, discrete sets of points are totally disconnected.  $\mathbb{Q}$ , the set of rationals in  $\mathbb{R}$ , is totally disconnected.

**Definition 5.3.4** A subset  $A \subset \mathbb{R}$  is said to be *perfect* if every point of  $A$  is a limit point of  $A$ .

**Theorem 5.3.5** *The Cantor set  $C$  is totally disconnected and perfect.*

**Proof.** If  $U$  is a non-empty open interval contained in  $C$ , then  $U$  is contained in  $S_n$  for each  $n \in \mathbb{Z}^+$ . But  $|S_n| = (2/3)^n \rightarrow 0$  as  $n \rightarrow \infty$ , so this is impossible as every non-empty open interval has positive length.

To see that  $C$  is perfect, let  $x \in C$  have ternary expansion

$$x = \cdot a_1 a_2 \dots a_n \dots \text{ where } a_i = 0, \text{ or } a_i = 2.$$

Set

$$x_n = \cdot a_1 a_2 \dots a_n * * * \dots,$$

chosen so that  $x_n \in C$ , and agrees with  $x$  in the first  $n$ -places, and  $x_n \neq x$ , and  $x_n \neq x_m$  for all  $n, m \in \mathbb{Z}^+$ ,  $n \neq m$ . Then

$$|x - x_n| \leq \sum_{k=n+1}^{\infty} \frac{2}{3^k} = \frac{1}{3^n} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

It follows that  $x$  is a limit point of  $C$ .

□

There are other “Cantor Sets” besides  $C$  (see Exercises 5.3). Any subset of  $\mathbb{R}$  that is closed, bounded, perfect, and totally disconnected is said to be a *Cantor Set*. We shall see that these arise quite naturally in dynamical systems theory.

### Exercises 5.3

1. (a) Find the ternary expansion of the following members of  $[0, 1]$ , and decide if they belong to the Cantor set: (i)  $x = 1/2$ , (ii)  $x = 1/4$ , (iii)  $x = 2/9$ .  
(b) Set  $C' = \{x \in [0, 1] : x \text{ does not have a ternary expansion involving } 1\}$  (e.g.,  $1/3 \notin C'$ ). Give two members of  $[0, 1]$  belonging to  $C'$ . Is  $C'$  open or closed? Is  $C'$  totally disconnected?
2. We have seen that the Cantor set is an uncountable set having measure zero. We generalize the notion of Cantor set to give an example of a set having empty interior which does not have measure zero:

Fix  $\alpha \in (0, 1)$ , and for each  $n \in \mathbb{Z}^+$ , let  $a_n = 2^{n-1}\alpha/3^n$ . Set  $S_0(\alpha) = [0, 1]$ . From its center, remove an open interval of length  $a_1$ , and denote by  $S_1(\alpha)$  the resulting set.  $S_1(\alpha)$  is a union of two disjoint closed intervals. From the center of each of these two intervals, remove an open interval of length  $a_2/2$  to obtain a set  $S_2(\alpha)$ , which is the union of  $2^2$  disjoint closed intervals of equal length. Continue in this way as in the construction of the Cantor set, so that at the  $n$ th stage we have a set  $S_n(\alpha)$

obtained by removing  $2^{n-1}$  disjoint closed intervals in  $S_{n-1}(\alpha)$ , an open interval of length  $a_{n-1}/2^{n-1}$ . Set

$$C(\alpha) = \bigcap_{n=1}^{\infty} S_n(\alpha).$$

$C(\alpha)$  is a *generalized Cantor set*. It is a subset of  $[0, 1]$  which can be shown to be uncountable and perfect. When  $\alpha = 1$  we obtain the usual Cantor set. Show:

- (a)  $C(\alpha)$  is a closed non-empty set.
- (b)  $C(\alpha)$  has empty interior. (Hint: Suppose that  $C(\alpha)$  contains an open interval  $I$ , and show this is impossible since the lengths of the intervals in  $C_n(\alpha)$  can be made arbitrarily small by making  $n$  large enough).
- (c)  $C(\alpha)$  is not a set of measure zero. (Hint: Find the total lengths of the intervals removed in the construction of  $C(\alpha)$ ).

3\*. *Cantor's Ternary Function* is defined as follows: Let  $x \in [0, 1]$  have ternary expansion  $x = \cdot x_1 x_2 x_3 \dots$  ( $x_i \in \{0, 1, 2\}$ ,  $i = 1, 2, 3, \dots$ ). Set  $N = \infty$  if  $x_n \neq 1$  for all  $n \in \mathbb{Z}^+$ , otherwise set  $N = \min\{n \in \mathbb{Z}^+ : x_n = 1\}$ . Let  $y_n = x_n/2$  for  $n < N$  and  $y_N = 1$ .

- (a) Show that if  $x$  has more than one ternary expansion, then  $\sum_{n=1}^N y_n/2^n$  is independent of which ternary expansion is used.
- (b) Show that the function  $\kappa : [0, 1] \rightarrow [0, 1]$  defined by

$$\kappa(x) = \sum_{n=1}^N \frac{y_n}{2^n}$$

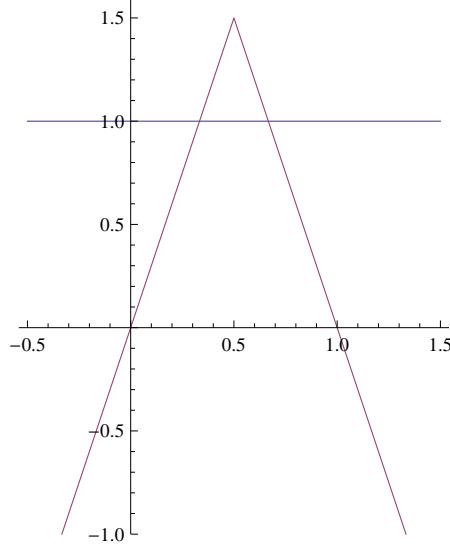
is a continuous, onto, and increasing.

- (c) Show that  $\kappa$  is constant on each interval contained in the complement of the Cantor set (for example,  $\kappa(x) = 1/2$  for all  $x \in (1/3, 2/3)$ ). Deduce that  $\kappa'(x) = 0$  everywhere except on a set of measure zero.

### 5.4 The Tent Map for $\mu = 3$ .

Now consider the tent family  $T_\mu$  for  $\mu = 3$ , where we think of  $T_3$  as a map defined on all of  $\mathbb{R}$  by

$$T_3(x) = \begin{cases} 3x; & x \leq 1/2, \\ 3(1-x); & x > 1/2. \end{cases}$$



The tent map  $T_\mu$  with  $\mu = 3$ .

Note that  $\{3/13, 9/13, 12/13\}$  and  $\{3/28, 9/28, 27/28\}$  are both 3-cycles for  $T_3$  and that if  $x > 1$ , then  $T_3^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ . We shall show that the orbit of  $x \in [0, 1]$  is bounded if and only if  $x \in C$ . In particular, if  $x \notin C$ , then  $T_3^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ . We also see that  $C$  is a set invariant under  $T_3$  ( $T_3(C) = C$ ), so we can consider  $T_3$  as a map  $T_3 : C \rightarrow C$ . This is where the interesting dynamics of  $T_3$  takes place.  $C$  is said to be the *attractor* of the map  $T_3$ .

**Proposition 5.4.1** *If  $\Lambda = \{x \in [0, 1] : T_3^n(x) \in [0, 1], \forall n \in \mathbb{Z}^+\}$ , then  $\Lambda = C$ , the Cantor set. In addition,  $T_3(C) \subseteq C$ .*

**Proof.**

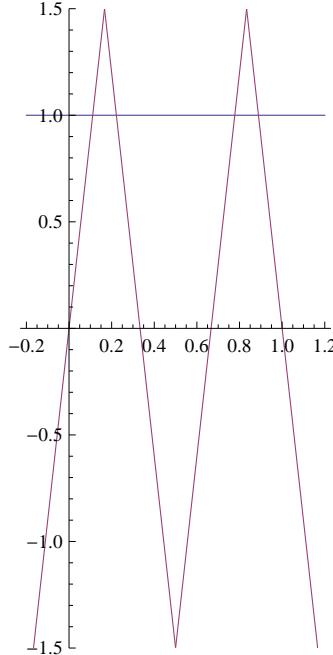
**Claim 1.** If  $x \in (1/3, 2/3)$ , then  $T_3^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ , since  $T_3$  is increasing on  $(1/3, 1/2]$ , and decreasing on  $[1/2, 2/3)$ , (if  $1/3 < x \leq 1/2$ , then  $T_3(1/3) < T_3(x) \leq T_3(1/2)$ , so  $1 < T_3(x) \leq 3/2$ ).

Similarly, if  $1/2 \leq x < 2/3$ , then  $1 < T_3(x) \leq 3/2$ , so  $T_3^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ .

**Claim 2.** If  $x \in (1/9, 2/9) \cup (7/9, 8/9)$ , then  $T_3^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ .

We use the fact that

$$T_3^2(x) = \begin{cases} 9x; & x \leq 1/6 \\ 3 - 9x; & 1/6 \leq x < 1/2 \\ 9x - 6; & 1/2 \leq x < 5/6 \\ 9 - 9x; & x \geq 5/6. \end{cases}$$



The tent map  $T_3^2$ .

$T_3^2$  is increasing on  $(1/9, 1/6]$  and decreasing on  $[1/6, 2/9)$ , so that if  $1/9 < x \leq 1/6$ , then  $T_3^2(1/9) < T_3^2(x) \leq T_3^2(1/6)$ , and  $1 < T_3^2(x) \leq 3/2$ . If  $1/6 < x \leq 2/9$  then  $1 < T_3^2(x) \leq 3/2$ . A similar argument also applies when  $x \in (7/9, 8/9)$ , to give  $T_3^2(x) > 1$  in each case, so that  $T_3^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ .

**Claim 3** If  $x \notin C$ , then  $T_3^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ .

We have seen that if  $x \in (1/3, 2/3)$ , then  $T_3^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ . Suppose instead  $x \in [0, 1/3] \cup [2/3, 1]$ , and has the ternary expansion

$$x = \cdot a_1 a_2 a_3 a_4 \dots \quad \text{where } a_i = 0, 1 \text{ or } 2,$$

then

$$T_3(x) = \begin{cases} \cdot a_2 a_3 a_4 \dots; & 0 \leq x \leq 1/3 \\ \cdot b_2 b_3 b_4 \dots; & 2/3 \leq x \leq 1, \end{cases}$$

where  $b_i = \begin{cases} 0; & a_i = 2 \\ 1; & a_i = 1 \\ 2; & a_i = 0 \end{cases}$ . Consequently, if  $x$  has a 1 in its ternary expansion, there exists  $k \in \mathbb{Z}^+$  with

$$T_3^k(x) = \cdot 1 a_{k+1} \dots, \quad \text{and so } T_3^k(x) \in [1/3, 2/3].$$

We can only have  $T_3^k(x) = 1/3$  when the ternary expansion of  $x$  consists of a sequence of 0's and 2's ( $k$  terms), followed by a 1, and then followed by an infinite string of 0's (in this case  $x \in C$  because the terms  $1000\dots$  can be written as  $0222\dots$ ). Thus,  $T_3^{k+1}(x) > 1$ , and the orbit of  $x$  will go to  $-\infty$ . Consequently, we deduce that if  $x \notin C$ , then  $T_3^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ .

**Claim 4.** If  $x \in C$ , then  $T_3(x) \in C$ .

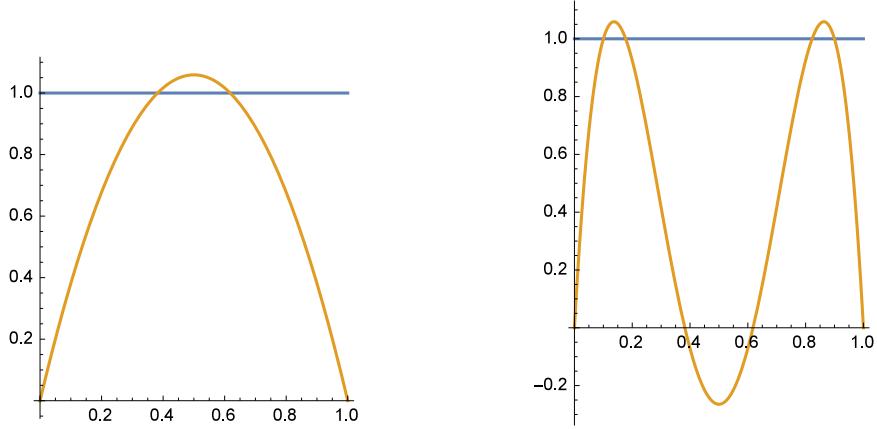
If each  $a_i = 0$  or 2, then each  $b_i$  (in Claim 3), is 0 or 2, so the result follows.  $\square$

In conclusion, we note that we may consider  $T_3$  as a map  $T_3 : C \rightarrow C$ . The Cantor set is where  $T_3$  behaves in a chaotic manner.  $T_3 : C \rightarrow C$  is an onto map, but it is clearly not one-to-one. A similar analysis can be made for  $T_\mu$  when  $\mu > 2$ , to deduce that there is some set  $C_\mu$ , a type of Cantor set, on which the dynamics is quite complicated. More complex reasoning of a similar type shows that the logistic map  $L_\mu = \mu x(1 - x)$ , for  $\mu > 4$ ,  $x \in \mathbb{R}$ , also has complicated dynamics on a Cantor set.

## 5.5 A Cantor Set Arising From the Logistic Map $L_\mu$ , $\mu > 4$ .

We outline the situation for the Cantor set arising from the logistic map  $L_\mu(x) = \mu x(1 - x)$  for  $\mu > 4$ . We think of  $L_\mu$  as a function defined on all of  $\mathbb{R}$ , but are mainly interested in its restriction to  $[0, 1]$ . In this case  $L_\mu(1/2) = \mu/4 > 1$ , and  $x = 1/2$  is a critical point for  $L_\mu$ . From the graph of  $L_\mu$ , we see that there are two points  $a_0$  and  $a_1 = 1 - a_0$  where  $L_\mu(a_0) = 1 = L_\mu(a_1)$ , and in addition,  $L_\mu(x) > 1$  for all  $x \in (a_0, a_1)$ . Just as for the tent map  $T_3$ , if  $x \in (a_0, a_1)$ , then  $L_\mu^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ . Set  $I_0 = [0, a_0]$ ,  $I_1 = [a_1, 1]$ , and  $\Lambda_1 = I_0 \cup I_1$ .  $\Lambda_1$  is a disjoint union of two closed intervals, and is the closed set:

$$\Lambda_1 = \{x \in [0, 1] : L_\mu(x) \in [0, 1]\}.$$



The graphs of  $L_\mu(x)$  and  $L_\mu^2(x)$  for  $\mu = 2 + \sqrt{5}$

We continue thus, as we did with the tent map  $T_3$ , to find  $\Lambda_2$ , a disjoint union of 4 closed intervals:

$$\Lambda_2 = \{x \in [0, 1] : L_\mu^2(x) \in [0, 1]\}.$$

At the  $n$ th stage,

$$\Lambda_n = \{x \in [0, 1] : L_\mu^n(x) \in [0, 1]\},$$

is the disjoint union of  $2^n$  closed intervals. We set  $\Lambda = \cap_{n=1}^{\infty} \Lambda_n$ .  $\Lambda \subset [0, 1]$  is a closed set because it is the intersection of closed sets.

Recall that if  $f : X \rightarrow X$  is a function on a metric space, the inverse image of a set  $U \subseteq X$  is the set  $f^{-1}(U) = \{x \in X : f(x) \in U\}$ . Clearly

$$\Lambda_1 = L_\mu^{-1}([0, 1]), \Lambda_2 = L_\mu^{-2}([0, 1]), \dots, \Lambda_n = L_\mu^{-n}([0, 1]),$$

$$\Lambda = \bigcap_{n=1}^{\infty} \Lambda_n = \bigcap_{n=1}^{\infty} L_\mu^{-n}[0, 1],$$

so that  $x \in \Lambda$  if and only if  $L_\mu^n(x) \in [0, 1]$  for all  $n \in \mathbb{Z}^+$ .  $\Lambda$  is a non-empty set because it must contain the periodic points of  $L_\mu$ . It is closed because it is the intersection of closed sets (which are the inverse image of closed sets). It can be shown to be a perfect, and a totally disconnected set of measure zero, so it is a Cantor set. As before, we can consider  $L_\mu$  as a map  $L_\mu : \Lambda \rightarrow \Lambda$ , where the chaotic nature of the map can be shown to reside.

The above analysis can be made easier in the case where  $\mu > 2 + \sqrt{5}$ . In this situation, we can find the points  $a_0$  and  $a_1$  by setting  $L_\mu(x) = 1$ , to give  $\mu x^2 - \mu x + 1 =$

0. Solving:

$$a_i = \frac{\mu \pm \sqrt{\mu^2 - 4\mu}}{2\mu}, \quad i = 0 \text{ or } 1.$$

For  $x \in \Lambda_1$ ,  $|L'_\mu(x)|$  is a minimum when  $x = a_0$ , or  $x = a_1$ , and then

$$L'_\mu(a_0) = \mu - 2\mu a_0 = \sqrt{\mu^2 - 4\mu}.$$

Thus for  $\mu > 0$ , and  $x \in \Lambda_1$ ,

$$L'_\mu(x) \geq L'_\mu(a_0) = \sqrt{\mu^2 - 4\mu} > 1,$$

when  $\mu^2 - 4\mu - 1 > 0$ , for which we require  $\mu > 2 + \sqrt{5}$ . This demonstrates that  $L_\mu$  cannot have any attracting periodic points if  $\mu > 2 + \sqrt{5}$ . It can now be shown that the individual intervals making up  $\Lambda_n$ , have length less than  $1/r^n$ , for some  $r > 1$ , and the fact that  $\Lambda$  is a Cantor set can now be deduced. A similar result holds for any  $\mu > 4$ , but the analysis is more difficult (see [32] for more details).



## CHAPTER 6

### Devaney's Definition of Chaos.

According to the dictionary, chaos means “complete confusion and disorder, a state in which behavior and events are not controlled by anything”. This is not type of chaos we study in this chapter. We consider the unpredictability that results from the iteration of certain functions, a type of “deterministic chaos”. For example, the great mathematician, John von Neumann, suggested that the logistic map  $L_4$ , could be used as a random number generator, since its iterations were seemingly random. An appropriate definition of deterministic chaos is still up for debate - comparisons of a number of different definitions of chaos are given in [16].

We shall investigate Devaney's definition of chaos for one-dimensional dynamical systems, and also for more general maps defined on metric spaces [32]. Devaney's original definition had three distinct requirements for a continuous map to be chaotic. It has since been determined that these conditions are not independent, and we investigate this situation. Continuous maps having points of period three need not be chaotic. For example, the logistic map  $L_4(x) = 4x(1 - x)$  has highly chaotic behavior as a dynamical system on  $[0, 1]$ .  $L_4$  has two 3-cycles, but as a dynamical system on  $\mathbb{R}$ , we regard it as not being chaotic: the behavior off  $[0, 1]$  is fairly mundane for if  $x > 1$ ,  $L^n(x) \rightarrow -\infty$  as  $n \rightarrow \infty$ .

For functions  $f : I \rightarrow I$ ,  $I \subseteq \mathbb{R}$ , it turns out that if  $f$  has a point with dense orbit, or if the periodic points form a dense subset of  $I$ , then  $f$  often has highly chaotic properties. Sharkovsky's Theorem tells us that if a continuous function has a point of period three, then it has points of all other periods, suggesting a very complicated dynamical behavior. Hence, Li and Yorke named their paper: “*Period three implies chaos*”. This is possibly the origin of the term “chaotic” in dynamical systems. Li and Yorke gave a different definition of chaos to the one of Devaney. Devaney's definition required the map to have a dense set of periodic points, to have a point with a dense orbit, and to have a type of sensitivity to initial conditions. We will investigate relations between these different conditions, and see how they apply to examples that we have already studied.

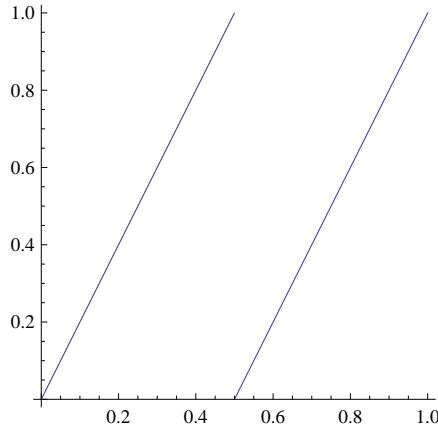
## 6.1 The Doubling Map and the Angle Doubling Map.

Unlike the examples we have considered so far, our first example in this chapter has a point of discontinuity. However, it illustrates quite nicely how a fairly simple map can have a dense set of periodic points, and consequently very complicated dynamics.

### Example 6.1.1 The Doubling Map.

The doubling map  $D : [0, 1) \rightarrow [0, 1)$  is defined by

$$D(x) = \begin{cases} 2x & ; \quad 0 \leq x < 1/2, \\ 2x - 1 & ; \quad 1/2 \leq x < 1. \end{cases}$$



The doubling map  $D$ .

It is instructive to describe  $D$  in terms of the binary expansion of a real number in  $[0, 1)$ . Any  $x \in [0, 1)$  can be represented by its (possibly non-unique), binary expansion

$$x = \cdot a_1 a_2 a_3 \dots \quad \text{where } a_i = 0 \text{ or } 1.$$

In other words,

$$x = \frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \dots = \sum_{i=1}^{\infty} \frac{a_i}{2^i}.$$

If  $x \in [0, 1/2)$ , then  $a_1 = 0$  in the binary expansion, and

$$D(x) = 2x = \frac{a_2}{2} + \frac{a_3}{2^2} + \dots = \cdot a_2 a_3 \dots$$

On the other hand, if  $x \in (1/2, 1)$ , then  $a_1 = 1$ , and

$$D(x) = 2x - 1 = (a_1 + \frac{a_2}{2} + \frac{a_3}{2^2} + \dots) - 1 = \frac{a_2}{2} + \frac{a_3}{2^2} + \dots = \cdot a_2 a_3 \dots$$

If  $x = 1/2$  there are two ways to represent  $x$ . Either  $x = \cdot 10000\dots$ , or  $x = \cdot 01111\dots$ , and there is some ambiguity. We see that in general, if  $x \neq 1/2$ , and  $x = \cdot a_1 a_2 a_3 \dots$ , then

$$D(\cdot a_1 a_2 a_3 \dots) = \cdot a_2 a_3 \dots$$

and more generally, if  $x$  is not a dyadic rational (i.e., does not have a terminating binary expansion), then

$$D^n(\cdot a_1 a_2 a_3 \dots) = \cdot a_{n+1} a_{n+2} a_{n+3} \dots$$

Consequently, if  $x = \cdot a_1 a_2 \dots a_n a_1 a_2 \dots a_n a_1 \dots$  has an expansion which repeats every  $n$  places, then  $D^n(x) = x$ , so that  $x$  is periodic of period  $n$ .

For example  $D^2(\cdot 010101\dots) = \cdot 010101\dots$ , so is a point of period 2. We can then show that the set of periodic points of  $D$  are dense in  $[0, 1]$ , and count the number of periodic points of period  $n$ . Notice that since  $D'(x) > 1$  everywhere it is defined, all of the periodic orbits of  $D$  are unstable.

**Proposition 6.1.2** *The periodic points of the doubling map are dense in  $[0, 1]$ .*

**Proof.** Let  $\epsilon > 0$ , and choose  $N$  so large that  $1/2^N < \epsilon$ . If  $x \in [0, 1)$ , it suffices to show that there is a periodic point  $y$  for  $D$  that is within  $\epsilon$  of  $x$ . Suppose that the binary expansion of  $x$  is

$$x = \cdot a_1 a_2 a_3 \dots = \sum_{i=1}^{\infty} \frac{a_i}{2^i}.$$

We set

$$y = \cdot a_1 a_2 \dots a_N a_1 a_2 \dots a_N a_1 \dots,$$

a point of period  $N$ . Then

$$|x - y| = \left| \sum_{j=N+1}^{\infty} \frac{b_j}{2^j} \right| \leq \sum_{j=N+1}^{\infty} \frac{1}{2^j} = \frac{1}{2^N} < \epsilon,$$

(where  $b_i = 0, 1$  or  $-1$ ).

**Example 6.1.3 The Angle Doubling Map.**

As before, we denote by  $\mathbb{C} = \{z = a + ib : a, b \in \mathbb{R}\}$ , the set of all complex numbers. If  $z = a + ib \in \mathbb{C}$ , then its *absolute value* (or *modulus*) is given by  $|z| = \sqrt{a^2 + b^2}$ . The conjugate of  $z$  is  $\bar{z} = a - ib$ , and we can check that  $z\bar{z} = |z|^2$ . We can also represent  $\mathbb{C}$  as the set of ordered pairs  $\{(a, b) : a, b \in \mathbb{R}\}$ , which we call the *complex plane*. The *unit circle*  $\mathbb{S}^1$  in  $\mathbb{C}$ , is the set

$$\mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}.$$

Points in  $\mathbb{S}^1$  may be represented as:

$$z = e^{i\theta} = \cos \theta + i \sin \theta, \quad -\pi < \theta \leq \pi.$$

Here  $\theta$  is the *principal argument* of  $z = a+ib$ , (written  $\text{Arg}(z)$ ), and it is the angle subtended by the ray from  $(0, 0)$  to  $(a, b)$  and the *real axis*, measured in the anticlockwise direction.

$\mathbb{S}^1$  is a metric space if the distance between two points  $z, w \in \mathbb{S}^1$  is defined to be the shortest distance between the two points, on the circle. We define a map  $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$  by  $f(z) = z^2$ . This map is called the *angle doubling map* because of the effect it has on  $\theta = \text{Arg}(z)$ :  $f(e^{i\theta}) = e^{2i\theta}$ . We see that the angle  $\theta$  is doubled. It is clear that there are some similarities between the doubling map, and the angle doubling map. Now we show that the periodic points of  $f$  are dense in  $\mathbb{S}^1$ .

Consider the periodic points of  $f(z) = z^2$ . Solving  $z^2 = z$  gives  $z = 1$  (we can disregard  $z = 0$ ),  $f^2(z) = z$  gives  $z^4 = z$  or  $z^3 = 1$ , and continuing in this way we see that the periodic points are some of the  $n$ th roots of unity.

**Proposition 6.1.4** *The periodic points of the angle doubling map  $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$  are dense in  $\mathbb{S}^1$ .*

**Proof.** If  $f^n(z) = z$  for some  $n \in \mathbb{Z}^+$ , then  $z^{2^n} = z$  or  $z^{2^n-1} = 1$ . Write  $z = e^{i\theta}$ , then we want to find the  $(2^n - 1)$ th roots of unity. Then:

$$e^{(2^n-1)i\theta} = e^{2k\pi i}, \quad \text{for some } k \in \mathbb{Z}^+,$$

giving the  $2^n - 1$  distinct roots:  $z_k = e^{2k\pi i/(2^n-1)}$ ,  $k = 0, 1, 2, \dots, 2^n - 2$ , showing that

$$\text{Per}_n(f) = \{e^{2k\pi i/(2^n-1)} : 0 \leq k < 2^n - 1\}, \quad n \in \mathbb{Z}^+.$$

These points are equally spaced around the circle, a distance  $2\pi/(2^n - 1)$  apart, which can be made arbitrarily small by taking  $n$  large enough. It follows that the periodic points are dense in  $\mathbb{S}^1$ . □

## 6.2 Transitivity.

Let  $X$  be a metric space and  $f : X \rightarrow X$  a function. Sometimes, when we iterate  $x_0 \in X$ , the orbit  $O(x_0) = \{x_0, f(x_0), \dots\}$ , spreads itself “uniformly” over  $X$ , so that  $O(x_0)$  is a dense set in  $X$ . This leads to:

**Definition 6.2.1** The function  $f : X \rightarrow X$  (or rather the dynamical system  $(X, f)$ ), is said to be (*topologically*) *transitive* if there exists  $x_0 \in X$  such that  $O(x_0)$  is a dense subset of  $X$ . A *transitive point* for  $f$  is a point  $x_0$  which has a dense orbit under  $f$ .

If  $f$  is transitive, then there is a dense set of transitive points, since each member of  $O(x_0)$  will be a transitive point.

**Example 6.2.2** The doubling map  $D : [0, 1] \rightarrow [0, 1]$  is transitive. To show this, we explicitly construct a point  $x_0 \in [0, 1]$  which has a dense orbit under  $D$ .  $x_0$  is defined using its binary expansion in the following way: first write down all possible “1-blocks”, i.e., 0 followed by 1. Then write down all possible “2-blocks”, i.e., 00, 01, 10, 11, then all possible ‘3-blocks’, i.e., 000, 001, 010, 011, 100, 101, 110, 111, and continue in this way with all possible “4-blocks” etc. (to be specific, we could write them down in the order in which they appear as in the binary expansion of the integers - lexicographical order). This gives:

$$x_0 = \cdot 01\ 00\ 01\ 10\ 11\ 000\ 001\ 010\ 011\ 100\ 101\ 110\ 111\ 0000\ 0001 \dots,$$

a point of  $[0, 1]$ .

To show that  $O(x_0)$  is dense in  $[0, 1]$ , let  $y \in [0, 1]$  with binary expansion

$$y = y_1 y_2 y_3 \dots = \sum_{i=1}^{\infty} \frac{y_i}{2^i}, \quad y_i = 0 \text{ or } 1,$$

and let  $\delta > 0$ .

Choose  $N$  so large that  $\frac{1}{2^N} < \delta$ . All possible finite strings of 0's and 1's appear in the binary expansion of  $x_0$ , so the string  $y_1 y_2 y_3 \dots y_N$  must also appear in the binary expansion of  $x_0$ .

It follows that for some  $r \in \mathbb{Z}^+$  we have

$$D^r(x_0) = \cdot y_1 y_2 y_3 \dots y_N b_{N+1} b_{N+2} \dots, \quad \text{for some } b_{N+1}, b_{N+2}, \dots,$$

so that

$$\begin{aligned} |D^r(x_0) - y| &= |\cdot y_1 y_2 \dots y_N b_{N+1} b_{N+2} \dots - \cdot y_1 y_2 \dots y_N y_{N+1} y_{N+2} \dots| \\ &\leq \sum_{i=N+1}^{\infty} \frac{1}{2^i} = \frac{1}{2^N} < \delta. \end{aligned}$$

This shows that any point  $y \in [0, 1]$  is arbitrarily close to the orbit of  $x_0$  under  $D$ , so the orbit of  $x_0$  is dense in  $[0, 1]$ . □

**Remark 6.2.3** Let  $(X, d)$  be a metric space that has no isolated points. This means that there is no point  $p \in X$  for which  $\{p\}$  is an open set. It is not too hard to see that for a continuous transitive map  $f : X \rightarrow X$ , and any non-empty open sets  $U$

and  $V$  in  $X$  there exists  $m \geq 1$  with

$$U \cap f^m(V) \neq \emptyset.$$

The converse of this statement holds for compact metric spaces (to be defined later). This result is known as the *Birkhoff Transitivity Theorem*, and will be proved in Chapter 17. The transitivity of the angle doubling map follows easily from this result.

### 6.3 Sensitive Dependence on Initial Conditions.

*Sensitive dependence on initial conditions* is the idea that small changes in a physical system can result in large changes down the road. For example, a butterfly flapping its wings in Japan can possibly affect the weather in the U.S. Popular literature sometimes calls this “the butterfly effect”.

Following, we give the definition of sensitive dependence on initial conditions for maps on a metric space.

**Definition 6.3.1** Let  $f : X \rightarrow X$  be defined on a metric space  $(X, d)$ . Then  $f$  has *sensitive dependence on initial conditions* if there exists  $\delta > 0$  such that for any  $x \in X$  and any open set  $U \subset X$  containing  $x$ , there is a point  $y \in U$  and  $n \in \mathbb{Z}^+$  with

$$d(f^n(x), f^n(y)) > \delta.$$

In other words, under iteration, points close to each other eventually move widely apart. A map has sensitive dependence on initial conditions if there exist points arbitrarily close to  $x$ , which are eventually at least distance  $\delta$  away from  $x$ . It is important to know whether we have sensitive dependence when doing computations, as round-off errors may be magnified after numerous iterations. For example, suppose we iterate the doubling map, starting with  $x_0 = 1/3$  and  $x_1 = .333$ . After 10 iterations we have  $D^{10}(x_0) = 1/3$  and  $D^{10}(x_1) = .992$ , more than distance  $1/2$  apart.

**Examples 6.3.2** 1. The linear map  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = ax$ ,  $|a| > 1$ , has sensitive dependence, for if  $x \neq y$ , then

$$|f^n(x) - f^n(y)| = |a^n x - a^n y| = a^n |x - y| \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

However, the dynamics of  $f$  are not complicated ( $f$  is not chaotic). As a specific example, consider Newton’s method applied to  $f(x) = x^{1/3}$ . The Newton function is  $N_f(x) = -2x$ , a linear function. If we iterate any non-zero initial guess  $x_0 \neq 0$ , it rapidly moves away from the zero of  $f$ .

2. The angle doubling map  $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $f(z) = z^2$  has sensitive dependence. If we iterate distinct points  $z = e^{i\theta}, w = e^{i\phi} \in \mathbb{S}^1$ , their distance apart doubles after each iteration.
3. The doubling map  $D : [0, 1] \rightarrow [0, 1]$  can be seen to have sensitive dependence on initial conditions. This follows from the fact that  $D'(x) = 2$  everywhere except at the point of discontinuity, so points that are close together are rapidly moved apart. In the next section we will show that the tent map  $T$  has sensitive dependence for similar reasons to that for the doubling map.

#### 6.4 The Definition of Chaos.

Devaney was the first to define the notion of chaos, saying that a continuous function is chaotic if it has a dense set of periodic points, is transitive, and also has sensitive dependence on initial conditions. Subsequently, it was shown that for continuous functions, the first two requirements imply the third. We define chaos for functions that may not be continuous as follows:

**Definition 6.4.1** Let  $f : X \rightarrow X$  be a function on the metric space  $X$ . Then  $f$  is said to be *chaotic* if:

- (i) The set of periodic points of  $f$  is dense in  $X$ .
- (ii)  $f$  is transitive.
- (iii)  $f$  has sensitive dependence on initial conditions.

**Examples 6.4.2** 1. Homeomorphisms and diffeomorphisms on an interval  $I \subseteq \mathbb{R}$  cannot be chaotic as they are never transitive. For example, if an order preserving homeomorphism  $f : [0, 1] \rightarrow [0, 1]$  has two fixed points  $0 < c_1 < c_2 < 1$ , and if  $c_1 < x < c_2$ , then  $c_1 < f^n(x) < c_2$ , for all  $n \geq 1$ , so the orbit of  $x$  is trapped between the fixed points. If the only fixed points are 0 and 1, then either  $f(x) > x$  for all  $x \in (0, 1)$ , or  $f(x) < x$  for all  $x \in (0, 1)$ , and again we can argue that  $f$  is not transitive, (in any case,  $f$  only has the two periodic points 0 and 1).

2. If  $f(x) = \sin x$ , then  $f$  has the single fixed point  $x = 0$  whose basin of attraction is all of  $\mathbb{R}$ . It follows that  $f$  cannot have any other periodic points:  $p$  being a periodic point is contradicted by  $f^n(p) \rightarrow 0$  as  $n \rightarrow \infty$ , so  $f$  is not chaotic. Similar considerations apply to functions such as  $\cos x$  and the logistic map  $L_\mu$  for  $0 < \mu < 3$ . We will show later that if  $f : I \rightarrow I$  is continuous ( $I = [a, b]$ , an interval), with

no period-2 points, then  $\lim_{n \rightarrow \infty} f^n(x)$  exists for all  $x \in I$ . So clearly transitivity is impossible.

3. We have shown that the doubling map  $D : [0, 1] \rightarrow [0, 1]$  has a dense set of periodic points and is transitive. In addition, in the last section, we showed that  $D$  has sensitive dependence on initial conditions. For these reasons, we regard it as a chaotic map.

4. The tent map  $T_2 = T : [0, 1] \rightarrow [0, 1]$  is chaotic. If  $x \in [0, 1]$  has a binary expansion  $x = \cdot a_1 a_2 a_3 \dots$ , then  $T(x) = \begin{cases} \cdot a_2 a_3 a_4 \dots; & a_1 = 0 \\ \cdot a'_2 a'_3 a'_4 \dots; & a_1 = 1, \end{cases}$  where  $a'_i = 1$  if  $a_i = 0$  and  $a'_i = 0$  if  $a_i = 1$ . More generally, we can see, using an induction argument from [119], that:

$$T^n(x) = \begin{cases} \cdot a_{n+1} a_{n+2} a_{n+3} \dots; & a_n = 0 \\ \cdot a'_{n+1} a'_{n+2} a'_{n+3} \dots; & a_n = 1. \end{cases}$$

We can use this to write down the periodic points of  $T$ . For example, the fixed points are  $x = 0$  and  $x = \cdot 1010\dots = 2/3$ , and the period 2-points are

$$x_1 = \cdot 01100110\dots = 2/5 \quad \text{and} \quad x_2 = \cdot 11001100\dots = 4/5.$$

The period 3-points are

$$\cdot 010010010010\dots = 2/7, \quad \cdot 100100100100\dots = 4/7, \quad \cdot 110110110110\dots = 6/7$$

and

$$\cdot 001110001110\dots = 2/9, \quad \cdot 011100011100\dots = 4/9, \quad \cdot 111000111000\dots = 8/9.$$

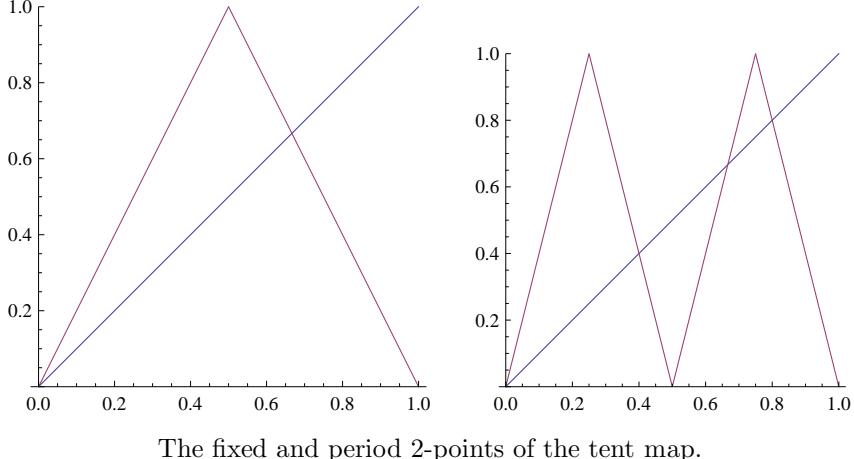
Notice that points of the form  $x = k/2^n \in (0, 1)$ ,  $k \in \mathbb{Z}^+$ , are almost fixed since they have a binary expansion of the form

$$x = \cdot a_1 a_2 a_3 \dots a_n 0 0 \dots,$$

where  $a_n = 0$  if  $k$  is even, and  $a_n = 1$  if  $k$  is odd. It follows that  $T^n(x) = 0$  if  $k$  is even, and  $T^n(x) = 1$  if  $k$  is odd. In particular

$$T^n \left[ \frac{k-1}{2^n}, \frac{k}{2^n} \right] = [0, 1].$$

The Intermediate Value Theorem implies that there is a fixed point of  $T^n$  in the interval  $[\frac{k-1}{2^n}, \frac{k}{2^n}]$ . Since such intervals can be made arbitrarily small and cover all of  $[0, 1]$ , the set of periodic points must be dense in  $[0, 1]$ .



The fixed and period 2-points of the tent map.

We use these ideas to show the following:

**Proposition 6.4.3** *The periodic points of the tent map  $T$  are those numbers in  $[0, 1]$  of the form  $x = r/s$  where  $r$  is an even integer and  $s$  is an odd integer.*

**Proof.** If  $x \in (0, 1)$  is a periodic point for  $T$ , then  $T^n(x) = x$  for some  $n \in \mathbb{Z}^+$ . There are two cases to consider. First suppose that  $x$  has a binary expansion

$$x = \cdot a_1 a_2 a_3 \dots a_n a_{n+1} \dots, \quad \text{where } a_n = 0,$$

then

$$T^n(x) = \cdot a_{n+1} a_{n+2} \dots a_{2n} \dots,$$

so we must have  $a_1 = a_{n+1}, a_2 = a_{n+2}, \dots, a_n = a_{2n} = 0$ , and

$$x = \cdot a_1 a_2 \dots a_n a_1 a_2 \dots a_n a_1 \dots, \quad \text{where } a_n = 0,$$

Rewriting gives

$$\begin{aligned} x &= \left( \frac{a_1}{2} + \frac{a_2}{2^2} + \dots + \frac{a_{n-1}}{2^{n-1}} \right) + \frac{1}{2^n} \left( \frac{a_1}{2} + \frac{a_2}{2^2} + \dots + \frac{a_{n-1}}{2^{n-1}} \right) + \dots \\ &= \left( \frac{a_1}{2} + \frac{a_2}{2^2} + \dots + \frac{a_{n-1}}{2^{n-1}} \right) \left( \frac{1}{1 - 1/2^n} \right) \\ &= \frac{a_1 2^{n-1} + a_2 2^{n-2} + \dots + a_{n-1} 2}{2^n - 1} = \frac{r}{s}, \end{aligned}$$

where  $r$  is even and  $s$  is odd.

In the second case,  $x = \cdot a_1 a_2 a_3 \dots a_n a_{n+1} \dots$ , where  $a_n = 1$ , so that

$$T^n(x) = \cdot a'_{n+1} a'_{n+2} \dots a'_{2n} \dots,$$

giving  $a_1 = a'_{n+1}$ ,  $a_2 = a'_{n+2}, \dots, a_n = a'_{2n}, \dots$ , and

$$x = \cdot a_1 a_2 \dots a_{n-1} 1 a'_1 a'_2 \dots a'_{n-1} 0 a_1 \dots$$

We complete the argument as before, but using the first  $2n$  terms of  $x$ .

Conversely, suppose that  $x = r/s \in (0, 1)$ , where  $r$  is an even integer and  $s$  is an odd integer. Since  $s$  and 2 are coprime, we can apply Euler's generalization of Fermat's Theorem to give

$$2^{\phi(s)} \equiv 1 \pmod{s}, \quad \text{or} \quad 2^p - 1 = ks \quad \text{for some } p, k \in \mathbb{Z}^+,$$

$(1 < kr < 2^p - 1)$ , where  $\phi$  is Euler's function. Write the binary expansion of  $kr$  (which is even) as

$$kr = a_1 2^{p-1} + a_2 2^{p-2} + \dots + a_{p-2} 2^2 + a_{p-1} 2, \quad a_i = 0 \text{ or } 1,$$

so

$$\begin{aligned} \frac{kr}{2^p - 1} &= (a_1 2^{p-1} + a_2 2^{p-2} + \dots + a_{p-2} 2^2 + a_{p-1} 2) \left( \frac{1/2^p}{1 - 1/2^p} \right) \\ &= \left( \frac{a_1}{2} + \frac{a_2}{2^2} + \dots + \frac{a_{p-1}}{2^{p-1}} \right) \left( \frac{1}{1 - 1/2^p} \right) \\ &= \cdot a_1 a_2 \dots a_{p-1} 0 a_1 a_2 \dots a_{p-1} 0 \dots, \end{aligned}$$

which is a point of period  $p$  (or less). □

Clearly  $T$  is transitive, since if we define  $x_0 \in (0, 1)$  having a binary expansion consisting of all 1-blocks, all 2-blocks, all 3-blocks etc., as before, except that we insert a single zero between every block, then it follows that if  $x = \cdot x_1 x_2 \dots x_n x_{n+1} \dots$ , then  $T^p(x_0) = \cdot x_1 x_2 \dots x_n b_{n+1} \dots$  for some  $p > 0$ , i.e., every block will appear in the iterates of  $x_0$ . Using a similar argument to that given earlier, we see that  $T$  is transitive. To show  $T$  is chaotic, we can show directly that  $T$  has sensitive dependence on initial conditions. Alternatively, we can use the fact that  $T$  is continuous, transitive, and has a dense set of periodic points, and then apply Theorem 6.6.1 (which is at the end of this chapter).

5. It is possible for a map to be transitive without being chaotic (although for continuous functions  $f$  on intervals in  $\mathbb{R}$ , this is not possible: see [124]). For example, consider the *irrational rotation*  $R_a : \mathbb{S}^1 \rightarrow \mathbb{S}^1$  defined by  $R_a(z) = a \cdot z$  for some (fixed)  $a \in \mathbb{S}^1$ . To say that  $R_a$  is an irrational rotation means that  $a^n \neq 1$  for any  $n \in \mathbb{Z}^+$ , i.e.,  $a$  is not an  $n$ th root of unity for any  $n \in \mathbb{Z}^+$ . We will show later that *every*  $z_0 \in \mathbb{S}^1$  has a dense orbit (a transformation with this property is said to be *minimal*).

However, suppose that  $R_a^n(z) = z$ , then  $a^n z = z$  or  $a^n = 1$ , a contradiction, so that  $R_a$  has no periodic points.  $R_a$  is an example of an *isometry*: points always stay the same distance apart:

$$|R_a(z) - R_a(w)| = |az - aw| = |a||z - w| = |z - w|.$$

Note that, if instead, we have  $a^n = 1$  for some  $n \in \mathbb{Z}^+$ , then  $R_a^n(z) = a^n z = z$  for all  $z \in \mathbb{S}^1$ , so that  $R_a^n$  is just the *identity map* (every point of  $\mathbb{S}^1$  is of period  $n$ ).

## 6.5 Symbolic Dynamics and the Shift Map.

Recall that the set of all one-sided infinite sequences of 0's and 1's, is denoted by  $\mathcal{A}^{\mathbb{Z}^+}$  where  $\mathcal{A} = \{0, 1\}$ . Denote this set by  $\Sigma$ , so

$$\Sigma = \{(s_1, s_2, s_3, \dots) : s_i = 0 \text{ or } 1, i \in \mathbb{Z}^+\},$$

a metric space with metric defined by

$$d(\omega_1, \omega_2) = \sum_{k=1}^{\infty} \frac{|s_k - t_k|}{2^k}, \quad \text{where } \omega_1 = (s_1, s_2, \dots), \quad \omega_2 = (t_1, t_2, \dots) \in \Sigma.$$

This metric has the following properties:

**Proposition 6.5.1** *If  $\omega_1 = (s_1, s_2, \dots)$ ,  $\omega_2 = (t_1, t_2, \dots) \in \Sigma$ , with  $s_i = t_i$ ,  $i = 1, 2, \dots, n$ , then  $d(\omega_1, \omega_2) \leq 1/2^n$ .*

**Proof.**

$$d(\omega_1, \omega_2) = \sum_{k=1}^{\infty} \frac{|s_k - t_k|}{2^k} = \sum_{k=n+1}^{\infty} \frac{|s_k - t_k|}{2^k} \leq \sum_{k=n+1}^{\infty} \frac{1}{2^k} = \frac{1}{2^n}.$$

□

**Proposition 6.5.2** *If  $d(\omega_1, \omega_2) < 1/2^n$ , then  $s_i = t_i$  for  $i = 1, 2, \dots, n$ .*

**Proof.** We give a proof by contradiction. Suppose that  $s_j = t_j$  for some  $1 \leq j \leq n$ , then

$$d(\omega_1, \omega_2) = \sum_{k=1}^{\infty} \frac{|s_k - t_k|}{2^k} \geq \frac{1}{2^j} \geq \frac{1}{2^n},$$

a contradiction.

□

The shift map  $\sigma$  (known as the *Bernoulli shift*), is an important function defined on  $\Sigma$ .

**Definition 6.5.3** The *Bernoulli shift*  $\sigma : \Sigma \rightarrow \Sigma$  is defined by

$$\sigma(s_1, s_2, s_3, \dots) = (s_2, s_3, \dots).$$

For example,  $\sigma(1, 0, 1, 0, \dots) = (0, 1, 0, 1, \dots)$  and  $\sigma^2(1, 0, 1, 0, \dots) = (1, 0, 1, 0, \dots)$ . If  $\omega_1 = (1, 0, 1, 0, \dots)$  and  $\omega_2 = (0, 1, 0, 1, \dots)$ , then  $\{\omega_1, \omega_2\}$  is a 2-cycle for  $\sigma$ . In this way, it is easy to write down all of the points of period  $n$ . Any sequence which is eventually constant is clearly an eventually fixed point of  $\sigma$ , and any sequence which is eventually periodic (such as  $(1, 1, 1, 0, 1, 0, 1, 0, 1, \dots)$ ), is an eventually periodic point.

**Proposition 6.5.4** *The shift map  $\sigma : \Sigma \rightarrow \Sigma$  is continuous and onto, but is not one-to-one.*

**Proof.** Clearly  $\sigma$  is onto but not one-to-one. To show that  $\sigma$  is continuous, let  $\epsilon > 0$ , then we want to find  $\delta > 0$  such that if

$$d(\omega_1, \omega_2) < \delta, \quad \text{then} \quad d(\sigma(\omega_1), \sigma(\omega_2)) < \epsilon.$$

It suffices to take  $\delta = 1/2^{n+1}$  if  $n$  is chosen so large that  $1/2^n < \epsilon$ .

In this case, if  $d(\omega_1, \omega_2) < \delta = 1/2^{n+1}$ , then from Proposition 6.5.2,  $s_i = t_i$  for  $i = 1, 2, \dots, n + 1$ . Clearly the first  $n$  terms of the sequences  $\sigma(\omega_1)$  and  $\sigma(\omega_2)$  are equal, so by Proposition 6.5.1,  $d(\sigma(\omega_1), \sigma(\omega_2)) \leq 1/2^n < \epsilon$ , so that  $\sigma$  is continuous.  $\square$

We can now prove:

**Theorem 6.5.5** *The Bernoulli shift  $\sigma : \Sigma \rightarrow \Sigma$  is chaotic.*

**Proof.** We first show that the periodic points are dense in  $\Sigma$ . Let  $\omega = (s_1, s_2, \dots) \in \Sigma$ . It is sufficient to show that there is a sequence of periodic points  $\omega_n \in \Sigma$  with  $\omega_n \rightarrow \omega$  as  $n \rightarrow \infty$ . Set

$\omega_1 = (s_1, s_1, s_1, s_1, \dots)$ , a period 1-point for  $\sigma$ ,

$\omega_2 = (s_1, s_2, s_1, s_2, \dots)$ , a period 2-point for  $\sigma$ ,

$\omega_3 = (s_1, s_2, s_3, s_1, s_2, s_3, \dots)$ , a period 3-point for  $\sigma$

Continue in this way (these points may have lesser period):

$\omega_n = (s_1, s_2, \dots, s_n, s_1, \dots)$ , a period  $n$ -point for  $\sigma$ .

Since  $\omega$  and  $\omega_n$  agree in the first  $n$  coordinates,  $d(\omega, \omega_n) \leq 1/2^n$ , so  $d(\omega, \omega_n) \rightarrow 0$  as  $n \rightarrow \infty$ , or  $\omega_n \rightarrow \omega$  as  $n \rightarrow \infty$ .

To show that  $\sigma$  is transitive, we explicitly construct a point  $\omega_0 \in \Sigma$  having a dense orbit under  $\sigma$ . Let

$$\omega_0 = (\underbrace{01}_{\text{1-blocks}} \underbrace{00011011}_{\text{all possible 2-blocks}} \underbrace{000001010101011\dots}_{\text{all possible 3 blocks}}).$$

Continue in this way so that all possible  $n$ -blocks appear in  $\omega_0$ . To see that  $\overline{\Omega(\omega_0)} = \Sigma$ , let  $\omega = (s_1, s_2, s_3, \dots) \in \Sigma$  be arbitrary. Let  $\epsilon > 0$  and choose  $n$  so large that  $1/2^n < \epsilon$ . Since  $\omega_0$  consists of all possible  $n$ -blocks, the sequence  $(s_1, s_2, \dots, s_n)$  must appear somewhere in  $\omega_0$ , i.e., there exists  $k > 0$  with

$$\sigma^k(\omega_0) = (s_1, s_2, \dots, s_n, \dots),$$

so that  $\omega$  and  $\sigma^k(\omega_0)$  agree on the first  $n$  coordinates. It follows that

$$d(\omega, \sigma^k(\omega_0)) \leq \frac{1}{2^n} < \epsilon.$$

This shows that the orbit of  $\omega_0$  comes arbitrarily close to any member of  $\Sigma$ , so it is dense in  $\Sigma$ . It follows that  $\sigma$  is transitive.

We now show that  $\sigma$  has sensitive dependence on initial conditions. If  $\omega_1 = (s_1, s_2, \dots), \omega_2 = (t_1, t_2, \dots) \in \Sigma$  with  $\omega_1 \neq \omega_2$ , then  $\omega_1$  and  $\omega_2$  must differ at some coordinate, say  $s_i \neq t_i$ . Thus

$$\sigma^{i-1}(\omega_1) = (s_i, s_{i+1}, \dots), \quad \text{and} \quad \sigma^{i-1}(\omega_2) = (t_i, t_{i+1}, \dots),$$

so that

$$d(\sigma^{i-1}(\omega_1), \sigma^{i-1}(\omega_2)) = \sum_{k=1}^{\infty} \frac{|s_{i+k-1} - t_{i+k-1}|}{2^k} = \frac{1}{2} + \text{other terms} \geq \frac{1}{2}.$$

### Exercises 6.5

1. Use induction to show that if  $T_2 : [0, 1] \rightarrow [0, 1]$  is the standard tent map, and  $x \in [0, 1]$  has binary expansion  $x = \cdot a_1 a_2 a_3 \dots$ ,  $a_i = 0$  or  $1$ , then

$$T_2^n(x) = \begin{cases} \cdot a_{n+1} a_{n+2} a_{n+3} \dots; & a_n = 0 \\ \cdot b_{n+1} b_{n+2} b_{n+3} \dots; & a_n = 1 \end{cases},$$

where  $b_i = 0$  if  $a_i = 1$ , and  $b_i = 1$  if  $a_i = 0$ . Deduce a 5-cycle for  $T_2$ .

2. (a) Let  $\Sigma = \{(a_1, a_2, a_3, \dots) : a_i = 0 \text{ or } 1\}$ , the sequence space of zeros and ones with the metric defined as in Examples 4.1.3. Let  $C$  be the Cantor set, and define a map  $f : \Sigma \rightarrow C$  by

$$f(a_1, a_2, a_3, \dots) = \cdot b_1 b_2 b_3 \dots, \quad \text{where } b_i = 0 \text{ if } a_i = 0 \text{ and } b_i = 2 \text{ if } a_i = 1,$$

giving the ternary expansion of a real number in  $[0, 1]$ . Show that  $f$  defines a homeomorphism between  $\Sigma$  and the Cantor set.

- (b) It can be shown that  $[0, 1]$  and  $C$  are not homeomorphic (for example,  $C$  is totally disconnected whilst  $[0, 1]$  is not). What goes wrong if we try to define a homeomorphism between  $[0, 1]$  and  $C$  by mapping the binary expansion of  $x \in [0, 1]$ , to the corresponding ternary expansion?

3. Let  $f : I \rightarrow I$  be a transitive map, where  $I$  is an interval. Show that if  $U$  and  $V$  are non-empty open sets in  $I$ , there exists  $m \in \mathbb{Z}^+$  with  $U \cap f^m(V) \neq \emptyset$ .

4. Let  $F : [0, 1] \rightarrow [0, 1]$  be the tripling map  $F(x) = 3x \bmod 1$ . Follow the proof for the doubling map (but use ternary expansions), to show that  $F$  is transitive and the period points are dense (find the periodic points).

5. Let  $D : [0, 1] \rightarrow [0, 1]$  be the doubling map. Show that  $|\text{Per}_n(D)| = 2^n - 1$ . (Recall that  $\text{Per}_n(D)$  consists of all  $x \in [0, 1]$  with  $f^n(x) = x$ ).

6. Let  $f_\lambda : [0, 1] \rightarrow [0, 1]$  be defined by  $f_\lambda(x) = 1 - |2x - 1|^\lambda$ , for  $0 < \lambda \leq 4$ .

- (a) Show that  $f_1 = T$ , the tent map, and  $f_2 = L_4$ , the logistic map.

- (b) Show that  $f_\lambda$  cannot be chaotic for  $\lambda < 1/2$ . (Hint: Examine the fixed point  $x = 0$ ).

- (c) Use a computer algebra system to do a manipulate plot for  $0 < \lambda \leq 4$ . Find the fixed points when  $\lambda = 1/2$  and when  $\lambda = 1/3$ .

## 6.6 For Continuous Maps, Sensitive Dependence is Implied by Transitivity and Dense Period Points.

We will now show that the original definition of chaos due to Devaney [32] can be simplified in the case of continuous mappings. Devaney's definition required *sensitive dependence on initial conditions*.

The following theorem, due to Banks, Brooks, Cairns, Davis and Stacey ([8], 1992), gives the promised implication that for continuous functions on a metric space, sensitivity of initial conditions follows from the periodic points being dense, together with transitivity. We need to assume that  $X$  is not a finite set as in this case the dynamical system could consist of a single periodic orbit. We shall see that no other two of these conditions imply the third.

**Theorem 6.6.1** *Let  $(X, d)$  be an infinite metric space. If  $f : X \rightarrow X$  is a continuous function which is transitive, and has a dense set of periodic points, then  $f$  has sensitive dependence on initial conditions.*

We first prove a preliminary result:

**Lemma 6.6.2** *Let  $f : X \rightarrow X$  be a transformation which has at least two different periodic orbits. Then there exists  $\epsilon > 0$  such that for any  $x \in X$  there is a periodic point  $p$  satisfying*

$$d(x, f^k(p)) > \epsilon, \quad \text{for all } k \in \mathbb{Z}^+.$$

**Proof.** Let  $a$  and  $b$  be two periodic points with different orbits. Then  $d(f^k(a), f^l(b)) > 0$  for all  $k$  and  $l$  (since we are dealing with finite sets).

Choose  $\epsilon > 0$  small enough that  $d(f^k(a), f^l(b)) > 2\epsilon$  for all  $k$  and  $l$ . Then if  $x \in X$ ,

$$d(f^k(a), x) + d(x, f^l(b)) \geq d(f^k(a), f^l(b)) > 2\epsilon \quad \forall k, l \in \mathbb{Z}^+,$$

by the triangle inequality.

If  $x$  is within  $\epsilon$  of any of the points  $f^l(b)$ , then it must be at a greater distance than  $\epsilon$  from all of the points  $f^k(a)$ . We choose  $x$  accordingly, and the result follows.  $\square$

**Proof of Theorem 6.6.1** Let  $x \in X$  and  $U$  be an open set in  $X$  containing  $x$ .

Let  $p$  be a periodic point of period  $r$  for  $f$ , whose orbit is a distance greater than  $4\delta$  from  $x$ .

The periodic points of  $f$  are dense in  $X$ , so there is a periodic point  $q$  of period  $n$  say, with

$$q \in V = U \cap B_\delta(x).$$

Write

$$W_i = B_\delta(f^i(p)),$$

then

$$f^i(p) \in W_i, \forall i \Rightarrow p \in f^{-i}(W_i), \forall i.$$

The continuity of  $f$  implies that the set

$$W = f^{-1}(W_1) \cap f^{-2}(W_2) \cap \cdots \cap f^{-n}(W_n)$$

is open, and from above, it is non-empty. Since  $f$  is transitive, there is a point  $z \in V$  with  $f^k(z) \in W$  for some  $k \in \mathbb{Z}^+$ .

Let  $j$  be the smallest integer with  $k < nj$ , or

$$1 \leq nj - k \leq n.$$

Then

$$f^{nj}(z) = f^{nj-k}(f^k(z)) \in f^{nj-k}(W).$$

But

$$f^{nj-k}(W) = f^{nj-k}(f^{-1}(W_1) \cap f^{-2}(W_2) \cap \cdots \cap f^{-n}(W_n)) \subset f^{nj-k}(f^{-(nj-k)}W_{nj-k}) = W_{nj-k},$$

so that  $d(f^{nj}(z), f^{nj-k}(p)) < \delta$ . Now  $f^{nj}(q) = q$ , and by the triangle inequality

$$d(f^{nj-k}(p), x) \leq d(f^{nj-k}(p), f^{nj}(z)) + d(f^{nj}(z), f^{nj}(q)) + d(f^{nj}(q), x)$$

so that

$$\begin{aligned} 4\delta &< d(f^{nj-k}(p), x) \leq d(f^{nj-k}(p), f^{nj}(z)) + d(f^{nj}(z), f^{nj}(q)) + d(q, x) \\ &< \delta + d(f^{nj}(z), f^{nj}(q)) + \delta. \end{aligned}$$

It follows that

$$d(f^{nj}(z), f^{nj}(q)) > 2\delta.$$

The above inequality implies that either

$$d(f^{nj}(x), f^{nj}(z)) \geq \delta,$$

or

$$d(f^{nj}(x), f^{nj}(q)) \geq \delta,$$

for if  $f^{nj}(x)$  were within distance  $< \delta$  from both of these points, the points would have to be within  $< 2\delta$  from each other, contradicting the inequality above. We see that one of the two,  $z$  or  $q$ , will serve as the  $y$  in the theorem with  $m = nj$ .

□

**Remarks 6.6.3** The following theorem, due Vellekoop and Berglund [124], shows that if a continuous real function is transitive, then it is chaotic:

**Theorem 6.6.4** *If  $f : I \rightarrow I$  is a continuous and transitive map on an interval  $I \subseteq \mathbb{R}$ , then the set of periodic points of  $f$  is dense in  $I$ , and hence  $f$  is chaotic on  $I$ .*

**Remark 6.6.5** It can be shown that (topological) properties such as being totally disconnected, perfect etc. are preserved by homeomorphisms.

The shift space  $\Sigma$  and the Cantor set  $C$  are homeomorphic metric spaces. In addition, the interval  $I = [0, 1]$  with the points 0 and 1 identified, and the unit circle  $\mathbb{S}^1$  in the complex plane are homeomorphic. In particular,  $\Sigma$  and  $C$  will have identical topological properties, and so will  $\mathbb{S}^1$  and  $I$ . It follows that  $\Sigma$  and  $I$  cannot be homeomorphic as  $C$  and  $I$  are not homeomorphic ( $C$  is totally disconnected, but  $I$  is not). Note that without identifying the end points of  $I$ ,  $I$  and  $\mathbb{S}^1$  will not be homeomorphic.

**Proof.**  $\Sigma$  is given its usual metric, and  $C$  has the metric induced from being a subset of  $\mathbb{R}$ , so that  $d(x, y) = |x - y|$  for  $x, y \in C$ .

We define a map  $h : C \rightarrow \Sigma$  by  $h(\cdot a_1 a_2 a_3 \dots) = (s_1, s_2, s_3, \dots)$ , where  $a_i = 0$  or 2 and  $s_i = a_i/2$ . Clearly  $h$  is both one-to-one and onto. We show that it is continuous at each  $x_0 \in C$ .

Let  $\epsilon > 0$  and choose  $n$  so large that  $1/2^n < \epsilon$ . Set  $\delta = 1/3^n$ , then if  $|x_0 - x| < \delta$ , both  $x_0$  and  $x$  must lie in the same component (sub-interval) of  $S_n$  of length  $1/3^n$ .  $x_0$  and  $x$  must have an identical ternary expansions in the first  $n$  places. Correspondingly,  $h(x_0)$  and  $h(x)$  must have the same first  $n$  coordinates. Consequently  $d(h(x_0), h(x)) \leq 1/2^n < \epsilon$ , so  $h$  is continuous at  $x_0$ . In a similar way we see that  $h^{-1}$  is continuous.

To see that  $I$  and  $\mathbb{S}^1$  are homeomorphic metric spaces, define  $h : I \rightarrow \mathbb{S}^1$  by  $h(x) = e^{2\pi i x}$ . Then  $h$  is one-to-one and onto (since we are identifying the end points of  $I$ ). The map  $h$  wraps the interval  $[0, 1]$  around the circle with  $h(0) = h(1)$ . In this way  $h$  becomes continuous.

□

## Exercises 6.6

1. Let  $T_3 : C \rightarrow C$  be the tent map  $T_3(x) = \begin{cases} 3x; & x < 1/2 \\ 3(1-x); & x \geq 1/2 \end{cases}$ , but restricted to the Cantor set  $C$ .

- (a) Using the formula from Section 5.4 and induction, show that if  $x = \cdot a_1 a_2 a_3 \dots$  is the ternary expansion of  $x \in C$ , then  $T_3^n(\cdot a_1 a_2 a_3 \dots) = \begin{cases} \cdot a_{n+1} a_{n+2} a_{n+3} \dots; & a_n = 0 \\ \cdot b_{n+1} b_{n+2} b_{n+3} \dots; & a_n = 2 \end{cases}$  where  $b_i = 0$  if  $a_i = 2$ , and  $b_i = 2$  if  $a_i = 0$ . Deduce the period-3 and period-5 points of  $T_3$ .
- (b) Show that the periodic points of  $T_3$  are dense in  $C$ .
- (c) Show that  $T_3$  is transitive on  $C$ .
- (d) Deduce that  $T_3$  is chaotic on  $C$ . (Hint: Use that fact that if  $f : X \rightarrow X$  is a continuous map on the metric space  $X$ , and  $A \subset X$  is a subspace of  $X$  invariant under  $f$ , then  $f : A \rightarrow A$  is continuous).

2. A piecewise linear function  $f : [0, 1] \rightarrow [0, 1]$  is defined so that  $f(0) = 1/2$ ,  $f(1/2) = 1$  and  $f(1) = 0$ , and then extended to be linear on  $[0, 1/2]$  and  $[1/2, 1]$ . Thus

$$f(x) = \begin{cases} x + 1/2; & 0 \leq x < 1/2 \\ 2 - 2x; & 1/2 \leq x \leq 1. \end{cases}$$

- (a) Write down what  $f$  does to  $x \in [0, 1/2]$  and  $x \in [1/2, 1]$ , where  $x = \cdot a_1 a_2 a_3 \dots$ ,  $a_i = 0$  or 1, is the binary expansion of  $x$ .
- (b) Note that  $\{0, 1/2, 1\}$  is a 3-cycle and that  $x = \cdot 101010 \dots = \overline{10}$  is a fixed point. Find the period-2 points.
- (c) Find the points of period 4 and 6, and deduce the points of period  $2 + 2k$ .
- (d) Find the points of period  $3 + 2k$  (see [119] and also [11] and [70]).

## CHAPTER 7

### Conjugacy of Dynamical Systems.

Two metric spaces  $X$  and  $Y$  are the “same” (homeomorphic) if there is a homeomorphism from one space to the other. In this chapter we study when two *dynamical systems* are the same. Given maps  $f : X \rightarrow X$  and  $g : Y \rightarrow Y$ , we require them to have the same type of dynamical behavior, e.g., there should be a one-to-one correspondence between their respective periodic points, if one is chaotic, then so is the other etc. One obvious requirement is that there should be a homeomorphism  $h : X \rightarrow Y$  between the underlying metric spaces. It is natural to require that this homeomorphism intertwine  $f$  and  $g$  in the sense that  $h \circ f = g \circ h$ . We have seen many similarities between the logistic map  $L_4(x) = 4x(1 - x)$  and the tent map  $T_2$ , and this will be examined in this chapter together with other examples, such as the shift map, and the angle doubling map. In particular, we use these ideas to show that the logistic map  $L_4$  is chaotic.

#### 7.1 Conjugate Maps.

This “sameness” is given by the idea of *conjugacy*, a notion borrowed from group theory, where two members  $a$  and  $b$  of a group  $G$  are *conjugate* if there exists  $g \in G$  with  $ag = gb$ . One of the central problems of one-dimensional dynamics and dynamical systems in general, is being able to tell whether or not two dynamical systems are conjugate. We will see that if one map has a 3-cycle, and another map has no 3-cycle (for example), then the maps cannot be conjugate. Also, if one map has 2 fixed points and the other has 3 fixed points, then the maps are not conjugate. These are examples of *conjugacy invariants*, which give criteria for maps to be non-conjugate. A generally more difficult problem is deciding if conjugacies exist between maps which have very similar dynamical properties.

**Definition 7.1.1** 1. Let  $f : X \rightarrow X$  and  $g : Y \rightarrow Y$  be maps of metric spaces. Then  $f$  and  $g$  are said to be *conjugate* if there is a homeomorphism  $h : X \rightarrow Y$  such that

$$h \circ f = g \circ h.$$

Strictly speaking, we mean that the dynamical systems,  $(X, f)$  and  $(Y, g)$  are conjugate, but we will often talk of conjugacy between the corresponding maps, so that the map  $h$  is a *conjugacy* between  $f$  and  $g$ . Obviously, conjugacy is an equivalence relation.

2. In the above definition, if instead of requiring  $h$  be a homeomorphism, we only require  $h : X \rightarrow Y$  to be continuous and onto, then we say that  $g$  is a *factor* of  $f$ .

**Examples 7.1.2** 1. If  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  are defined by  $f(x) = 2x(1 - x)$  and  $g(x) = x^2$ , then we can check that  $f$  and  $g$  are conjugate via  $h(x) = -2x + 1$  as follows:

$$h(f(x)) = h(2x(1 - x)) = -4x(1 - x) + 1 = 4x^2 - 4x + 1$$

and

$$g(h(x)) = g(-2x + 1) = (-2x + 1)^2 = 4x^2 - 4x + 1,$$

so that  $h \circ f = g \circ h$ . Since  $h$  is a homeomorphism of  $\mathbb{R}$ ,  $f$  and  $g$  are conjugate. We can also check that  $h : [0, 1] \rightarrow [-1, 1]$  gives a conjugacy between  $f : [0, 1] \rightarrow [0, 1]$  and  $g : [-1, 1] \rightarrow [-1, 1]$ .

2. Define  $f_a : \mathbb{R} \rightarrow \mathbb{R}$  by  $f_a(x) = ax$ , for  $a \in \mathbb{R}$ . If  $h(x) = x^{1/3}$ , then

$$h(f_8(x)) = h(8x) = (8x)^{1/3} = 2x^{1/3}, \quad \text{and} \quad f_2(h(x)) = f_2(x^{1/3}) = 2x^{1/3},$$

so  $f_2$  and  $f_8$  are conjugate, since  $h$  is a homeomorphism.

3. The logistic map  $L_4 : [0, 1] \rightarrow [0, 1]$  is a factor of the doubling map  $D : [0, 1] \rightarrow [0, 1]$ . If we define  $h : [0, 1] \rightarrow [0, 1]$  by  $h(x) = \sin^2(\pi x)$ , then  $h$  is continuous and onto (since  $h(0) = 0$  and  $h(1/2) = 1$ ). Note that  $h$  is not one-to-one. Let us check that

$$h(D(x)) = L_4(h(x)) \quad \text{for all } x \in [0, 1].$$

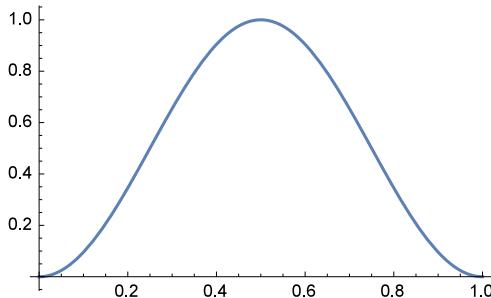
We have

$$h(D(x)) = \begin{cases} \sin^2(2\pi x); & 0 \leq x < 1/2 \\ \sin^2(\pi(2x - 1)); & 1/2 \leq x < 1 \end{cases} = \sin^2(2\pi x),$$

and

$$L_4(h(x)) = 4 \sin^2(\pi x)(1 - \sin^2(\pi x)) = 4 \sin^2(\pi x) \cos^2(\pi x) = \sin^2(2\pi x),$$

and the result follows.



The factor map  $h : [0, 1] \rightarrow [0, 1]$ .

### Exercises 7.1

1. Prove that conjugacy is an *equivalence relation*, i.e., show that:
  - (i) A function  $f : X \rightarrow X$  is conjugate to itself.
  - (ii) If  $f : X \rightarrow X$  is conjugate to  $g : Y \rightarrow Y$ , then  $g$  is conjugate to  $f$ .
  - (iii) If  $f : X \rightarrow X$  is conjugate to  $g : Y \rightarrow Y$ , and  $g$  is conjugate to  $k : Z \rightarrow Z$ , then  $f$  is conjugate to  $k$ .
  
2. (a) Define  $f_a : \mathbb{R} \rightarrow \mathbb{R}$  by  $f_a(x) = ax$ , for  $a \in \mathbb{R}$ . Show that  $f_{1/2}$  and  $f_{1/4}$  are conjugate via the map  $h(x) = \begin{cases} \sqrt{x}; & x \geq 0 \\ -\sqrt{-x}; & x < 0 \end{cases}$ .
   
 (b) More generally, show that if  $f_a, f_b : [0, \infty) \rightarrow [0, \infty)$ , ( $0 < a, b < 1$ ), then  $f_a$  and  $f_b$  are conjugate via a map of the form  $h(x) = x^p$ ,  $p > 0$ , and similarly if  $a, b > 1$ .
   
 (c) Discuss the cases where  $a > 1$  and  $0 < b < 1$ . What happens when  $a = 1/2$  and  $b = 2$ ?
  
3. Prove that if  $f : X \rightarrow X$  and  $g : Y \rightarrow Y$  are conjugate maps of metric spaces, then  $f$  is one-to-one if and only if  $g$  is one-to-one, and  $f$  is onto if and only if  $g$  is onto.
  
4. Prove that if  $f$  and  $g$  are conjugate via  $h$ , and  $f$  has a local maximum at  $x_0$ , then  $g$  has a local maximum or minimum at  $h(x_0)$ .

5. Suppose that  $h : [0, 1] \rightarrow [0, 1]$  is a conjugacy between  $f, g : [0, 1] \rightarrow [0, 1]$  where  $f(0) = f(1) = 0$ , and  $g(0) = g(1) = 0$ . Show that  $h$  is increasing on  $[0, 1]$ . Deduce that  $h$  maps the zero's of  $f$  to the zero's of  $g$ .
6. (a) Let  $f : X \rightarrow X$  be a homeomorphism on a metric space  $X$ , with inverse  $f^{-1}$ . Prove that  $f$  is conjugate to  $f^{-1}$  via an involution  $g$  (i.e.,  $f \circ g = g \circ f^{-1}$  where  $g^2(x) = x$  for all  $x \in X$ ), if and only if  $f$  is the product of two involutions ( $f = h \circ k$  where  $h^2 = k^2 = \text{identity map}$ ).
- (b) Let  $f(x) = 1/(3-x)$ . By writing  $f$  as the product of two involutions, show that  $f$  is conjugate to  $f^{-1}$ . (We can think of  $f$  as a homeomorphism of  $\mathbb{R} \cup \{\infty\}$ , by setting  $f(3) = \infty$  and  $f(\infty) = 0$ ).
7. The function  $T_n(x) = \cos(n \arccos(x))$  is the  $n$ th *Chebyshev polynomial*. Show that  $T_n$  is conjugate to the map  $\Lambda_n : [0, 1] \rightarrow [0, 1]$ , the piecewise linear continuous map defined by joining the points  $(0, 0), (1/n, 1), (2/n, 0), (3/n, 1), \dots$ , ending with  $(1, 1)$  if  $n$  is odd, or  $(1, 0)$  if  $n$  is even. Use the conjugacy map  $h : [0, 1] \rightarrow [0, 1]$ ,  $h(x) = \cos(\pi x)$ . (See [21], where there is a generalization of this to maps  $T_\lambda$ , where  $\lambda > 1$  is a real number).

## 7.2 Properties of Conjugate Maps and Chaos Through Conjugacy.

It is often easier to show indirectly that certain dynamical systems are chaotic, by showing that they are conjugate to chaotic systems, and using the following result:

**Proposition 7.2.1** *Let  $f : X \rightarrow X$  and  $g : Y \rightarrow Y$  be maps of metric spaces. If there is a conjugacy  $h : X \rightarrow Y$ :  $h \circ f = g \circ h$ , then*

1.  $h \circ f^n = g^n \circ h$  for all  $n \in \mathbb{Z}^+$ , (so  $f^n$  and  $g^n$  are also conjugate).
2. If  $c$  is a point of period  $m$  for  $f$ , then  $h(c)$  is a point of period  $m$  for  $g$ .  $c$  is attracting if and only if  $h(c)$  is attracting.
3.  $f$  is transitive if and only if  $g$  is transitive.
4.  $f$  has a dense set of periodic points if and only if  $g$  has a dense set of periodic points.
5. Let  $f$  and  $g$  be continuous maps. Then  $f$  is chaotic if and only if  $g$  is chaotic.

**Proof.** 1.  $h \circ f^2 = h \circ f \circ f = g \circ h \circ f = g \circ g \circ h = g^2 \circ h$ , and in the same way  $h \circ f^3 = g^3 \circ h$ , and continuing inductively the result follows.

2. Suppose that  $f^i(c) \neq c$  for  $0 < i < m$  and  $f^m(c) = c$ . Then  $h \circ f^i(c) \neq h(c)$  for  $0 < i < m$  since  $h$  is one-to-one, and so  $g^i \circ h(c) \neq h(c)$  for  $0 < i < m$ . In addition,  $h \circ f^m(c) = g^m \circ h(c)$ , or  $h(c) = g^m(h(c))$ , so  $h(c)$  is a period- $m$  point for  $g$ .

We shall show only that if  $p$  is an attracting fixed point of  $f$  (so that there is an open ball  $B_\epsilon(p)$  such that if  $x \in B_\epsilon(p)$  then  $f^n(x) \rightarrow p$  as  $n \rightarrow \infty$ ), then  $h(p)$  is an attracting fixed point of  $g$ .

Let  $V = h(B_\epsilon(p))$ . Then since  $h$  is a homeomorphism,  $V$  is open in  $Y$  and contains  $h(p)$ . Let  $y \in V$ , then  $h^{-1}(y) \in B_\epsilon(p)$ , so that  $f^n(h^{-1}(y)) \rightarrow p$  as  $n \rightarrow \infty$ .

Since  $h$  is continuous,  $h(f^n(h^{-1}(y))) \rightarrow h(p)$  as  $n \rightarrow \infty$ , i.e.,

$$g^n(y) = h \circ f^n \circ h^{-1}(y) \rightarrow h(p), \quad \text{as } n \rightarrow \infty,$$

and this holds for any  $y \in V$ . Finally, for any ball  $B_\delta(h(p)) \subset V$  with  $\delta > 0$ , we must have  $g^n(y) \rightarrow h(p)$ , for all  $y \in B_\delta(h(p))$ , as  $n \rightarrow \infty$ . So  $h(p)$  is an attracting fixed point.

3. Suppose that  $O(z) = \{z, f(z), f^2(z), \dots\}$  is dense in  $X$ . Let  $V \subset Y$  be a non-empty open set. Since  $h$  is a homeomorphism,  $h^{-1}(V)$  is open in  $X$ , so there exists  $k \in \mathbb{Z}^+$  with  $f^k(z) \in h^{-1}(V)$ .

It follows that  $h(f^k(z)) = g^k(h(z)) \in V$ , so that

$$O(h(z)) = \{h(z), g(h(z)), g^2(h(z)), \dots\}$$

is dense in  $Y$ , i.e.,  $g$  is transitive. Similarly, if  $g$  is transitive, then  $f$  is transitive.

4. Suppose that  $f$  has a dense set of periodic points. Let  $V \subset Y$  be non-empty and open. Then  $h^{-1}(V)$  is open in  $X$ , and so contains periodic points of  $f$ . As in (3), we see that  $V$  contains periodic points of  $g$ . Similarly, if  $g$  has a dense set of periodic points, so does  $f$ .

5. If  $f$  is continuous and chaotic, it is both transitive and has a dense set of periodic points. From (3) and (4), the same is true for  $g$ . Now Theorem 6.6.1 implies that  $g$  is chaotic.

□

**Example 7.2.2** We remark that sensitive dependence on initial conditions is not a conjugacy invariant. It is possible for two maps on metric spaces to be conjugate, one to have sensitive dependence, but the other not: Consider  $T : (0, \infty) \rightarrow (0, \infty)$ ,

$T(x) = 2x$  and  $S : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $S(x) = x + \ln 2$ . If  $H : (0, \infty) \rightarrow \mathbb{R}$  is defined by  $H(x) = \ln x$ , then  $H$  is a homeomorphism, and we can check that  $H \circ T = S \circ H$ , so  $T$  and  $S$  are conjugate.  $T$  has sensitive dependence, but  $S$  does not.

It can be shown however, that if  $T : X \rightarrow X$  is a continuous map on a *compact metric space*  $X$  (for example  $X = [0, 1]$  - see Chapter 17), having sensitive dependence, then any map conjugate to  $T$  also has sensitive dependence.

It can also be shown that the property of having negative Schwarzian derivative is not a conjugacy invariant.

**Examples 7.2.3** 1. *The logistic map*  $L_4 : [0, 1] \rightarrow [0, 1]$ ,  $L_4(x) = 4x(1-x)$  is conjugate to the tent map  $T : [0, 1] \rightarrow [0, 1]$ ,  $T(x) = \begin{cases} 2x; & 0 \leq x \leq 1/2 \\ 2(1-x); & 1/2 < x \leq 1. \end{cases}$

**Proof.** Define  $h : [0, 1] \rightarrow [0, 1]$  by  $h(x) = \sin^2(\pi x/2)$ . We can see that  $h$  is a homeomorphism because it is one-to-one, onto and both  $h$  and  $h^{-1}$  are continuous. ( $h$  is not a diffeomorphism since  $h'(1) = 0$ ). Also,

$$L_4 \circ h(x) = L_4\left(\sin^2\left(\frac{\pi x}{2}\right)\right) = 4\sin^2\left(\frac{\pi x}{2}\right)\left(1 - \sin^2\left(\frac{\pi x}{2}\right)\right) = \sin^2(\pi x),$$

and

$$h \circ T(x) = h(Tx) = \begin{cases} h(2x); & 0 \leq x \leq 1/2 \\ h(2-2x); & 1/2 < x \leq 1 \end{cases} = \sin^2(\pi x),$$

so  $L_4 \circ h = h \circ T$  and  $L_4$  and  $T$  are conjugate. □

2. *The logistic map*  $L_4$  is a factor of the angle doubling map  $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $f(z) = z^2$ .

**Proof.** Define  $h : \mathbb{S}^1 \rightarrow [0, 1]$  by  $h(e^{ix}) = \sin^2 x$ . Then

$$L_4 \circ h(e^{ix}) = L_4(\sin^2 x) = 4\sin^2(1 - \sin^2 x) = \sin^2(2x),$$

and

$$h \circ f(e^{ix}) = h(e^{2ix}) = \sin^2(2x).$$

$h$  is clearly well defined, onto and continuous, but it is not one-to-one:  $h(e^{ix}) = h(e^{-ix})$ , so  $L_4$  is a factor of  $f$ , but  $h$  is not a conjugacy. □

We can now show that many of the above maps are chaotic. In order to do this, we need to weaken the conditions of Proposition 7.2.1. If we drop the requirement that  $h$  is a homeomorphism, but just require it to be continuous and onto, then we can show that for continuous functions, if  $f$  is chaotic then so is  $g$ . In other words, if

$g$  is a factor of  $f$  where  $f$  is chaotic, then  $g$  is also chaotic. This result will be useful in showing that a number of well known maps are chaotic.

**Proposition 7.2.4** *Let  $h : X \rightarrow Y$  be continuous and onto. If  $f : X \rightarrow X$  and  $g : Y \rightarrow Y$  are both continuous, and satisfy  $h \circ f = g \circ h$  with  $f$  chaotic, then  $g$  is chaotic.*

Before proving this proposition we need a lemma concerning continuous functions on metric spaces:

**Lemma 7.2.5** *Let  $h : X \rightarrow Y$  be a continuous function where  $X$  and  $Y$  are metric spaces, and  $A \subseteq X$ , then  $h(\overline{A}) \subseteq \overline{h(A)}$ .*

**Proof.** Let  $y \in h(\overline{A})$ . Then there exists  $x \in \overline{A}$  with  $y = h(x)$ . We can find a sequence  $x_n \in A$  with  $\lim_{n \rightarrow \infty} x_n = x$ .

Then  $h(x_n) \in h(A)$ , and since  $h$  is continuous

$$\lim_{n \rightarrow \infty} h(x_n) = h(x) = y, \quad \text{so that } y \in \overline{h(A)}.$$

□

**Proof of Proposition 7.2.4** Use  $\text{Per}(f)$  and  $\text{Per}(g)$  to denote the periodic points of  $f$  and  $g$  respectively. We saw earlier that  $h(\text{Per}(f)) \subseteq \text{Per}(g)$ . Since  $f$  is chaotic,  $\overline{\text{Per}(f)} = X$ , and since  $h$  is onto,  $h(X) = Y$ . Then, using the lemma we have

$$Y = h(X) = h(\overline{\text{Per}(f)}) \subseteq \overline{h(\text{Per}(f))} \subseteq \overline{\text{Per}(g)},$$

so that  $\overline{\text{Per}(g)} = Y$ . In other words, the periodic points of  $g$  are dense in  $Y$ .

$f$  is transitive, so there exists  $x_0 \in X$  with  $\overline{O_f(x_0)} = X$  (where we use the subscript to distinguish the orbits with respect to  $f$  and  $g$ ). Now

$$\begin{aligned} h(O_f(x_0)) &= h\{f^n(x_0) : n \in \mathbb{Z}^+\} = \{h \circ f^n(x_0) : n \in \mathbb{Z}^+\} \\ &= \{g^n \circ h(x_0) : n \in \mathbb{Z}^+\} = O_g(h(x_0)). \end{aligned}$$

Thus,

$$Y = h(X) = h(\overline{O_f(x_0)}) \subseteq \overline{h(O_f(x_0))} = \overline{O_g(h(x_0))},$$

and  $h(x_0)$  is a transitive point for  $g$ . Since  $g$  is continuous, transitive, and has a dense set of periodic points, it follows from Theorem 6.6.1 that  $g$  is chaotic.

□

It is easily seen that Proposition 7.2.4 remains true if we replace the requirement that  $h$  be onto by requiring that  $h(X)$  be dense in  $Y$ .

**Theorem 7.2.6** *The tent map  $T : [0, 1] \rightarrow [0, 1]$ , the logistic map  $L_4(x) = 4x(1 - x)$ , the angle doubling map  $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $f(z) = z^2$ , and the shift map  $\sigma$ , are all chaotic.*

**Proof.** If we can show that the angle-doubling map  $f$  is a factor of the shift map  $\sigma$ , then we have:

- (i) The tent map  $T$  is conjugate to the logistic map  $L_4$ .
- (ii)  $L_4$  is a factor of  $f(z)$ ,
- (iii)  $f(z)$  is factor of  $\sigma$ .

It has been shown that the shift map is chaotic. The result now follows from Theorem 7.2.1, since all of these maps are continuous.

We need only show that if  $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $f(z) = z^2$  and  $\sigma : \Sigma \rightarrow \Sigma$  is the shift map, then  $f$  is a factor of  $\sigma$ . Define  $h : \Sigma \rightarrow \mathbb{S}^1$  by

$$h(a_1, a_2, a_3, \dots) = e^{2\pi i(\cdot a_1 a_2 a_3 \dots)}.$$

$$\begin{aligned} f \circ h(a_1, a_2, a_3, \dots) &= f(e^{2\pi i(\cdot a_1 a_2 a_3 \dots)}) = e^{4\pi i(\cdot a_1 a_2 a_3 \dots)} \\ &= e^{2\pi i 2(a_1/2 + a_2/2^2 + a_3/2^3 + \dots)} = e^{2\pi i a_1} e^{2\pi i(\cdot a_2 a_3 \dots)} = e^{2\pi i(\cdot a_2 a_3 \dots)}, \end{aligned}$$

and

$$h \circ \sigma(a_1, a_2, a_3, \dots) = h(a_2, a_3, \dots) = e^{2\pi i(\cdot a_2 a_3 \dots)},$$

or  $h \circ \sigma = f \circ h$ .

It is clear that  $h$  is onto  $\mathbb{S}^1$ . We leave the proof of the continuity as an exercise.  $h$  is not one-to-one since for example  $h(0, 1, 1, 1, \dots) = -1 = h(1, 0, 0, \dots)$ . □

## Exercises 7.2

1. (a) Prove that the map  $h : \mathbb{S}^1 \rightarrow [0, 1]$ ,  $h(e^{ix}) = \sin^2(x)$  is well defined, continuous and onto.
- (b) Do the same for the map  $h : \Sigma \rightarrow \mathbb{S}^1$ ,  $h(a_1, a_2, a_3, \dots) = e^{2\pi i(\cdot a_1 a_2 a_3)}$ , defined in Theorem 7.2.6.
  
2. If  $D : [0, 1] \rightarrow [0, 1]$  is the doubling map,  $D(x) = 2x \pmod{1}$ , and  $f : S^1 \rightarrow S^1$  is the angle doubling map,  $f(z) = z^2$ , show that  $D$  is a factor of  $f$ .

3. (a) If  $g(z) = z^3$  on  $S^1$ , show that  $g$  is the angle tripling map.
- (b) Find the periodic points of  $g$ , and show that they are dense in  $S^1$ .
- (c) Show that the map  $F : [0, 1] \rightarrow [0, 1]$ ,  $F(x) = 3x \bmod 1$  (of Exercises 6.5 # 4), is a factor of  $g$ .
4. If  $T_3 : C \rightarrow C$  is the tent map with  $\mu = 3$ , but restricted to the Cantor set, show that  $T_3$  is conjugate to the shift map  $\sigma : \Sigma \rightarrow \Sigma$ . Deduce that  $T_3$  is chaotic on  $C$ . (See Exercise 6.6 # 1 for a different proof).
5. Is the shift map  $\sigma : \Sigma \rightarrow \Sigma$  conjugate to the doubling map  $D$ ?
6. Let  $U : [-1, 1] \rightarrow [-1, 1]$  be defined by  $U(x) = 1 - 2x^2$ , and  $T : [0, 1] \rightarrow [0, 1]$  be the tent map. Prove that  $h : [0, 1] \rightarrow [-1, 1]$ ,  $h(x) = -\cos(\pi x)$ , defines a conjugacy between  $U$  and  $T$ .
7. Suppose that  $f$  and  $g$  are real functions conjugate via a diffeomorphism  $h$ . Show that  $h$  preserves the value of the derivative at the fixed point (i.e., if  $p$  is a fixed point of  $f$ ,  $f'(p) = g'(h(p))$ ). What is the situation with periodic points?
8. Prove that the map  $G : [-1, 1] \rightarrow [-1, 1]$ ,  $G(x) = 4x^3 - 3x$  is conjugate to  $F : [0, 1] \rightarrow [0, 1]$ ,  $F(x) = \begin{cases} 3x; & 0 \leq x < 1/3 \\ -3x + 2; & 1/3 \leq x < 2/3 \\ 3x - 2; & 2/3 \leq x \leq 1, \end{cases}$  via the conjugacy  $h : [0, 1] \rightarrow [-1, 1]$ ,  $h(x) = \cos(\pi x)$ . (Hint:  $\cos(3x) = 4\cos^3(x) - 3\cos(x)$ ). Deduce that  $G$  is chaotic. This fact is related to Exercise 7.1, # 7, concerning the Chebyshev polynomials.
9. Prove that Theorem 7.2.4 remains true if we replace the condition that  $h$  is onto, by  $\overline{h(X)} = Y$ .

10. Show that the notion of stable fixed point is preserved by conjugacy (see Definition 4.3.5(i)).

### 7.3 Linear Conjugacy.

It is sometimes the case that the conjugacy between two real (or complex) functions, is given by a map with a straight line graph (an affine map). This is called a linear conjugacy, and is stronger than the usual notion of conjugacy.

**Definition 7.3.1** For functions  $f : I \rightarrow I$  and  $g : J \rightarrow J$  defined on subintervals of  $\mathbb{R}$ , we say that  $f$  and  $g$  are *linearly conjugate*, and that  $h$  is a *linear conjugacy*, if  $h$  maps  $I$  onto  $J$  where  $h(x) = ax + b$  for some  $a, b \in \mathbb{R}$ ,  $a \neq 0$ , and  $h \circ f = g \circ h$ .

The following example gives a criterion for two quadratic functions to be linearly conjugate.

**Example 7.3.2** Let  $F(x) = ax^2 + bx + c$  and  $G(x) = rx^2 + sx + t$ , where  $a \neq 0$  and  $r \neq 0$ . If

$$c = \frac{b^2 - s^2 + 2s - 2b + 4rt}{4a},$$

then  $F$  and  $G$  are linearly conjugate via the affine map

$$h(x) = \frac{a}{r}x + \frac{b-s}{2r}.$$

**Proof.**

$$\begin{aligned} h \circ F(x) &= h(ax^2 + bx + c) = \frac{a(ax^2 + bx + c)}{r} + \frac{b-s}{2r} \\ &= \frac{a^2}{r}x^2 + \frac{ab}{r}x + \frac{2ac + b - s}{2r}, \end{aligned}$$

and

$$\begin{aligned} G \circ h(x) &= G\left(\frac{a}{r}x + \frac{b-s}{2r}\right) = r\left(\frac{a}{r}x + \frac{b-s}{2r}\right)^2 + s\left(\frac{a}{r}x + \frac{b-s}{2r}\right) + t \\ &= r\left(\frac{a^2}{r^2}x^2 + 2\frac{a(b-s)}{2r^2}x + \frac{(b-s)^2}{4r^2}\right) + \frac{sa}{r}x + \frac{bs - s^2}{2r} + t \\ &= \frac{a^2}{r}x^2 + \frac{ab}{r}x + \frac{(b-s)^2 + 2bs - 2s^2 + 4rt}{4r}. \end{aligned}$$

We see that these are equal if

$$c = \frac{b^2 - s^2 + 2s - 2b + 4rt}{4a}.$$

□

**Examples 7.3.3** 1. If  $F(x) = ax^2 + bx + c$  is a dynamical system on the interval  $[0, 1]$  (see Exercise 1.1 # 7, where conditions are given for maps of this type to be dynamical systems), then the conjugacy between  $F$  and  $G$  of Example 7.3.2 has the property

$$h(0) = \frac{b-s}{2r} \quad \text{and} \quad h(1) = \frac{2a+b-s}{2r}.$$

So if  $a/r > 0$ , then  $F$  is conjugate to  $G$  on the interval  $[\frac{b-s}{2r}, \frac{2a+b-s}{2r}]$ .

2. If  $L_\mu(x) = \mu x(1-x)$ ,  $f_c(x) = x^2 + c$ , and  $c = \frac{2\mu - \mu^2}{4}$ , then  $L_\mu$  restricted to  $[0, 1]$  is linearly conjugate to  $f_c$  restricted to  $[-\mu/2, \mu/2]$ . In particular,  $L_4(x) = 4x(1-x)$  on  $[0, 1]$ , is conjugate to  $f_{-2}(x) = x^2 - 2$  on  $[-2, 2]$ . It follows that  $f_{-2}$  restricted to  $[-2, 2]$  is chaotic. The conjugacy is given by  $h(x) = -\mu x + \mu/2$ .

**Proof.** We apply Example 7.3.2 with

$$a = -\mu, \quad b = \mu, \quad c = 0, \quad r = 1, \quad s = 0, \quad t = c.$$

In this case  $h(0) = \mu/2$ ,  $h(1) = -\mu/2$ , and we can check that the conditions of the example hold when  $c = \frac{2\mu - \mu^2}{4}$  (also see Exercises 7.3 # 3). □

3. On the other hand, if  $\mu = 2$ , we see that  $L_2(x) = 2x(1-x)$  on  $[0, 1]$  is conjugate to  $f_0(x) = x^2$  on  $[-1, 1]$ . Recall in Exercises 1.1 # 3, the difference equation  $x_{n+1} = 2x_n(1-x_n)$  transforms to  $y_{n+1} = y_n^2$  on setting  $x_n = (1-y_n)/2$ . This reflects the fact that  $L_2$  and  $f_0$  are conjugate via  $h(x) = -2x + 1$ .

4. We can check that the logistic map  $L_4$  is conjugate to  $F : [-1, 1] \rightarrow [-1, 1]$ ,  $F(x) = 2x^2 - 1$ .

### Exercises 7.3

1. Let  $(X, f)$  and  $(Y, g)$  be conjugate dynamical systems, with conjugacy  $h : X \rightarrow Y$ . If  $A \subset X$  is invariant under  $f$ , then we may think of  $(A, f)$  as a dynamical system. Show that  $(h(A), g)$  is a dynamical system.
2. Check that for  $0 < \mu \leq 4$ , if  $f_c(x) = x^2 + c$  with  $c = (2\mu - \mu^2)/4$ , then  $f_c$  is a dynamical system on  $[-\mu/2, \mu/2]$ .
3. In this question we are looking at  $f_c$  and  $L_\mu$  as dynamical systems on  $\mathbb{R}$ .

- (a) Show that if  $c > 1/4$ , there is no  $\mu \in \mathbb{R}$  for which  $f_c$  is linearly conjugate to  $L_\mu$ .
- (b) Show that if  $c = 1/4$ , then  $f_c$  is linearly conjugate to  $L_1$ .
- (c) Show that if  $c < 1/4$ , then  $f_c$  is linearly conjugate to both  $L_{\mu_1}$  and  $L_{\mu_2}$ , where  $\mu_1 = 1 + \sqrt{1 - 4c}$  and  $\mu_2 = 1 - \sqrt{1 - 4c}$ . Deduce that  $L_{\mu_1}$  is conjugate to  $L_{\mu_2}$  (for example,  $L_{-2}$  is conjugate to  $L_4$ ).
- (d) From (c), we see that  $L_{2/3}$  is conjugate to  $L_{4/3}$ . Why does this not contradict familiar properties of  $L_\mu$ ?
- (e) Deduce that if  $c > 0$ , then there is a unique  $\mu > 0$  such that  $f_c$  is linearly conjugate to  $L_\mu$ .
4. (a) Let  $f_a(x) = ax$ ,  $f_b(x) = bx$ ,  $a, b \in \mathbb{R}$ , be defined on  $\mathbb{R}$ . Under which conditions are  $f_a$  and  $f_b$  linearly conjugate?
- (b) Show that any conjugation  $h$ , between  $f_a$  and  $f_b$ , cannot be a diffeomorphism unless  $a = b$ . (Hint: Differentiate the conjugacy equation and deduce that  $h'(0) = 0$ ).
- (c) Let  $0 < a, b < 1$  and  $f_a, f_b : [0, 1] \rightarrow [0, 1]$ . Show that any conjugacy between  $f_a$  and  $f_b$ ,  $h : [0, 1] \rightarrow [0, 1]$  must satisfy  $h(0) = 0$ ,  $h(1) = 1$ , and  $h(a^n) = b^n$  for all  $n \in \mathbb{Z}^+$ .
5. Show that every quadratic polynomial  $p(x) = ax^2 + bx + d$  is linearly conjugate to a unique polynomial of the form  $f_c(x) = x^2 + c$ .
6. Prove that the logistic map  $L_4$  is conjugate to  $F : [-1, 1] \rightarrow [-1, 1]$ ,  $F(x) = 2x^2 - 1$ .

7. (a) Show that the logistic map  $L_\mu(x) = \mu x(1 - x)$ ,  $x \in [0, 1]$  is conjugate to the logistic type map  $F_\mu(x) = (2 - \mu)x(1 - x)$  ( $\mu \neq 2$ ), via the linear conjugacy (which is defined on the interval with end points  $\frac{1-\mu}{2-\mu}$  and  $\frac{1}{2-\mu}$ ):

$$h(x) = \frac{\mu}{2 - \mu}x + \frac{1 - \mu}{2 - \mu}.$$

- (b) Deduce that  $L_{-2} : [-1/2, 3/2] \rightarrow [-1/2, 3/2]$  is conjugate to  $L_4 : [0, 1] \rightarrow [0, 1]$ .
- (c) Show that  $L_{4/3} : [0, 1] \rightarrow [0, 1]$  is conjugate to  $L_{2/3} : [-1/2, 3/2] \rightarrow [-1/2, 3/2]$ .
- (d) Use part (b) to find a closed form solution to the difference equation  $x_{n+1} = -2x_n(1 - x_n)$ . (Hint: Use the conjugacy between  $L_{-2}$  and  $L_4$ , and also Exercise 1.1 # 3(ii)).



## CHAPTER 8

### Singer's Theorem.

The Schwarzian derivative was introduced in Chapter 1 to give a criterion for a non-hyperbolic fixed point of a map  $f$  to be attracting or repelling. The hypotheses of that result suggests that having a negative Schwarzian derivative is important for a dynamical system. In Section 8.1 we give some criteria for a function  $f$  to have a negative Schwarzian derivative. As a consequence, we see the important role the critical points of  $f$  (where  $f'(x) = 0$ ), play in the theory of iteration, something that will be emphasized in our study of complex iterations in Chapter 14. Chapter 8 uses definitions and results from Chapters 1 and 2, requiring an understanding of the notions of the basin of attraction of an attracting periodic point, and the Schwarzian derivative. It is essentially independent of Chapters 3 through 7, and the remaining chapters in the book.

The main result is a version of Singer's Theorem (1978, [115]), which says that for a map having negative Schwarzian derivative, the basin of attraction of an attracting periodic point must contain a critical point of the map. We use this result to show the surprising fact that the logistic map  $L_\mu : [0, 1] \rightarrow [0, 1]$ ,  $L_\mu(x) = \mu x(1 - x)$ , can have at most one attracting cycle for any value of  $\mu$ , with  $0 < \mu < 4$ . Singer's result is actually a real version of a general result of Fatou, which says that the basin of attraction of an attracting periodic point of a complex map must contain a critical point of the map. However, the real version requires a different proof.

Throughout Chapter 8, we require the map  $f : \mathbb{R} \rightarrow \mathbb{R}$  to be a  $C^3$ -function on  $\mathbb{R}$ . This means that  $f'''(x)$  exists, and is continuous for  $x \in \mathbb{R}$ .

#### 8.1 The Schwarzian Derivative Revisited.

Recall that the Schwarzian derivative of  $f(x)$  is:

$$Sf(x) = \frac{f'''(x)}{f'(x)} - \frac{3}{2} \left[ \frac{f''(x)}{f'(x)} \right]^2.$$

Set  $F(x) = \frac{f''(x)}{f'(x)}$ . Then a computation (see the exercises), shows that

$$Sf(x) = F'(x) - \frac{1}{2} [F(x)]^2.$$

Given a quadratic  $p(x) = ax^2 + bx + c$ ,  $a \neq 0$ , we see that

$$Sp(x) = -\frac{6a^2}{(2ax+b)^2} < 0,$$

except at the critical point of  $p(x)$ ,  $x = -b/2a$ . We regard  $Sp(-b/2a) = -\infty$  (since  $\lim x \rightarrow -b/2a Sp(x) = -\infty$ ), so that the Schwarzian derivative is negative everywhere.

Our first goal is to show that many polynomials have negative Schwarzian derivatives.

**Lemma 8.1.1** *Let  $f(x)$  be a polynomial of degree  $n$  for which all the roots of its derivative  $f'(x)$  are distinct and real. Then  $Sf(x) < 0$  for all  $x$ .*

**Proof.** Suppose that the derivative of  $f(x)$  is given by

$$f'(x) = a(x - r_1)(x - r_2) \cdots (x - r_{n-1}),$$

where  $a \in \mathbb{R}$  and  $r_1, r_2, \dots, r_{n-1}$  are all real and distinct. Then

$$F(x) = \frac{f''(x)}{f'(x)} = (\ln f'(x))' = \sum_{i=1}^{n-1} \frac{1}{x - r_i},$$

and so

$$F'(x) = -\sum_{i=1}^{n-1} \frac{1}{(x - r_i)^2}.$$

Now substitute  $F'(x)$  into the Schwarzian derivative formula to obtain:

$$Sf(x) = F'(x) - \frac{1}{2}[F(x)]^2 < 0.$$

□

The next result shows that the Schwarzian derivative has some pleasing properties.

**Lemma 8.1.2** *Assume that  $f$  is a  $C^3$  map on  $\mathbb{R}$ , then*

- (i)  $S(f \circ g)(x) = Sf(g(x)) \cdot (g'(x))^2 + Sg(x).$

(ii) If  $Sf < 0$  and  $Sg < 0$ , then  $S(f \circ g) < 0$ .

(iii) If  $Sf < 0$ , then  $Sf^k < 0$  for all  $k \in \mathbb{Z}^+$ .

**Proof.** (i) As above we have  $F(x) = \frac{f''(x)}{f'(x)}$ , so set  $G(x) = \frac{g''(x)}{g'(x)}$  and  $H(x) = \frac{h''(x)}{h'(x)}$ , where  $h = f \circ g$ . Then

$$h'(x) = f'(g(x)) \cdot g'(x), \quad h''(x) = (f''(g(x))(g'(x))^2 + (f'(g(x))g''(x)),$$

so that

$$\begin{aligned} H(x) &= \frac{f''(g(x))(g'(x))^2 + f'(g(x))g''(x)}{f'(g(x))g'(x)} \\ &= \frac{f''(g(x))g'(x)}{f'(g(x))} + \frac{g''(x)}{g'(x)} = F(g(x))g'(x) + G(x). \end{aligned}$$

This gives

$$H'(x) = F'(g(x))(g'(x))^2 + F(g(x))g''(x) + G'(x),$$

and

$$\begin{aligned} S(f \circ g)(x) &= H'(x) - \frac{1}{2}[H(x)]^2 \\ &= \left[ F'(g(x)) - \frac{1}{2}[F(g(x))]^2 \right] (g'(x))^2 + F(g(x))g''(x) - F(g(x))g'(x)G(x) + G'(x) - \frac{1}{2}[G(x)]^2 \\ &= Sf(g(x)) \cdot (g'(x))^2 + Sg(x), \end{aligned}$$

since  $G(x) = g''(x)/g'(x)$ .

(ii) is now immediate, and (iii) follows by induction. □

**Example 8.1.3** Let  $g(x) = \frac{ax+b}{cx+d}$ ,  $a, b, c, d \in \mathbb{R}$ , be a *linear fractional transformation*. A direct calculation shows that  $Sg(x) = 0$  everywhere in its domain. It follows from Lemma 1 that if  $h(x) = g(f(x)) = \frac{af(x)+b}{cf(x)+d}$ , then  $Sh(x) = Sf(x)$ .

We now prove in Theorem 8.1.4, a version of Singer's Theorem. Recall that if  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a continuous function, and  $c \in \mathbb{R}$  is an attracting fixed point or attracting periodic point, then the basin of attraction  $B_f(c)$  is an open set. Denote by  $W$  the immediate basin of attraction of  $c$  (the maximal open interval contained in  $B_f(c)$ , containing  $c$ ).

**Theorem 8.1.4** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a  $C^3$  map with a negative Schwarzian derivative. If  $c$  is an attracting periodic point of  $f$ , then either:

- (i) the immediate basin of attraction  $W$  of  $c$  is an unbounded interval (of one of the three forms:  $(-\infty, \infty)$ ,  $(-\infty, a)$  or  $(a, \infty)$ ) for some  $a \in \mathbb{R}$ , or
- (ii) there is a critical point of  $f$  (where  $f'(x) = 0$ ), whose orbit is attracted to the orbit of  $c$  under  $f$ .

**Proof.** We first look at the case where  $c$  is a fixed point of  $f$ . Suppose that its immediate basin of attraction is the open interval  $W$ , and that (i) does not hold.

Then  $W$  is a bounded set, so  $W = (a, b)$  for some  $a, b \in \mathbb{R}$ . Using arguments given in Chapter 2, the continuity of  $f$ , and the fact that  $a, b \notin (a, b)$  gives  $f(a), f(b) \notin (a, b)$ , so there are three possibilities for  $f(a)$  and  $f(b)$ .

**Case 1:**  $f(a) = f(b)$  - this will occur, for example, if  $a$  is a fixed point, and  $b$  is an eventual fixed point. It follows from the Mean Value Theorem, that  $(a, b)$  contains a critical point of  $f$ .

**Case 2:**  $f(a) = a$  and  $f(b) = b$ . By the Mean Value Theorem, there are points  $x_1 \in (a, c)$  and  $x_2 \in (c, b)$  satisfying

$$f'(x_1) = \frac{f(c) - f(a)}{c - a} = 1, \quad f'(x_2) = \frac{f(b) - f(c)}{b - c} = 1,$$

or  $f'(x_1) = f'(x_2) = 1$ . But since  $c$  is an attracting fixed point,  $|f'(c)| \leq 1$ . It follows that either  $f'(x_0) = 0$  for some  $x_0 \in (x_1, x_2)$ , or  $f'(x)$  has a minimum value  $f'(x_0) > 0$ . In the latter case, we have  $f'(x_0) > 0$ ,  $f''(x_0) = 0$ , and  $f'''(x_0) > 0$ , so that  $Sf(x_0) > 0$ , contradicting the Schwarzian derivative being everywhere negative.

**Case 3:**  $f(a) = b$  and  $f(b) = a$ . Then  $a, b$  and  $c$  are fixed by  $f^2$ , so that  $(f^2)'$  has a zero  $x_0$  in  $(a, b)$  (as in Case 1). But

$$(f^2)'(x_0) = (f'(f(x_0)))f'(x_0) = 0,$$

so either  $x_0$  or  $f(x_0)$  is a root of  $f'$ , but both lie in  $(a, b)$ .

Now suppose that  $c$  is a point of period  $k$ , then  $f^k(c) = c$ , an attracting fixed point for  $f^k$ . From our earlier arguments, the immediate basin of attraction of  $c$  (for  $f^k$ ), contains a critical point of  $f^k$ , say  $x_0$ :

$$(f^k)'(x_0) = f'(x_0)f'(f(x_0)) \cdots f'(f^{k-1}(x_0)) = 0,$$

so that  $f'(f^m(x_0)) = 0$  for some  $0 \leq m < k$ . In this case,  $f^m(x_0) \in f^m(W) \subset W$ , the basin of attraction of  $c$ .

□

**Example 8.1.5** 1. Consider the polynomial  $f(x) = x - x^5$ . We see that  $f'(x)$  has only two real roots:  $\pm(1/5)^{1/4}$ , so Lemma 8.1.1 is not applicable. We have  $f''(x) = -20x^3$  and  $f'''(x) = -60x^2$ . The fixed point  $x = 0$  is non-hyperbolic, and since  $f''(0) = f'''(0) = 0$ , none of our earlier criteria are applicable.

Substituting into the Schwarzian derivative gives:

$$Sf(x) = \frac{-60x^2}{(1-5x^4)} - \frac{3}{2} \left[ \frac{-20x^3}{1-5x^4} \right]^2 = \frac{-60x^2(1+5x^4)}{(1-5x^4)^2},$$

which is always negative. We can check that  $x = 0$  is an attracting fixed point. In fact,  $f$  has a repelling 2-cycle  $\{-2^{1/4}, 2^{1/4}\}$ , and the basin of attraction of  $x = 0$  is  $(-2^{1/4}, 2^{1/4})$ , containing both critical points (see the exercises).

2. The map  $f(x) = 3x/4 + x^3$  has a positive Schwarzian derivative on the interval  $(-1/2\sqrt{2}, 1/2\sqrt{2})$ . It has fixed points at  $x = 0$  and  $x = \pm 1/2$ , with  $x = 0$  attracting. The basin of attraction of  $x = 0$  is  $(-1/2, 1/2)$ , but  $f$  has no critical points.

## 8.2 Singer's Theorem.

An immediate consequence of the last result, to be applied to the logistic maps is:

**Corollary 8.2.1** *Let  $f : [0, 1] \rightarrow [0, 1]$  be a  $C^3$  map with  $Sf(x) < 0$  for all  $x$ . The basin of attraction of an attracting cycle contains 0, 1 or a critical point of  $f(x)$ .*

**Proof.** If  $J = (a, b)$ ,  $0 < a < b < 1$ , is the basin of attraction of an attracting cycle, then we have seen above that it must contain a critical point of  $f$ . If the basin of attraction is not of this form, then it must be of the form  $[0, a)$  or  $(b, 1]$ , so will contain 0 or 1.

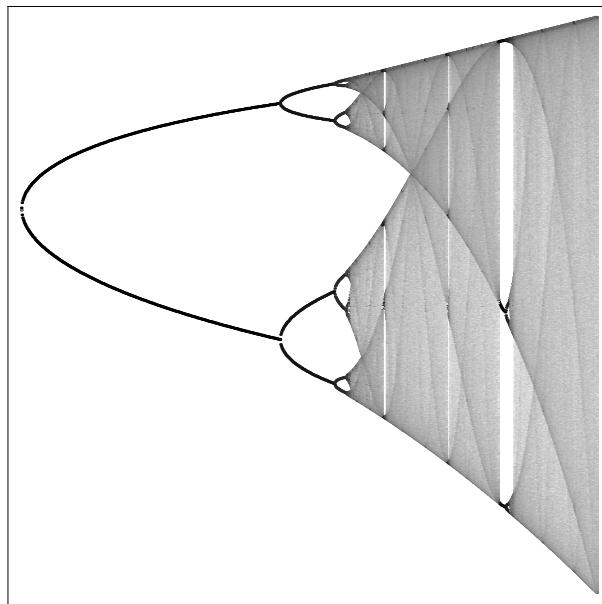
□

**Example 8.2.2** We now see that the logistic map  $L_\mu(x) = \mu x(1-x)$ ,  $0 < \mu < 4$ ,  $x \in [0, 1]$ , has at most one attracting periodic cycle. If  $0 < \mu \leq 1$ , then 0 is the only attracting fixed point, having basin of attractions  $[0, 1]$ .

For  $1 < \mu < 4$ ,  $L_\mu$  has exactly one critical point  $x_0 = 1/2$ .

Since  $L'_\mu(0) = \mu > 1$ , the fixed point 0 is unstable; therefore  $[0, a)$  cannot be the basin of attraction of 0. Furthermore,  $L_\mu(1) = 0$  and hence,  $(b, 1]$  is not a basin of attraction either. Since  $SL_\mu(x) < 0$  everywhere (at  $x = 1/2$ ,  $\lim_{x \rightarrow 1/2} L_\mu(x) = -\infty$ ),

we conclude that there is at most one attracting periodic cycle in  $(0, 1)$ , and the result follows.



The Bifurcation Diagram for  $\mu > 3$

Where we see exactly six horizontal lines in the bifurcation diagram of the logistic map, we must have an attracting 6-cycle, and not two attracting 3-cycles, since only one attracting cycle can exist at any one time.

**Remarks 8.2.3** 1. We have seen that the tent map  $T_2$  and the logistic map  $L_4$  are conjugate. We might ask the question: Are there parameter values  $1 < \lambda < 2$  and  $3 < \mu < 4$  such that  $T_\lambda$  and  $L_\mu$  are conjugate? We have seen that  $L_\mu$  has at most one attracting cycle. It has been shown that there are values of  $\mu$  for which  $L_\mu$  has no attracting cycle, and in this case  $L_\mu$  is conjugate to a Tent map (see [24] where criteria for unimodal maps having no attracting periodic orbits to be conjugate, are given). Clearly when  $L_\mu$  has an attracting cycle, it cannot be conjugate to such a tent map since all the periodic points of these tent maps are repelling.

2 In a similar way we arrive at Singer's Theorem [115], proved by David Singer in 1978. Singer's theorem is actually a real version of a theorem about complex functions proved by the French mathematician Gaston Julia in 1918 [72]:

**8.2.4 Singer's Theorem** Let  $f$  be a  $C^3$  map on a closed interval  $I$  with  $Sf(x) < 0$ , for all  $x \in I$ . If  $f$  has  $n$  critical points in  $I$ , then  $f$  has at most  $n+2$  attracting cycles.

**Example 8.2.5** If  $f_\lambda : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f_\lambda(x) = x^3 - \lambda x$ , then  $Sf_\lambda(x) < 0$  when  $\lambda > 0$ . The fixed points are  $x = 0$  and  $x = \pm\sqrt{\lambda+1}$ ,  $x = 0$  being attracting when  $|\lambda| < 1$ , and the other two being repelling.  $f_\lambda$  has two critical points,  $\pm\sqrt{\lambda/3}$ . The basin of attraction of  $x = 0$  is  $(-\sqrt{\lambda+1}, \sqrt{\lambda+1})$ , which includes the critical points.

## Exercises 8.2

1. Show that  $F(x) = x^4 - 2x^2 - 3$  has a negative Schwarzian derivative everywhere. (Hint: Look at the roots of  $F'(x)$ ).
2. If  $Sf(x)$  is the Schwarzian derivative of  $f(x)$ , a  $C^3$  function, and  $F(x) = \frac{f''(x)}{f'(x)}$ , show that  $Sf(x) = F'(x) - (F(x))^2/2$ .
3. Show that if  $c$  is an attracting fixed point of a  $C^1$ -function,  $f$ , then  $|f'(c)| \leq 1$ .
4. If  $f(x) = x - x^5$ , show that  $x = 0$  is an attracting fixed point, and that  $\{-2^{1/4}, 2^{1/4}\}$  is a repelling 2-cycle. Show that the basin of attraction of  $x = 0$  is the interval  $(-2^{1/4}, 2^{1/4})$ , and that it contains the critical points of  $f$ . (Hint: If  $(a, b)$  is the basin of attraction of  $x = 0$ , argue that the only possibility is  $f(a) = b$  and  $f(b) = a$ ).
5. (a) Show that if  $p$  is a polynomial of degree  $n$  having  $n$  distinct fixed points, and negative Schwarzian derivative, then not all of the fixed points can be attracting.  
(b) On the other hand, show that the logistic maps  $L_\mu : \mathbb{R} \rightarrow \mathbb{R}$  (for  $\mu > 2 + \sqrt{5}$ ), have negative Schwarzian derivative, but have no attracting periodic orbits.

6. (a) Prove that if  $a^2 - 3b > 0$ , then the cubic  $p(x) = x^3 + ax^2 + bx + c$  has a negative Schwarzian derivative.
- (b) Use (a) to show that if  $a^2 - 3b < 0$ , there is an interval on which  $p$  has a positive Schwarzian derivative.
- (c) Deduce that if  $p(x) = x^3 + ax^2 + bx + c$ , then  $Sp(x) > 0$  on an interval, if and only if  $p(x)$  is strictly increasing.
- (d) Use the above to show that  $p(x) = x^3 - x^2 + x/2$  has positive Schwarzian derivative on the interval  $(1/6, 1/2)$ . Show that  $x = 0$  is an attracting fixed point, but  $p(x)$  has no critical points.
7. How many attracting cycles can  $f(x) = ax^2 + bx + c$  have?
8. Here we give an example of a map which is strictly increasing (so has no critical points), having negative Schwarzian derivative ([32]).  
 Let  $G(x) = \lambda \arctan(x)$ , ( $\lambda \neq 0$ ). Show that the Schwarzian derivative is
- $$SG(x) = \frac{-2}{(1+x^2)^2},$$
- and that  $G(x)$  has no critical points.
- (a) Show that if  $|\lambda| < 1$ , then  $x = 0$  is an asymptotically stable fixed point with basin of attraction  $(-\infty, \infty)$ .
- (b) If  $\lambda > 1$ , then show that  $G$  has two attracting fixed points  $x_1$  and  $x_2$ , with basins of attraction  $(-\infty, 0)$  and  $(0, \infty)$ , respectively.
- (c) Show that if  $\lambda < -1$ , then  $G$  has an attracting 2-cycle  $\{\bar{x}_1, \bar{x}_2\}$  with basin of attraction  $(-\infty, 0) \cup (0, \infty)$ .
- (d) Does Theorem 8.1.4 apply to this example?
9. (a) Let  $f(x) = 2x^3 + bx^2 + cx + d$ . Find  $b, c$  and  $d$  so that the following hold: (i)  $x = 0$  is a period-2 point, (ii)  $f^2(1) = f(1)$ , (iii) both  $x = 0$  and  $x = 1$  are critical points of  $f$ . Show that there is a unique map with these properties.

- (b) Use Theorem 8.1.1 to show that  $f$  has a negative Schwarzian derivative.  
(c) What does Theorem 8.1.4 say about  $f$ ?

10. If  $f(x) = \frac{ax + b}{cx + d}$ , we have seen that the Schwarzian derivative satisfies  $Sf(x) = 0$  (Exercises 1.5). Now, suppose that  $f$  is a function for which  $Sf(x) = 0$ . Show that

- (i)  $(f''(x))^2/(f'(x))^3 = \text{constant}$ ,  
(ii)  $f(x)$  is of the form  $f(x) = \frac{ax + b}{cx + d}$ . (Hint: Set  $y = f'$  to obtain a separable differential equation of the form  $y' = cy^{3/2}$ ).

11. ([98]) Let  $N_f(x) = x - f(x)/f'(x)$  be the Newton function of a four times continuously differentiable function  $f$ . If  $f(\alpha) = 0$ , show that  $N_f'''(\alpha) = 2Sf(\alpha)$ , where  $Sf$  is the Schwarzian derivative of  $f$ .

12. Show that if  $p(x)$  is a polynomial of the form

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_4 x^4 + a_3 x^3 + a_1 x + a_0,$$

where  $a_3/a_1 > 0$ , then  $Sp(x) > 0$  on an interval containing 0.



## CHAPTER 9

### **Conjugacy, Fundamental Domains and the Tent Family.**

In this chapter we look at conjugacies of a more technical nature. First, we examine conjugacy between homeomorphisms defined on a subinterval of  $\mathbb{R}$ . Then we shall consider various questions involving conjugacy for the tent family  $T_\mu$ . Very often it is not possible to give a specific formula for the conjugacy map between two dynamical systems, but we are able to show that one exists. In this way we are able to answer various questions concerning conjugacies for distinct logistic maps.

#### **9.1 Conjugacy and Fundamental Domains.**

We have seen that two dynamical systems  $f$  and  $g$  with different dynamical properties cannot be conjugate. On the other hand, sometimes we have dynamical systems having very similar dynamical properties, which we would like to show are conjugate. In some cases, this can be done using the notion of *fundamental domain*, a set on which we construct a homeomorphism  $h$  in an arbitrary manner and show that it extends to a conjugacy on the whole space. We illustrate this idea with homeomorphisms  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ . We look at a fairly straightforward case where both homeomorphisms are order preserving, and have no fixed points (their graphs lie strictly above the line  $y = x$ ).

**Proposition 9.1.1** *Let  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  be homeomorphisms satisfying  $f(x) > x$  and  $g(x) > x$  for all  $x \in \mathbb{R}$ . Then  $f$  and  $g$  are conjugate.*

**Proof.** The idea for the proof is as follows: Select  $x_0 \in \mathbb{R}$  arbitrarily, and consider the *2-sided orbit*

$$O_f(x_0) = \{f^n(x_0) : n \in \mathbb{Z}\} = \{\dots, x_{-1}, x_0, x_1, x_2, \dots\}.$$

Since  $f(x) > x$  for all  $x$ ,  $(f^n(x_0))$  is an increasing sequence:  $\dots x_{-1} < x_0 < x_1 < x_2 < \dots$ , so that the sets

$$\dots, [x_{-1}, x_0), [x_0, x_1), [x_1, x_2), \dots,$$

are disjoint, and their union is all of  $\mathbb{R}$ . We must have  $\lim_{n \rightarrow \infty} x_n = \infty$ , since otherwise the limit would exist and it would have to be a fixed point. There are no fixed points since  $f(x) > x$ , always.

The set  $I = [x_0, f(x_0)) = [x_0, x_1)$ , is called a *fundamental domain* for  $f$ . Set  $J = [x_0, g(x_0))$ , and define a map  $h : I \rightarrow J$  to be continuous, increasing and onto, but otherwise arbitrary (e.g., we can set  $h(x_0) = x_0$  and  $h(f(x_0)) = g(x_0)$ , and then linearly from  $I$  to  $J$ ).

Every other orbit of  $f$  intertwines with  $O_f(x_0)$ : if  $y_0 \in (x_0, x_1)$ , then  $y_n = f^n(y_0) \in f^n(I)$ , so lies between  $x_n$  and  $x_{n+1}$ . It follows that every orbit has a unique member in the interval  $[x_0, x_1)$ , and we use this to extend the definition of  $h$  to all of  $\mathbb{R}$ .

If  $x \in f^n(I)$ , we define  $h(x)$  by mapping  $x$  back to  $I$  via  $f^{-n}$ , then using  $h(f^{-n}(x))$ , which is well defined, and then mapping back to  $g^n(J)$  using  $g^n$ . In other words, if  $x \in f^n(I)$ ,  $n \in \mathbb{Z}$ , set

$$h(x) = g^n \circ h \circ f^{-n}(x).$$

In this way,  $h$  is defined on all of  $\mathbb{R}$ . We can check that  $h$  is one-to-one. It is onto because  $h(f^n(I)) = g^n(J)$  for each  $n$ , and we can check that it is continuous. Finally, because of the definition of  $h$ , if  $x \in \mathbb{R}$ , then  $x \in f^n(I)$  for some  $n \in \mathbb{Z}$ , so  $x = f^n(y)$  for some  $y \in I$ . Then

$$g \circ h(x) = g(g^n \circ h \circ f^{-n}(x)) = g^{n+1} \circ h \circ f^{-(n+1)}(f(x)) = h \circ f(x),$$

so that  $f$  and  $g$  are conjugate. □

**Examples 9.1.2** 1. The above argument, which is due to Sternberg (see [122]), can be generalized to the situation where  $f(x)$  and  $g(x)$  are homeomorphisms with corresponding fixed points. To illustrate this, consider a homeomorphism  $f : [0, 1] \rightarrow [0, 1]$  which is order preserving, so that  $f(0) = 0$ ,  $f(1) = 1$  and  $f$  is increasing. Suppose that  $f$  has fixed points (in addition to 0 and 1), at  $c_1, c_2, \dots, c_n$ . Then  $f^2$  has the same collection of fixed points ( $f$  cannot have points of period 2 or higher). If  $f(x) > x$  for  $c_k < x < c_{k+1}$ , then we can use the argument of Proposition 9.1.1 to construct a homeomorphism between  $f$  and  $f^2$ , and do the same for each interval  $[c_i, c_{i+1}]$  (treating the case where  $f(x) < x$  in an analogous way). We thus see that  $f$  and  $f^2$  are conjugate maps (see also [61]).

2. Consider the logistic maps  $L_\mu(x) = \mu x(1 - x)$  for various values of  $\mu \in (0, 4]$  and  $x \in [0, 1]$ . We first show that for  $0 < \mu < \lambda \leq 1$ ,  $L_\mu$  and  $L_\lambda$  are conjugate. There is a slight complication here as these maps are not increasing, but they do have an

attracting fixed point at 0. We saw earlier that the basin of attraction is all of  $[0, 1]$ . We deal first with the interval on which the maps are increasing,  $[0, 1/2]$ , and look at the restriction of the functions to this interval.

Our aim is to construct a homeomorphism  $h : [0, 1] \rightarrow [0, 1]$  with the property  $L_\lambda \circ h = h \circ L_\mu$ . Take  $(L_\mu(1/2), 1/2] = (\mu/4, 1/2]$  as a fundamental domain for  $L_\mu$ , and  $(L_\lambda(1/2), 1/2] = (\lambda/4, 1/2]$  as a fundamental domain for  $L_\lambda$ . Define  $h : (\mu/4, 1/2] \rightarrow (\lambda/4, 1/2]$  by  $h(1/2) = 1/2$  and  $h(\mu/4) = \lambda/4$ , and then linearly on the remainder of the interval.

Set  $I = (\mu/4, 1/2]$  and  $J = (\lambda/4, 1/2]$ . Then since 0 is an attracting fixed point, the intervals  $L_\mu^n(I)$  and  $L_\lambda^n(J)$  are disjoint for  $n \in \mathbb{Z}^+$ , and their union is all of  $(0, 1/2]$ . Extend the definition of  $h$  so that it is defined on  $(0, 1/2]$  by;

$$h(x) = L_\lambda^n \circ h \circ L_\mu^{-n}(x), \quad \text{for } x \in L_\mu^n(I).$$

We can now check that  $h$  is continuous and increasing on  $[0, 1/2]$ , when we set  $h(0) = 0$ .

Now define  $h$  on  $(1/2, 1]$  by setting  $h(1 - x) = 1 - h(x)$  for  $x \in [0, 1/2)$  (giving a homeomorphism on  $[0, 1]$ ). Then

$$L_\lambda(h(1 - x)) = L_\lambda(1 - h(x)) = L_\lambda(h(x)) = h(L_\mu(x)) = h(L_\mu(1 - x)),$$

so that  $h$  is the required conjugation. □

3. A similar proof shows that  $L_\mu$  and  $L_\lambda$  are conjugate whenever  $1 < \mu < \lambda < 2$ . Look at the intervals  $[0, 1 - 1/\mu]$  and  $[1 - 1/\mu, 1/2]$  separately. Use the fact that  $1 - 1/\mu$  is an attracting fixed point, and then use the symmetry about the point  $x = 1/2$ .

However, these maps cannot be conjugate to  $L_2$  since any conjugating map  $h : [0, 1] \rightarrow [0, 1]$  must have the property that  $h(1/2) = 1/2$  (see the exercises). This leads to a contradiction.

4. The maps  $L_4$  and  $L_\mu$ ,  $\mu \in (0, 4)$  cannot be conjugate since  $L_4 : [0, 1] \rightarrow [0, 1]$  is an onto map, but  $L_\mu$  is not (see the exercises).

## Exercises 9.1

1. (a) Let  $a, b \in (0, 1)$  and  $f_a(x) = ax, f_b(x) = bx$  be dynamical systems on  $[0, 1]$ . We saw in Exercises 7.3 that these maps need not be linearly conjugate. Use the method of examples in this section, to prove that  $f_a$  and  $f_b$  are conjugate.

- (b) Let  $g : [0, 1] \rightarrow [0, 1]$  be continuous, strictly increasing with  $g(0) = 0$  and  $g(x) < x$  for all  $x \in (0, 1]$ . Prove that  $g$  is conjugate to  $f_a$  for any  $a \in (0, 1)$ .
2. Let  $0 < \lambda, \mu < 1$ . If  $h : [0, 1] \rightarrow [0, 1]$  is an orientation preserving homeomorphism with  $h \circ L_\mu(x) = L_\lambda \circ h(x)$  for all  $x \in [0, 1]$ , show that
- (a)  $h(1/2) = 1/2$ .
  - (b)  $h(\mu/4) = \lambda/4$ .
- (Hint:  $h$  is a conjugation between two different logistic maps with  $h \circ L_\mu(x) = L_\lambda \circ h(x)$ . Note that this equation also holds if we replace  $x$  by  $1 - x$ . Use this to deduce that  $h(x) + h(1 - x) = 1$  for all  $x \in [0, 1]$ ).
3. Let  $f, g : [0, 1] \rightarrow [0, 1]$  be homeomorphisms. Can  $f$  and  $g$  be conjugate in the following situations?
- (a) When  $f$  is order preserving and  $g$  is order reversing.
  - (b) When  $f(x) > x$  and  $g(x) < x$  for all  $x \in (0, 1)$ .
4. Let  $S_\mu(x) = \mu \sin(x)$ . Prove that if  $0 < \mu < \lambda < 1$ , then  $S_\mu$  and  $S_\lambda$  are conjugate maps.
5. Let  $R_a : \mathbb{S}^1 \rightarrow \mathbb{S}^1$  be the rotation  $R_a(z) = az$ .
- (a) Can  $R_a$  be conjugate to  $R_b$  for  $a \neq b$ ?
  - (b) If  $T_\alpha : [0, 1] \rightarrow [0, 1]$ ,  $T_\alpha(x) = x + \alpha \pmod{1}$ , and  $a = e^{2\pi i \alpha}$ , show that  $\phi \circ T_\alpha(x) = R_a \circ \phi(x)$  for  $x \in [0, 1)$ , but  $T_\alpha$  and  $R_a$  are not conjugate (the underlying spaces are not homeomorphic). ( $\phi : [0, 1] \rightarrow \mathbb{S}^1$ , is defined by  $\phi(x) = e^{2\pi i x}$ ).
6. Prove that  $T_\alpha : [0, 1] \rightarrow [0, 1]$ ,  $T_\alpha(x) = x + \alpha \pmod{1}$  is conjugate to its inverse map  $T_\alpha^{-1}(x) = x - \alpha \pmod{1}$ . Can  $T_\alpha$  be conjugate to  $T_\alpha^2$ ?

7\*. Prove that if  $0 < \mu \leq 2$ , the logistic map  $L_\mu$  is conjugate to  $L_\mu^2$ . Show that this is not true for  $\mu > 2$ . What is the corresponding result for the tent family?

8\*. The aim of this exercise is to show the uniqueness of the conjugacy between the tent map  $T_2$  and the logistic map  $L_4$ .

(i) Check that a conjugacy between  $T_2$  and  $L_4$  is given by  $k : [0, 1] \rightarrow [0, 1]$ ,

$$k(x) = \frac{2}{\pi} \arcsin(\sqrt{x}); \quad T_2 \circ k = k \circ L_4.$$

(ii) Suppose that  $h : [0, 1] \rightarrow [0, 1]$  is another conjugacy between  $T_2$  and  $L_4$ . Then  $h(0) = 0$ ,  $h(1) = 1$ , and  $h$  is a strictly increasing continuous function (why?), i.e.,  $h$  is an order preserving homeomorphism of  $[0, 1]$ .

(iii) Show that  $h$  maps the local maxima (respectively minima), of  $T_2^n$  to local maxima (respectively minima), of  $L_4^n$  (see Exercises 7.1# 4).

(iv) Use the fact that any such conjugation is order preserving, to show that  $h(x) = k(x)$  at all local maxima and local minima.

(v) Use the continuity of  $h$  and  $k$  to deduce that  $h(x) = k(x)$  for all  $x \in [0, 1]$ .

(vi) Deduce that there is no  $C^1$ -conjugacy between  $T_2$  and  $L_4$ .

9. Use the previous exercise to show that if  $L_4(x) = 4x(1 - x)$ , then  $L_4^n$  has critical points at  $\sin^2(k\pi/2^{n+1})$ , for  $k = 1, 2, \dots, 2^{n+1} - 1$ .

10. Use the fact that the conjugation between  $T_2$  and  $L_4$  is unique, to show that, if  $\phi : [0, 1] \rightarrow [0, 1]$  is a homeomorphism satisfying  $L_4 \circ \phi = \phi \circ L_4$ , then  $\phi(x) = x$  for all  $x \in [0, 1]$ , i.e.,  $\phi$  is the identity map. (Hint: first show that  $k \circ \phi$  is also a conjugation between  $T_2$  and  $L_4$ , where  $k(x) = \frac{2}{\pi} \arcsin(\sqrt{x})$ ).

## 9.2 Conjugacy, the Tent Map and Periodic Points of the Tent Family.

We saw in Section 2.7 that for  $\mu \geq (1 + \sqrt{5})/2$ , the tent map  $T_\mu$  has a point of period three, so by Sharkovsky's Theorem, it will have points of all possible periods. In this section we use a certain conjugacy to show that for  $\mu > 1$ ,  $T_\mu$  will have points of period  $2^n$  for each  $n \geq 1$ . Our argument is based on the result in [67]. We first show that the interval  $[1/(1+\mu), \mu/(1+\mu)]$  is invariant under  $T_\mu^2$  when  $1 < \mu \leq \sqrt{2}$ .

The formula for  $T_\mu^2$  in Section 2.7 gives

$$T_\mu^2(x) = \begin{cases} \mu^2 x; & 0 \leq x \leq \frac{1}{2\mu}, \\ \mu - \mu^2 x; & \frac{1}{2\mu} < x \leq \frac{1}{2}, \\ \mu^2 x + \mu - \mu^2; & \frac{1}{2} < x \leq 1 - \frac{1}{2\mu}, \\ \mu^2 - \mu^2 x; & 1 - \frac{1}{2\mu} < x \leq 1. \end{cases}$$

**Proposition 9.2.1** *For  $1 < \mu \leq \sqrt{2}$ , the restriction of  $T_\mu^2$  to the set  $[\frac{1}{1+\mu}, \frac{\mu}{1+\mu}]$  is well defined, so that the map*

$$T_\mu^2 : \left[ \frac{1}{1+\mu}, \frac{\mu}{1+\mu} \right] \rightarrow \left[ \frac{1}{1+\mu}, \frac{\mu}{1+\mu} \right],$$

*is a dynamical system.*

**Proof.** Note that for  $1 < \mu$ ,  $1/(1+\mu) < 1/2$  and  $\mu/(1+\mu) > 1/2$ . In addition,

$$T_\mu(\mu/(1+\mu)) = \mu/(1+\mu),$$

so  $\mu/(1+\mu)$  is a fixed point of  $T_\mu$ . We see that  $1/(1+\mu)$  is an eventual fixed point,  $1/(1+\mu)$  and  $\mu/(1+\mu)$ , being equally spaced on opposite sides of  $x = 1/2$ .

Let  $x \in [1/(1+\mu), \mu/(1+\mu)]$ . Then from the formula for  $T_\mu^2$ , and the fact that

$$\frac{1}{2\mu} < \frac{1}{1+\mu} < \frac{\mu}{1+\mu} < 1 - \frac{1}{2\mu},$$

we see that on this interval the minimum value of  $T_\mu^2$  occurs at  $x = 1/2$ . This gives

$$T_\mu^2(x) \geq T_\mu^2(1/2) = \mu(1 - \mu/2) \geq \frac{1}{1+\mu},$$

since this is equivalent to

$$\mu^3 - \mu^2 - 2\mu + 2 \leq 0, \quad \text{or} \quad (\mu - 1)(\mu^2 - 2) \leq 0,$$

where  $1 < \mu \leq \sqrt{2}$ .

On the other hand, assuming that  $x \in [1/(1+\mu), \mu/(1+\mu)]$ , we have that if  $x \leq 1/2$ , then  $T_\mu(x) = \mu x > \mu/(1+\mu) > 1/2$ , and  $T_\mu^2(x) = \mu(1-\mu x) < \mu(1-\mu/(1+\mu)) = \mu/(1+\mu)$ , so  $T_\mu^2(x) \in [1/(1+\mu), \mu/(1+\mu)]$ .

If instead we have  $x > 1/2$ , then

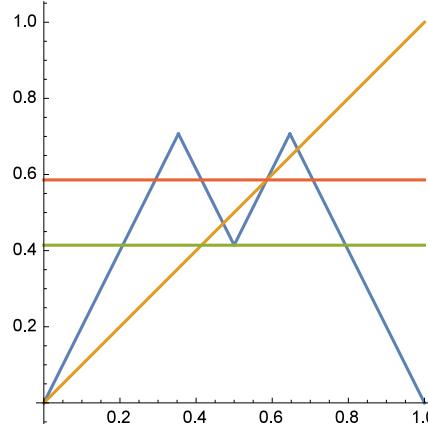
$$T_\mu(x) = \mu(1-x) > \mu(1 - \frac{\mu}{1+\mu}) = \frac{\mu}{1+\mu} > \frac{1}{2},$$

so

$$T_\mu^2(x) = \mu(1 - \mu(1-x)) = \mu(1 - \mu + \mu x) \leq \mu(1 - \mu + \frac{\mu^2}{1+\mu}) = \frac{\mu}{1+\mu},$$

so again  $T_\mu^2(x) \in [1/(1+\mu), \mu/(1+\mu)]$ . □

We use the Proposition 9.2.1, to show that  $T_\mu$  and  $T_{\sqrt{\mu}}^2$  are conjugate when  $T_{\sqrt{\mu}}^2$  is restricted to a suitable invariant subinterval.



When  $\mu = \sqrt{2}$ , we see an inverted version of the graph of  $T_2$  in the graph of  $T_{\sqrt{2}}^2$ .

**Proposition 9.2.2** *For  $1 < \mu \leq \sqrt{2}$ ,  $T_\mu^2$  restricted to the interval*

$$\left[ \frac{1}{1+\mu}, \frac{\mu}{1+\mu} \right],$$

*is conjugate to  $T_{\mu^2}$  on  $[0, 1]$ .*

**Proof.** From Proposition 9.2.1, we see that the given interval is invariant under  $T_\mu^2$ . We now show that we actually have a linear conjugacy  $h$ :

$$h \circ T_\mu^2 = T_{\mu^2} \circ h.$$

$h$  is a linear map of the form  $h(x) = ax + b$

$$h : \left[ \frac{1}{1+\mu}, \frac{\mu}{1+\mu} \right] \rightarrow [0, 1],$$

where

$$a = \frac{1+\mu}{1-\mu}, \quad b = \frac{\mu}{\mu-1}.$$

Then it can be seen that

$$h\left(\frac{\mu}{1+\mu}\right) = 0, \quad \text{and} \quad h\left(\frac{1}{1+\mu}\right) = 1.$$

If  $0 \leq x \leq 1/2$ , then since  $h^{-1}(x) = x/a - b/a$ , we can check that  $1/2 \leq h^{-1}(x) \leq \mu/(1+\mu) < 1 - 1/2\mu$ , so that

$$\begin{aligned} h \circ T_\mu^2 \circ h^{-1}(x) &= h \circ T_\mu^2 \left( \frac{x}{a} - \frac{b}{a} \right) \\ &= h \left( \mu^2 \left( \frac{x}{a} - \frac{b}{a} \right) + \mu - \mu^2 \right) = \mu^2 x - \mu^2 b + a(\mu - \mu^2) + b = \mu^2 x = T_{\mu^2}(x). \end{aligned}$$

Similarly, we can check that if  $1/2 < x \leq 1$ , then  $1/2\mu < 1/(1+\mu) \leq h^{-1}(x) \leq 1/2$ , and

$$\begin{aligned} h \circ T_\mu^2 \circ h^{-1}(x) &= h \circ T_\mu^2 \left( \frac{x}{a} - \frac{b}{a} \right) = h \left( \mu - \mu^2 \left( \frac{x}{a} - \frac{b}{a} \right) \right) \\ &= a\mu - \mu^2 x + \mu^2 b + b = \mu^2(1-x) = T_{\mu^2}(x), \end{aligned}$$

i.e., in both cases we have  $h \circ T_\mu^2 \circ h^{-1}(x) = T_{\mu^2}(x)$ , giving the desired conjugacy.  $\square$

**Corollary 9.2.3** For  $1 < \mu \leq 2$ ,  $T_{\sqrt{\mu}}$  restricted to the interval

$$\left[ \frac{\sqrt{\mu}-1}{\mu-1}, \frac{\mu-\sqrt{\mu}}{\mu-1} \right],$$

is conjugate to  $T_\mu$  on  $[0, 1]$ .

**Proof.** Replace  $\mu$  by  $\sqrt{\mu}$  in the previous result.  $\square$

We apply these results to obtain information about the periodic points of  $T_\mu$ . Unlike the situation for the logistic family, all the  $2^n$ -cycles are created at the same time.

**Theorem 9.2.4** For  $1 < \mu \leq 2$ ,  $T_\mu$  has a  $2^n$ -cycle for each  $n \in \mathbb{Z}^+$ .

**Proof.** We have seen that for each  $\mu > 1$ ,  $T_\mu$  has a period-2 point distinct from the fixed point of  $T_\mu$ . In particular, as  $\mu^2 > 1$ ,  $T_{\mu^2}$  has a period-2 point distinct from the fixed point of  $T_{\mu^2}$ . But by Corollary 9.2.3,  $T_{\mu^2}$  and  $T_\mu^2$  (suitably restricted), are conjugate, so  $T_\mu^2$  has a period-2 point distinct from the fixed point of  $T_\mu^2$ . This point must be a period-4 point for  $T_\mu$ , for if not, it would be a period-2 point, giving a fixed point for  $T_\mu^2$ .

Continuing this argument, starting with a period-2 point for  $T_{\mu^4}$ , and the conjugacy between  $T_{\mu^2}^2$  and  $T_{\mu^4}$ , we deduce that  $T_\mu$  has a period-8 point. In this way, for each  $n \in \mathbb{Z}^+$ ,  $T_\mu$  has a period- $2^n$  point.

□

**Example 9.2.5** Consider the case where  $\mu = 2$ . Then we see that  $T_2$ , the standard tent map, is conjugate to  $T_{\sqrt{2}}^2$  restricted to the interval  $[\sqrt{2}-1, 2-\sqrt{2}]$ . This implies that  $T_{\sqrt{2}}^2$  has the same dynamics as  $T_2$  on this subinterval. For example, it must have a 3-cycle, say  $\{c_1, c_2, c_3\}$ , where the  $c_i$ 's are distinct and  $T_{\sqrt{2}}^6(c_1) = c_1$ . It follows that  $c_1$  is a point of period 6 for  $T_{\sqrt{2}}$ , and in this way we deduce that  $T_{\sqrt{2}}$  has  $2k$ -cycles for each  $k \in \mathbb{Z}^+$ . We saw in Section 2.8 that for the tent family, a period-3 point is born when  $\mu = (1 + \sqrt{5})/2$ . In particular,  $T_{\sqrt{2}}$  has no 3-cycle, but if  $\alpha = (1 + \sqrt{5})/2$ , and since  $T_{\sqrt{\alpha}}^2$  (suitably restricted) is conjugate to  $T_\alpha$ , it follows that  $T_{\sqrt{\alpha}}$  must have points of period 6. Note that  $\sqrt{\alpha} = 1.27202\dots < \sqrt{2}$ , and this is where period-6 first appears.

**Remark 9.2.6** 1. Suppose that  $\mu > 1$  and  $\frac{\mu^2}{1+\mu^3} \leq \frac{1}{2}$ . Then  $\frac{\mu^3}{1+\mu^3} = 1 - \frac{1}{1+\mu^3} \geq \frac{1}{2}$ , so that

$$T_\mu\left(\frac{\mu}{1+\mu^3}\right) = \frac{\mu^2}{1+\mu^3}, \quad T_\mu\left(\frac{\mu^2}{1+\mu^3}\right) = \frac{\mu^3}{1+\mu^3} \quad \text{and} \quad T_\mu\left(\frac{\mu^3}{1+\mu^3}\right) = \frac{\mu}{1+\mu^3}.$$

This gives the 3-cycle:

$$\left\{ \frac{\mu}{1+\mu^3}, \frac{\mu^2}{1+\mu^3}, \frac{\mu^3}{1+\mu^3} \right\}.$$

This 3-cycle appears when  $\mu > 1$ , and  $\frac{\mu^2}{1+\mu^3} \leq \frac{1}{2}$ , or equivalently

$$\mu^3 - 2\mu^2 + 1 \geq 0, \quad \text{or} \quad (\mu-1)(\mu^2 - \mu - 1) \geq 0.$$

For the 3-cycle to appear, we require  $\mu \geq (1 + \sqrt{5})/2$ . A similar analysis can be done for other periodic orbits. For example, if  $\mu > 1$  and  $\frac{\mu^3}{1+\mu^4} \leq \frac{1}{2}$ ,  $\mu^3/(1+\mu^4)$  lies

on a 4-cycle, which occurs when  $\mu^3 - \mu^2 - \mu - 1 \geq 0$  ( $\mu \geq 1.83929$  approximately). However, there must be other 4-cycles, as we have seen that  $2^n$ -cycles ( $n \geq 1$ ) are created when  $\mu > 1$ .

2. Suppose that  $1 < \mu < 2$ . Then if  $x \in [\mu - \mu^2/2, \mu/2]$ , we can check that  $T_\mu(x) \in [\mu - \mu^2/2, \mu/2]$ , showing that the interval is an invariant set. For  $1 < \mu < \sqrt{2}$ , the smallest set invariant under  $T_\mu$  is a collection of subintervals of  $[\mu - \mu^2/2, \mu/2]$ . If  $\mu > \sqrt{2}$ , this smallest set becomes all of the interval  $[\mu - \mu^2/2, \mu/2]$ , called the *Julia set* of  $T_\mu$ . The Julia set is named after one of the early pioneers of chaotic dynamics, Gaston Julia, who worked on complex dynamics in the early 1900's. For  $\mu = 2$ , the Julia set is all of  $[0, 1]$ . The bifurcation diagram for  $T_\mu$ ,  $\mu > 1$  gives us some insight into the dynamics in this situation.
3. The conjugacy between  $T_2$  and  $L_4$  can be constructed by consideration of the periodic points of these maps. Since the period points are dense for each of these maps, by carefully ordering them according to their ordering in  $[0, 1]$ , we can define a map  $h$  by defining it on the periodic points.  $h$  is then defined on a dense subset of  $[0, 1]$ , into a dense subset. This map can be continuously extended to a homeomorphism of  $[0, 1]$  with  $h(0) = 0, h(1) = 1$ . Using these ideas, it can be shown that the conjugation between  $T_2$  and  $L_4$  is unique (see Exercises 9.1 # 8).
4. Proposition 9.2.2 shows that for  $1 < \mu \leq \sqrt{2}$ ,  $T_\mu^2$  restricted to the interval

$$\left[ \frac{\mu - 1}{\mu^2 - 1}, \frac{\mu^2 - \mu}{\mu^2 - 1} \right] = \left[ \frac{1}{\mu + 1}, \frac{\mu}{\mu + 1} \right],$$

is conjugate to  $T_{\mu^2}$  on  $[0, 1]$ . How do we arrive at this conjugacy? Notice that for  $\mu > 1$ ,  $T_\mu$  has a fixed point  $p_\mu = \mu/(\mu + 1)$ , and another point  $\hat{p}_\mu = 1/(\mu + 1)$  with  $T_\mu(\hat{p}_\mu) = T_\mu(p_\mu)$ , that is eventually fixed. If we look at the graph of  $T_\mu^2$  restricted to the interval  $[\hat{p}_\mu, p_\mu]$ , we see an “upside-down” version of  $T_\mu$ , and we consider the possibility that  $T_\mu^2$  restricted to the interval  $[\hat{p}_\mu, p_\mu]$  is actually conjugate to  $T_\mu$  (or in fact  $T_{\mu^2}$ ).

Define a linear map  $h_\mu : [\hat{p}_\mu, p_\mu] \rightarrow [0, 1]$  of the form  $h_\mu(x) = ax + b$  in such a way that  $h_\mu(p_\mu) = 0$  and  $h_\mu(\hat{p}_\mu) = 1$ . We can check that

$$h_\mu(x) = \frac{1}{\hat{p}_\mu - p_\mu}(x - p_\mu), \quad \text{and} \quad h_\mu^{-1}(x) = (\hat{p}_\mu - p_\mu)x + p_\mu.$$

$h_\mu$  expands the interval  $[\hat{p}_\mu, p_\mu]$  onto the interval  $[0, 1]$ , and changes the orientation. This is exactly the conjugacy defined in Proposition 9.2.2.

We define a *renormalization operator* of  $T_\mu$  by

$$(RT_\mu)(x) = h_\mu \circ T_\mu^2 \circ h_\mu^{-1}(x).$$

What we actually showed in the previous section is that  $(RT_\mu)(x) = T_{\mu^2}(x)$ , giving us the conjugacy claimed. This procedure can be continued for  $T_\mu^4$ ,  $T_\mu^8$  etc, and similar considerations can be made with the logistic map  $L_\mu$  (see [32] for more details).

### Exercises 9.2

1. (a) If  $\mu > 1$  and  $\frac{\mu^3}{1 + \mu^4} \leq \frac{1}{2}$ , show that

$$\left\{ \frac{\mu^3}{1 + \mu^4}, \frac{\mu^4}{1 + \mu^4}, \frac{\mu}{1 + \mu^4}, \frac{\mu^2}{1 + \mu^4} \right\},$$

is a 4-cycle for  $T_\mu$ , which occurs when  $\mu^3 - \mu^2 - \mu - 1 \geq 0$  ( $\mu \geq 1.83929$  approximately). Why does this not contradict our proof that  $T_\mu$  has  $2^n$ -cycles for  $n \geq 1$ , when  $\mu > 1$ ?

- (b) Do a similar analysis to find a 5-cycle.

- (c) Deduce that if  $\frac{\mu^{n-1}}{1 + \mu^n} \leq \frac{1}{2}$ , then  $\frac{\mu^{n-1}}{1 + \mu^n}$  is period- $n$  point for  $T_\mu$ , when  $\mu$  satisfies  $\mu^{n-1} - \mu^{n-2} - \mu^{n-3} - \dots - 1 \geq 0$ .

2. Show that when  $\mu > 1$ ,  $\frac{\mu - \mu^2 + \mu^4}{1 + \mu^4}$  is a period-4 point for  $T_\mu$ , and find the corresponding 4-cycle.

3. Consider the tent map  $T_{\sqrt{2}}$ .

- (i) Show that  $x = 1/2$  is an eventual fixed point for  $T_{\sqrt{2}}$ .

- (ii) Use the ideas of Example 9.2.5 to show that there is a subinterval of  $[0, 1]$  on which  $T_{\sqrt{2}}$  is chaotic, and show that  $T_{\sqrt{2}}$  has  $2k$ -cycles for any  $k \in \mathbb{Z}^+$ .

- (iii) Does  $T_{\sqrt{2}}$  have points of odd period greater than 1?

4. Prove Proposition 9.2.1 by using the fact that the maximum and minimum values a continuous function  $f(x)$  may take, occur either at an end point, or where  $f'(x) = 0$ , or where  $f'(x)$  does not exist.

5. Let  $1 < \mu < 2$ . Show that  $T_\mu : [\mu - \mu^2/2, \mu/2] \rightarrow [\mu - \mu^2/2, \mu/2]$  defines a dynamical system. Show that for  $1 < \mu < \sqrt{2}$ ,

$$\left[ \frac{1}{1+\mu}, \frac{\mu}{1+\mu} \right] \subseteq [\mu - \mu^2, \mu/2].$$

6\*. Let  $T_\mu$  be the tent map. Show that if  $\mu > \sqrt{2}$ , then for each open interval  $U \subset [0, 1]$ , there exists  $n > 0$  such that

$$[T_\mu^2(1/2), T_\mu(1/2)] \subseteq T_\mu^n(U).$$

(Hint: Use the fact that  $|T_\mu(U)| \geq \mu|U|$  if  $U$  does not contain  $1/2$ , so that the length keeps increasing. We claim there exists  $m > 0$  such that  $T^m(U)$  and  $T^{m+1}(U)$  both contain  $1/2$ , for if not,  $|T^{m+2}(U)| \geq \mu^2|U|/2$  for all  $m \in \mathbb{Z}^+$ , a contradiction, since this eventually exceeds 1).

## CHAPTER 10

### Fractals.

Many curves and surfaces that arise in nature, have been modeled mathematically by approximating them with smooth curves or surfaces. This is done so that calculus may be applied to their study. However, in recent years it has been realized that for certain of these objects, the calculus is not the best tool for their study.

For example, the motion of a particle suspended in a fluid (Brownian motion), the length of the coastline of an island or the surface area of the human lung. The length of a coast line is dependent on how carefully it is measured; for such an object the more closely you look at it, the more irregular it appears, and the greater the length appears to be. The study of such objects has resulted in a new area of mathematics called *Fractal Geometry*. Fractal geometry was popularized by the mathematician Benoit Mandelbrot, and it was he, who coined the term *fractal* in 1977 ([88]). Mandelbrot was originally from Poland, educated in France, and later moved to the United States. He is particularly famous for what is now called the *Mandelbrot set*, which we shall define in Chapter 14.

Much of the current interest in fractals is a consequence of Mandelbrot's work. His computer simulations of maps of the complex plane have resulted in extremely complicated and beautiful fractals. The mathematical work was initiated by Cayley, Fatou, and Julia in the late 19th and early 20th centuries, but progress slowed until the development of the electronic computer. Later, we shall see how fractals arise from the study of complex analytic maps and also from substitutions.

#### 10.1 Examples of Fractals.

Probably the first published example of a fractal was given by Karl Weierstrass in 1872. He constructed a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  which is continuous everywhere, but nowhere differentiable. Its graph has a *self-similarity* typical of fractals. The graph appears much the same no matter at what scale it is viewed. Until 1872, mathematicians believed that continuous functions had to be differentiable at “most” points, and this was often tacitly assumed. The Cantor set, named after George Cantor (but actually discovered by Henry Smith), is probably the simplest example

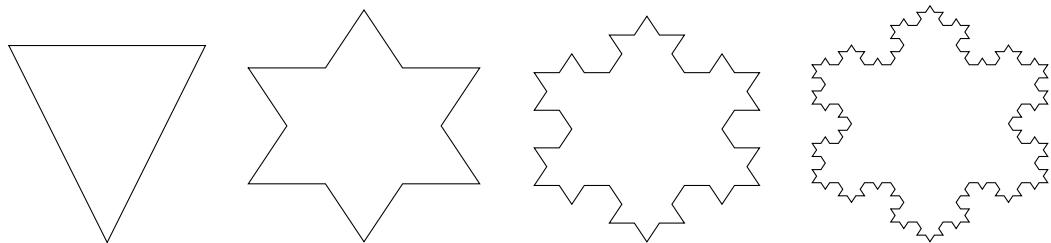
of a fractal. Henry Smith was an Oxford professor whose paper on the Riemann integral and discontinuous functions ([116]), appeared in 1874, whereas Cantor's paper appeared in 1883.

Another early example is due to Helge von Koch (1904), who constructed the famous *Koch Snowflake*. Starting with an equilateral triangle of side length 1 unit, three equilateral triangles are constructed (one on each side as shown), each having side length  $1/3$ . This construction is then continued so that at each stage, an equilateral triangle of side length one-third that of the previous triangles, is added to each exposed line. The Koch Snowflake itself is a limiting curve in a sense to be described shortly.

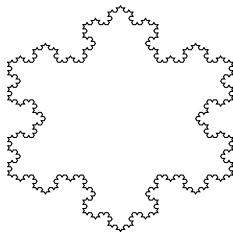
More explicitly, the Koch Snowflake curve is constructed inductively in the following way:

- (i) Start with an equilateral triangle with side length = 1. The perimeter of the resulting curve is  $L(0) = 3$ .
- (ii) Now add equilateral triangles of side length  $1/3$ . The perimeter is now  $L(1) = 3 \times 4 \times 1/3 = 4$ .
- (iii) Add new equilateral triangles as seen below by taking  $n = 2$ . The perimeter is now  $L(2) = 3 \times 4^2 \times (1/3^2) = 4^2/3$ .
- (iv) Continue in this way so that at the  $n$ th stage the perimeter is  $L(n) = 4^n/3^{n-1}$ .

For  $n = 1, 2, 3, 4, \dots$ , let us graph the first few steps in the construction, as these give us a good idea of what the limiting curve, (the Koch snowflake) looks like. As  $n \rightarrow \infty$ ,  $L(n) \rightarrow \infty$ , so that the Koch Snowflake is a curve having infinite length. In addition, the distance between any two points on the curve is infinite. The limiting curve is continuous, and it can be shown to be nowhere differentiable. The area enclosed by the curve is easily found and is clearly finite.



The first 4 iterations of the Koch Snowflake.



The Koch Snowflake.

The Koch snowflake is an example of a fractal, in the sense that it has the self-similarity property typical of fractals - as we zoom in on any part of the curve (no matter how far), it resembles (gives an exact copy), of what we saw previously. It is a fractal curve in  $\mathbb{R}^2$ . Typically these are continuous curves, differentiable nowhere and have infinite length.

## 10.2 An Intuitive Introduction to the Idea of Fractal Dimension.

The idea of (topological) dimension of an object in Euclidean space is intuitively clear. For example, a point or a finite number of points is 0-dimensional; a curve is 1-dimensional; a surface is 2-dimensional, etc. Thus, the topological dimension of the snowflake curve is 1. More precisely:

**Definition 10.2.1** A non-empty set  $K \subseteq \mathbb{R}^n$  has *topological dimension 0*, if for every point  $x \in K$  there is an open ball  $B_\delta(x)$  in  $\mathbb{R}^n$  having arbitrarily small radius, whose boundary does not intersect  $K$ .

$K$  has topological dimension  $k \in \mathbb{Z}^+$ , if every point  $x \in K$  is surrounded by an open ball  $B_\delta(x)$  having arbitrarily small radius, whose boundary intersects  $K$  in a set of topological dimension  $k - 1$ , and  $k$  is the least positive integer with this property.

Any discrete set such as the Cantor set or the set of rationals, will have topological dimension 0. The topological dimension of a line, a circle or the Koch curve is 1. A filled in circle or square will have topological dimension 2, and for a solid cube or sphere it will be 3. A smooth curve that we meet in calculus becomes a straight line upon repeated magnification (what we call *local linearity*). Fractals do not have this property, and the notion of dimension for fractals is not so simple.

It turns out that for sets such as the snowflake curve and the Cantor set, there is another very useful idea of dimension, originally called *Hausdorff-Besicovitch dimension*, and which we will call *fractal dimension* (following B. Mandelbrot). We motivate its definition as follows:

- (i) Given a piece of string, two copies of it result in a string “twice the size”.
- (ii) For a square we need 4 copies.
- (iii) For a cube we need 8 copies.
- (iv) For a 4-dimensional cube we need 16 copies. i.e., to double the (side length) of a  $d$ -dimensional cube we need  $c = 2^d$  copies, so that  $d = (\log c)/(\log 2)$ ,

Thus for example, if we have an object whose size doubles if three copies are stuck together, then it would have fractal dimension  $\log 3/\log 2$ .

Returning to the snowflake curve, notice that (one side of) it is made up of 4 copies of itself; each  $1/3$  of the size, so  $a = 3$ ,  $c = 4$  and

$$d = \frac{\log c}{\log a} = \frac{\log 4}{\log 3} = 1.2616\dots$$

The number  $d = \frac{\log 4}{\log 3}$  is called the *fractal dimension* of the snowflake curve.

**Definition 10.2.2** If we have a geometric object with a self-similarity for which  $c$  copies increase the size by a factor of  $a$ , then  $d = \frac{\log c}{\log a}$  is defined to be the *fractal dimension* of the object.

The Cantor set is seen to be made up of two copies of itself, each reduced in size by a factor of three, so that  $c = 2$ ,  $a = 3$  and  $d = \log 2/\log 3$ , again a fractal since the topological dimension of  $C$  is zero.

Roughly speaking, a fractal is defined to be a geometrical object whose fractal dimension is strictly greater than its (topological) dimension (this was the original definition due to Mandelbrot). The coastline of Britain has  $d = 1.25$  (approximately), so it is very rough. That of the U.S. is closer to 1 (fairly smooth).

Most fractals we shall meet, have the property of being self-similar, i.e., they do not change their appearance significantly when viewed under a microscope of arbitrary magnifying power. This self similarity can take a linear form as in the snow flake curve, or a non-linear form as in the Mandelbrot set (to be defined in Chapter 14). Unfortunately, the fractal dimension that we have defined, only makes sense for self-similar sets. We introduce a more general definition in the next section.

### 10.3 Box Counting Dimension.

We try to make the discussion of the last section more precise by introducing the notion of *box dimension*, first introduced in 1928 by Georges Bouligand, based on an

idea of Hermann Minkowski (1864 - 1909), who used balls rather than boxes. A *box* in  $\mathbb{R}^n$  is a set of the form

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] = \{(x_1, x_2, \dots, x_n) : a_i \leq x_i \leq b_i, i = 1, 2, \dots, n\}.$$

Let  $K \subseteq \mathbb{R}^n$  be a non-empty set, and let  $N_\delta(K)$  = the minimum number of boxes of equal side length  $\delta > 0$  needed to cover the set  $K$ . As  $\delta \rightarrow 0^+$ , more boxes will be needed, so that  $N_\delta(K)$  will increase. The idea is that as  $\delta$  decreases, the number of boxes required to cover  $K$  increases as some power of the length of the boxes. In particular, there is some  $d > 0$ , for which  $N_\delta \delta^d$  is asymptotic to 1, as  $\delta$  approaches 0.

This power  $d$  is the *box counting dimension*, which is defined more formally as follows:

**Definition 10.3.1** The *box counting dimension* of a non-empty set  $K \subseteq \mathbb{R}^n$  is given by

$$\dim(K) = \lim_{\delta \rightarrow 0^+} \frac{\log N_\delta(K)}{\log 1/\delta},$$

if this limit exists (it is independent of the base of the logarithm since  $\log_b(a) = \ln(a)/\ln(b)$  for  $a, b > 0, b \neq 1$ ).

Using balls instead of boxes (what is called *Minkowski dimension*), has the advantage that it can be used to define fractal dimension in a metric space. For our purposes it is more difficult to calculate. If the limit in Definition 10.3.1 does not exist, it is sometimes replaced by  $\limsup$  or  $\liminf$ .

**Examples 10.3.2** 1. Let  $I = [0, 1]$ , the unit interval in  $\mathbb{R}$ . If  $\delta > 0$  is small, we can cover  $I$  with  $1/\delta$  intervals of equal length. Thus  $N_\delta = 1/\delta$ , and then  $\frac{\log N_\delta}{\log 1/\delta} = 1$ , so the fractal dimension is one. In practice, we calculate this limit for some subsequence  $\delta_n \rightarrow 0$ , in order to avoid difficulties in dealing with the limit through continuous values (see Exercises 10.3 # 9). For  $I = [0, 1]$  we could proceed as follows:

We can cover  $I$  with one interval of length  $\delta = 1$ ; two intervals of length  $\delta = 1/2$ . In general, we need  $2^n$  intervals of length  $\delta = 2^{-n}$  so

$$\dim([0, 1]) = \lim_{n \rightarrow \infty} \frac{\log 2^n}{-\log 2^{-n}} = 1,$$

which is what we would expect. Strictly speaking, this only proves that  $\dim([0, 1]) \leq 1$ , since we have not shown that the covering gives the minimum value for  $N$ . We see that using sequences in this way is not an entirely satisfactory way of proceeding, but it works quite well in practice.

In a similar way, it can be shown that if  $K$  is a smooth curve, it will have box counting dimension 1, and for a smooth surface it will be 2, the same as the topological dimension. In particular,  $[0, 1]^2$  will have box counting dimension 2, and  $[0, 1]^3$  will have box counting dimension 3.

2. The Cantor set  $C$  has  $N_1(C) = 1$  (using the single interval  $[0, 1]$  to cover  $C$ ). Dividing  $[0, 1]$  into three equal subintervals, we can cover  $C$  with two intervals of length  $1/3$  and  $N_{1/3}(C) = 2$ .

Continuing in this way, we can cover  $C$  with  $2^n$  intervals of length  $3^{-n}$ , so that  $N_{3^{-n}}(C) = 2^n$  and

$$\dim(C) = \lim_{n \rightarrow \infty} \frac{\log 2^n}{-\log 3^{-n}} = \frac{\log 2}{\log 3}.$$

3. The *Menger sponge*  $M$ , is a 3-dimensional analog of the Cantor set. Start with the unit cube  $[0, 1]^3$ . Remove an (open) cube having one-third the side length in the center of each of the six faces, and also remove the central cube  $(1/3, 2/3)^3$ , so we have removed a total volume of  $7 \times 1/3^3$  cubic units. Continue in this way on each of the 20 cubes of side length  $1/3$  remaining, removing a cube of side length  $1/3^2$  from each face and also the central cube. Continue indefinitely. The resulting solid, the Menger sponge, can be shown to have box counting dimension equal to  $\dim(M) = \lim_{n \rightarrow \infty} \frac{\log 20^n}{\log 3^{-n}} = \frac{\log 20}{\log 3} \sim 2.73$ , since when  $\delta = 1$ ,  $N_1(M) = 1$ ; when  $\delta = 1/3$ ,  $N_\delta(M) = 20$ , and continuing this process gives the result.

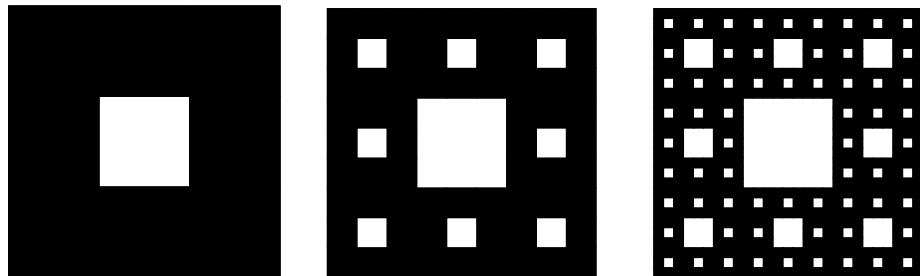
**Remark 10.3.3** A more complicated and more general notion than box dimension, due to Felix Hausdorff (1918), called the *Hausdorff dimension* is defined in more advanced treatise (see [52]).

### Exercises 10.3

1. Show that the length of the Koch Snowflake is infinite, and find the area enclosed by the curve.
2. A type of Cantor set  $K$  is obtained by removing *open middle halves* from the interval  $[0, 1]$ . First remove the interval  $(1/4, 3/4)$ , leaving  $[0, 1/4] \cup [3/4, 1]$ . Now remove the intervals  $(1/16, 3/16)$ , and  $(13/16, 15/16)$ , and continue in this way so

that at each stage we remove half of what is remaining. Find the fractal dimension of the resulting Cantor set  $K$ .

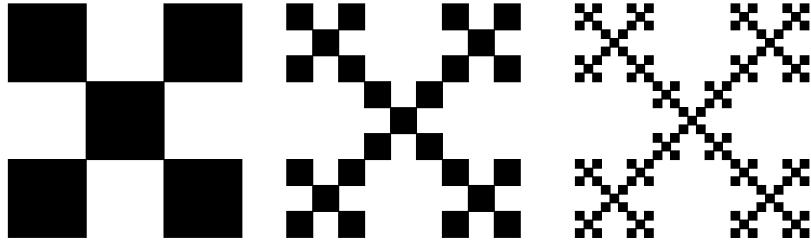
3. (a) Find the box counting dimension of the set  $\mathbb{Q} \cap [0, 1]$ .
- (b) Use the method of Example 10.3.2 # 1, to show that the unit square in  $\mathbb{R}^2$ ,  $[0, 1]^2$ , has box counting dimension 2.
4. The Sierpinski Carpet is a 2-dimensional version of the Cantor set and the Menger sponge. Start with the unit square  $[0, 1] \times [0, 1]$  partitioned into 9 equal squares. Remove the “open middle third” square  $(1/3, 2/3) \times (1/3, 2/3)$ . From each of the eight remaining squares of side length  $1/3$ , remove the open middle third squares, and continue indefinitely.
  - (a) Show that the resulting area removed, is equal to one square unit.
  - (b) Show that the box counting dimension of the Sierpinski Carpet is  $(\log 8)/(\log 3)$ .
  - (c) Show that the Sierpinski Carpet has no interior (i.e., contains no open balls in  $\mathbb{R}^2$ ).



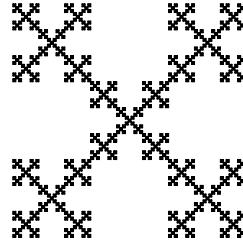
The first 3 iterations of the Sierpinski Carpet.

5. Show that the total volume removed in the construction of the Menger sponge  $M$ , is 1 cubic unit (so that its volume is 0). Show the box counting dimension of  $M$  is  $\log(20)/\log(3)$ .
6. Find the box counting dimension of the set  $\{0, 1, 1/2, 1/3, 1/4, \dots\}$ .

7. The first three steps in the construction of the fractal shown are indicated below. Determine the fractal dimension.



The first 3 iterations of the fractal.



The generated fractal.

8. The *Kronecker product*  $A \otimes B$  of an  $n \times m$  matrix  $A = [a_{ij}]$  and a  $k \times \ell$  matrix  $B$  is

$$\text{defined to be } A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & \dots & \dots & a_{2m}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}B & \dots & \dots & a_{nm}B \end{bmatrix}, \text{ an } nk \times m\ell \text{ matrix. Kronecker}$$

products have been used to generate fractals (see for example [66]). For example, let  $K = [0, 1]^2$  be the unit square in  $\mathbb{R}^2$ , and subdivide  $K$  into  $9 = 3 \times 3$  equal squares

and color black (1) or white (0) according to the matrix  $A_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ . Set

$A_2 = A_1 \otimes A_1$  and  $A_n = A_1 \otimes A_1 \otimes A_1 \otimes \dots \otimes A_1$  ( $n$ -times). Find  $A_2$  and  $A_3$  and deduce that coloring equal  $3^{2n} = 3^n \times 3^n$  squares of  $K$  according to the matrix  $A_n$  gives the  $n$ th approximation to the Sierpinski gadget.

- (a) A fractal is defined using the Kronecker product of the matrix  $G_1 = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ .

Find the Kronecker product  $G_1 \otimes G_1$ , and find the box dimension of the generated fractal.

- (b) Find matrices whose Kronecker product gives rise to (i) the Cantor set, (ii) the Sierpinski triangle.

9. The aim of this exercise is to justify the use of discrete sequences to calculate the box counting dimension of a fractal. Suppose that  $\{r_n\}$  is a positive decreasing sequence satisfying  $\lim_{n \rightarrow \infty} r_n = 0$ ,  $\lim_{n \rightarrow \infty} \frac{\log r_{n+1}}{\log r_n} = 1$  and  $\lim_{n \rightarrow \infty} \frac{\log N(r_n)}{-\log r_n} = \alpha$ . Show that  $\lim_{r \rightarrow 0^+} \frac{\log N(r)}{-\log r} = \alpha$ . (Hint: Show that  $\frac{\log N(r_n)}{-\log r_{n+1}} \leq \frac{\log N(r)}{-\log r} \leq \frac{\log N(r_{n+1})}{-\log r_n}$  for  $r_{n+1} < r \leq r_n$ , and that the first and last expressions both have limit  $\alpha$ ).

## 10.4 The Mathematical Theory of Fractals.

In the remainder of this chapter, we attempt to give a rigorous definition of the idea of a limiting set. For example, the Koch Snowflake is the limiting curve of the sequence of curves in  $\mathbb{R}^2$  shown in our first figure. We will give a precise definition of this limiting curve using properties of metric spaces. In order to do this, we need to define a metric on a space consisting of “sets of sets”, known as the *Hausdorff metric*. Then we can look at the convergence of sets (or rather sequences of sets), in the resulting space.

### 10.4.1 Complete Metric Spaces.

Given a sequence  $(x_n)$  in a metric space  $X$ , it is possible that it converges, but not to a point of  $X$ . For example, consider the sequence in  $\mathbb{Q}$  (the set of rationals with its usual metric  $d(x, y) = |x - y|$  for  $x, y \in \mathbb{Q}$ ), defined as:

$x_0 = 1$ ,  $x_1 = 1.1$ ,  $x_2 = 1.14$ ,  $x_3 = 1.141$ , and  $x_n$  being the member of  $\mathbb{Q}$  equal to  $1.141\dots$ , with a decimal expansion consisting of 1 followed by the first  $n$  terms in the decimal expansion of  $\sqrt{2}$ , followed by 0's. It is clear that  $(x_n)$  is a sequence in  $\mathbb{Q}$ , with  $\lim_{n \rightarrow \infty} x_n = \sqrt{2} \notin \mathbb{Q}$ . In the reals  $\mathbb{R}$  or the complex numbers  $\mathbb{C}$ , this type of difficulty cannot occur.  $\mathbb{R}$  and  $\mathbb{Q}$  are examples of what we call *complete metric spaces*, which we now define.

**Definition 10.4.2** Let  $(x_n)$  be a sequence in a metric space  $(X, d)$ . Then  $(x_n)$  is a *Cauchy sequence* if given any  $\epsilon > 0$ , there exists  $N \in \mathbb{Z}^+$ , for which

$$m, n > N \Rightarrow d(x_n, x_m) < \epsilon.$$

**Definition 10.4.3** The metric space  $(X, d)$  is said to be *complete* if every Cauchy sequence  $(x_n)$  in  $X$  converges to a member of  $X$ .

**Proposition 10.4.4** *In a metric space  $(X, d)$ , any convergent sequence is a Cauchy sequence.*

**Proof.** Let  $(x_n)$  be a convergent sequence, converging to  $x \in X$ . Then given  $\epsilon > 0$ , there exists  $N \in \mathbb{Z}^+$  such that

$$n > N \Rightarrow d(x, x_n) < \epsilon/2.$$

By the triangle inequality, if  $m, n > N$ , then

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \epsilon/2 + \epsilon/2 = \epsilon,$$

so that  $(x_n)$  is a Cauchy sequence. □

**Example 10.4.5** The sequence  $x_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}$ , is not a Cauchy sequence in  $\mathbb{R}$ .

**Proof.** A sequence  $(x_n)$  in a metric space  $(X, d)$  fails to be Cauchy if we can find  $\epsilon > 0$  such that for all  $N \in \mathbb{Z}^+$ , we can find  $x_n, x_m$  with  $n, m \geq N$  and  $d(x_n, x_m) \geq \epsilon$ . We shall see that it suffices to take  $\epsilon = 1/2$ . In this case,  $d(x_n, x_m) = |x_n - x_m|$ .

Set  $n = 2N$  and  $m = N$ , then

$$|x_{2N} - x_N| = \frac{1}{N+1} + \frac{1}{N+2} + \frac{1}{N+3} + \dots + \frac{1}{2N} > \frac{N}{2N} = \frac{1}{2},$$

so that  $(x_n)$  is not a Cauchy sequence. For this sequence,  $|x_n - x_{n-1}| = 1/n \rightarrow 0$  as  $n \rightarrow \infty$ , so we see that this property need not give rise to a Cauchy sequence. In particular, the sequence does not converge. □

The converse of Proposition 10.4.4 is not true. It can be seen that the rational numbers do not constitute a complete metric space, since for example, the sequence mentioned above, where  $x_n$  is the first  $n$  terms of the decimal expansion of  $\sqrt{2}$ , followed by 0's, can be seen to be a Cauchy sequence in  $\mathbb{Q}$  (with respect to the

usual metric), but does not converge in  $\mathbb{Q}$  (see also Exercises 10.4 # 2(c)). The reals  $\mathbb{R}$ ,  $\mathbb{R}^n$ , and the complex numbers  $\mathbb{C}$ , are examples of complete metric spaces. The completeness of  $\mathbb{R}$  is a consequence of the *Completeness Axiom* (see Appendix A), which says that any non-empty, bounded set  $S \subset \mathbb{R}$  has a *least upper bound* (called the *supremum*), and the completeness of  $\mathbb{R}^n$  and  $\mathbb{C}$  can be deduced from this. The supremum  $\alpha$  of a non-empty set  $S$  which is bounded above, is defined as follows:

$$\alpha = \sup(S) \text{ if } \alpha \geq x \text{ for all } x \in S, \text{ and if } \beta \geq x \text{ for all } x \in S, \text{ then } \alpha \leq \beta.$$

## Exercises 10.4

1. Use the definition to show that the following sequences  $(x_n)$  in  $\mathbb{R}$  are Cauchy sequences: (i)  $x_n = \frac{n+1}{n}$ , (ii)  $x_n = 1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n}$ , (iii)  $x_n = \sum_{k=1}^n \frac{(-1)^k}{k}$ .
2. (a) A sequence  $(x_n)$  in  $\mathbb{R}$  has the property that  $|x_{n+1} - x_{n+2}| \leq k|x_n - x_{n+1}|$ , for some  $0 < k < 1$  and  $n = 1, 2, \dots$ . Show that  $(x_n)$  is a Cauchy sequence.  
(b) Show that  $x_n = a^n$  is a Cauchy sequence for  $0 < a < 1$ .  
(c) Newton's method for  $f(x) = x^2 - 2$  gives rise to the sequence  $(x_n)$  satisfying  $x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n}$ . Set  $x_1 = 1$ . Show that  $(x_n)$  is a Cauchy sequence. Deduce the limit of the sequence. (Hint: Show that  $(x_{n+1})^2 - 2 \geq 0$ , ( $n \geq 2$ ), and then use (a)).
3. Show that the sequence  $(x_n)$ ,  $x_n = (-1)^n$  is not a Cauchy sequence.

## 10.5 The Contraction Mapping Theorem and Self-Similar Sets.

We shall use an important property of complete metric spaces concerned with those functions  $f : X \rightarrow X$  having the effect of bringing points closer together.

**Definition 10.5.1** A function  $f : X \rightarrow X$  on the metric space  $(X, d)$ , is called a *contraction mapping*, if there exists a real number  $\alpha$ ,  $0 < \alpha < 1$ , satisfying

$$d(f(x), f(y)) \leq \alpha d(x, y), \quad \text{for all } x, y \in X.$$

$\alpha$  is called the *contraction constant* of  $f$ .

We shall see that contraction mappings always have a fixed point. The following theorem is called the *Contraction Mapping Theorem*, or the *Banach Fixed-Point Theorem*, after the Polish mathematician Stefan Banach. Banach stated, and proved the theorem in 1922. The proof we shall give is due to R. Palais [97].

**Theorem 10.5.2** *Let  $f : X \rightarrow X$  be a contraction mapping on a non-empty, complete metric space  $X$ . Then  $f$  has a unique fixed point. In addition, if  $p$  is the fixed point,  $p = \lim_{n \rightarrow \infty} f^n(x)$  for every  $x \in X$ .*

**Proof.** Choose  $x_0 \in X$  arbitrarily, and set  $x_n = f^n(x_0)$ . We shall show that  $(x_n)$  is a Cauchy sequence, and hence is convergent. Suppose that  $0 < \alpha < 1$  has the property that  $d(f(x), f(y)) \leq \alpha d(x, y)$  for  $x, y \in X$ . Then inductively, we see that for any  $k \in \mathbb{Z}^+$

$$d(f^k(x), f^k(y)) \leq \alpha^k d(x, y).$$

By the triangle inequality, we have for  $x, y \in X$ :

$$\begin{aligned} d(x, y) &\leq d(x, f(x)) + d(f(x), f(y)) + d(f(y), y) \\ &\leq d(x, f(x)) + \alpha d(x, y) + d(f(y), y). \end{aligned}$$

Solving for  $d(x, y)$  gives the *Fundamental Contraction Inequality*:

$$d(x, y) \leq \frac{d(f(x), x) + d(f(y), y)}{1 - \alpha}.$$

Note that, if  $x$  and  $y$  are both fixed points, then  $d(x, y) = 0$ . So any fixed point is unique. If, in the Fundamental Inequality, we replace  $x$  and  $y$  by  $x_n = f^n(x_0)$  and  $x_m = f^m(x_0)$  respectively, then

$$\begin{aligned} d(x_n, x_m) &\leq \frac{d(f(x_n), x_n) + d(f(x_m), x_m)}{1 - \alpha} \\ &= \frac{d(f^n(f(x_0)), f^n(x_0)) + d(f^m(f(x_0)), f^m(x_0))}{1 - \alpha} \\ &\leq \frac{\alpha^n d(f(x_0), x_0) + \alpha^m d(f(x_0), x_0)}{1 - \alpha} \\ &= \frac{\alpha^n + \alpha^m}{1 - \alpha} d(f(x_0), x_0), \end{aligned}$$

Since  $\alpha^k$  becomes arbitrarily small for  $k$  large enough, the sequence  $\{x_n\}$  is Cauchy and hence  $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} f^n(x_0)$  exists. We showed earlier that if such a limit exists, the sequence must converge to a fixed point, say  $p$ , which must be unique by our earlier remark.

□

If we let  $m \rightarrow \infty$  in the last inequality, we get the rate at which  $f^n(x_0)$  converges to  $p$ :

$$\text{Corollary 10.5.3 } d(f^n(x_0), p) \leq \frac{\alpha^n}{1 - \alpha} d(f(x_0), x_0).$$

**Examples 10.5.4** 1. Let  $f : I \rightarrow I$  where  $I \subseteq \mathbb{R}$  is an interval. If  $|f(x) - f(y)| \leq \alpha|x - y|$  for all  $x, y \in I$ , where  $0 < \alpha < 1$ , then  $f$  is a contraction mapping, and  $f$  has a unique fixed point. The basin of attraction of  $f$  is all of  $I$ . This is no longer true if  $\alpha = 1$  (see Exercises 10.5). The functions  $L_\mu(x) = \mu x(1 - x)$ ,  $0 < \mu < 1$ , and  $f(x) = \cos(x)$  restricted to  $[0, 1]$  are easily seen to be contractions, so have a unique fixed point (as we have seen).

2. If  $f : X \rightarrow X$  is a map on a complete metric space for which  $f^k$  is a contraction for some  $k > 1$ , then it can be shown that  $f$  has a unique fixed point  $p$ , and for every  $x \in X$  we have  $f^n(x) \rightarrow p$  as  $n \rightarrow \infty$  (see Exercises 10.5). Note that a contraction mapping has to be continuous, but it is possible that  $f^k$  is a continuous contraction mapping without  $f$  being continuous ([28]).

The next theorem gives us the notion of a *self-similar set*. Here, we follow Hutchinson [71] (see also [76]). We use the concept of *compact set*, which we will not define for general metric spaces until Chapter 17. In the case of our main examples,  $X = \mathbb{R}^n$  (usually with  $n = 2$ ), by a compact set we mean one which is closed and bounded.

**Theorem 10.5.5 (Hutchinson's Theorem)** *Let  $(X, d)$  be a complete metric space. For each  $1 \leq i \leq N$ , let  $f_i : X \rightarrow X$  be a contraction mapping. Then there exists a unique non-empty compact subset  $K$  of  $X$  that satisfies*

$$K = f_1(K) \cup \cdots \cup f_N(K).$$

*K is called the self-similar set with respect to  $\{f_1, \dots, f_N\}$ .*

Before we can prove Theorem 10.5.5, we need some preliminary notions. Define a function on the subsets of  $X$  by

$$F(A) = \bigcup_{i=1}^N f_i(A), \quad \text{for } A \subseteq X.$$

The idea of the proof is to show the existence of a fixed point of  $F$  using the Contraction Mapping Theorem. This is what is called an *iterated function system* (IFS). To make this work, we must define a distance  $D$  on the set of non-empty subsets of  $X$ , so that with this distance, we have a complete metric space. We use the notion

of *Hausdorff metric*. Set

$$\mathcal{C}(X) = \{A : A \text{ is a non-empty compact subset of } X\},$$

so that  $F : \mathcal{C}(X) \rightarrow \mathcal{C}(X)$ . In Chapter 17, we will prove that  $F$  is well defined (the image of a compact set is a compact set).

$(\mathcal{C}(X), D)$  is a metric space if we define a metric  $D$  in the following way:

Suppose  $A \in \mathcal{C}(X)$  and  $\bar{B}_\delta(x) = \{y \in X : d(x, y) \leq \delta\}$  is the closed ball of radius  $\delta$  centered on  $x$ . Then set

$$U_\delta(A) = \bigcup_{y \in A} \bar{B}_\delta(y) = \{x \in X : d(x, y) \leq \delta \text{ for some } y \in A\}.$$

Think of  $U_\delta(A)$  as a closed set containing  $A$ , and whose boundary lies within  $\delta$  of  $A$ .

**Definition 10.5.6** The *Hausdorff metric*  $D$  is defined on  $\mathcal{C}(X)$  by

$$D(A, B) = \inf\{\delta > 0 : A \subseteq U_\delta(B) \text{ and } B \subseteq U_\delta(A)\}, \quad A, B \in \mathcal{C}(X).$$

We will show that  $D(A, B)$  defines a metric on  $\mathcal{C}(X)$ , but will omit the proof of completeness.

**Theorem 10.5.7** Let  $(X, d)$  be a metric space. Then the Hausdorff metric  $D(A, B)$  defines a metric on  $\mathcal{C}(X)$ . If the metric space  $(X, d)$  is complete, then so is the metric space  $(\mathcal{C}(X), D)$ .

**Proof.** From the symmetry of the definition of  $D$ , we see that  $D(A, B) = D(B, A) \geq 0$ , and also that  $D(A, A) = 0$ .

Suppose that  $D(A, B) = 0$ . Then for any  $n \in \mathbb{Z}^+$ ,  $A \subseteq U_{1/n}(B)$ . So that for any  $x \in A$ , we can find a sequence  $(x_n)$ :  $x_n \in B$  with  $d(x, x_n) \leq 1/n$ . Since  $B$  is a closed set (being compact), we have  $x \in B$  so that  $A \subseteq B$ . Similarly  $B \subseteq A$ , and  $A = B$ .

To prove the triangle inequality, let  $A, B$  and  $C$  belong to  $\mathcal{C}(X)$ , and suppose that  $D(A, B) < \delta$ , and  $D(B, C) < \epsilon$ . Then  $U_{\delta+\epsilon}(A) \supseteq C$  and  $U_{\delta+\epsilon}(C) \supseteq A$ , so that

$$D(A, C) < \delta + \epsilon = D(A, B) + \delta' + D(B, C) + \epsilon',$$

where  $\delta' = \delta - D(A, B) > 0$ , and  $\epsilon' = \epsilon - D(B, C) > 0$  are arbitrary. This gives

$$D(A, C) \leq D(A, B) + D(B, C).$$

□

We now show that a contraction on the metric space  $(X, d)$  gives rise to a contraction on  $(\mathcal{C}(X), D)$ .

**Proposition 10.5.8** Let  $f : X \rightarrow X$  be a contraction mapping on  $(X, d)$  with

$$d(f(x), f(y)) \leq \alpha \cdot d(x, y), \quad (0 < \alpha < 1), \quad \text{for all } x, y \in X.$$

Then

$$D(f(A), f(B)) \leq \alpha \cdot D(A, B) \quad \text{for all } A, B \in \mathcal{C}(X).$$

**Proof.** If  $\delta > D(A, B)$ , then  $U_\delta(A) \supseteq B$ , so that  $f(U_\delta(A)) \supseteq f(B)$ . Let  $z = f(y) \in f(U_\delta(A))$ . Then there exists  $x \in U_\delta(A)$  with  $d(x, y) \leq \delta$ . Since  $f$  is a contraction, this implies  $d(f(x), f(y)) \leq \alpha d(x, y) \leq \alpha\delta$ . Consequently,  $z = f(y) \in U_{\alpha\delta}(f(A))$  so

$$U_{\alpha\delta}(f(A)) \supseteq f(B).$$

In a similar way,  $U_{\alpha\delta}(f(B)) \supseteq f(A)$ , so that  $D(f(A), f(B)) \leq \alpha\delta$  and the result follows.  $\square$

**Proof of Theorem 10.5.5** We first show that if  $A_1, A_2, B_1, B_2 \in \mathcal{C}(X)$ , then

$$D(A_1 \cup A_2, B_1 \cup B_2) \leq \max\{D(A_1, B_1), D(A_2, B_2)\}.$$

To see this, suppose that  $\delta > \max\{D(A_1, B_1), D(A_2, B_2)\}$ , then

$$\delta > D(A_1, B_1) \geq \inf\{\delta' : U_{\delta'}(A_1) \supseteq B_1\}.$$

In particular,

$$U_\delta(A_1) \supseteq B_1 \quad \text{and} \quad U_\delta(A_2) \supseteq B_2,$$

implying that  $U_\delta(A_1 \cup A_2) \supseteq B_1 \cup B_2$ . A similar argument implies that  $U_\delta(B_1 \cup B_2) \supseteq A_1 \cup A_2$ , and  $\delta > D(A_1 \cup A_2, B_1 \cup B_2)$ . It follows that  $D(A_1 \cup A_2, B_1 \cup B_2) \leq \max\{D(A_1, B_1), D(A_2, B_2)\}$ .

Now using this result repeatedly, we have

$$D(F(A), F(B)) = D(\bigcup_{j=1}^N f_j(A), \bigcup_{j=1}^N f_j(B)) \leq \max_{1 \leq j \leq N} D(f_j(A), f_j(B)).$$

By Proposition 10.5.5,  $D(f_i(A), f_i(B)) \leq \alpha_i D(A, B)$ , where  $\alpha_i$  are the contraction constants for the maps  $f_i$ . If  $\alpha = \max_{1 \leq i \leq N} \alpha_i$ , then

$$D(F(A), F(B)) \leq \alpha D(A, B),$$

so that  $F$  is a contraction on  $(\mathcal{C}(X), D)$ , a complete metric space. The result follows by the contraction mapping theorem.  $\square$

**10.5.9 Examples of Iterated Function Systems.** 1. Take  $X = \mathbb{R}$  and denote by  $C$  the Cantor middle-thirds set. If we define  $f_i : \mathbb{R} \rightarrow \mathbb{R}$ ,  $i = 1, 2$ , by

$$f_1(x) = \frac{x}{3}, \quad f_2(x) = \frac{x}{3} + \frac{2}{3},$$

then  $f_1$  and  $f_2$  are contractions on  $\mathbb{R}$ . Define  $F$  on  $\mathcal{C}(\mathbb{R})$  by  $F(A) = f_1(A) \cup f_2(A)$  for all compact subsets  $A \subset \mathbb{R}$ , again a contraction. According to Hutchinson's Theorem, if  $J$  is any compact subset of  $\mathbb{R}$ ,  $\lim_{n \rightarrow \infty} F^n(J) = K$  is a unique fixed point of  $F$  in  $\mathcal{C}(\mathbb{R})$ . It is easy to check that  $C$  is a fixed point of  $F$ , so we must have  $K = C$ , the Cantor set. If we start with  $J = [0, 1]$ , then  $f_1(J) = [0, 1/3]$ ,  $f_2(J) = [2/3, 1]$  and  $F(J) = S_1$ , which is the first step in the construction of the Cantor set. Continuing in this way, we obtain  $F^n(J) = S_n$ , thus  $C$  is the limit of the sequence  $(S_n)$ , in the Hausdorff metric.

2. Set  $X = \mathbb{R}^2$ , and let  $K$  be the interior and boundary of the triangle having vertices  $(0, 0)$ ,  $(1, 0)$  and  $(1/2, \sqrt{3}/2)$ , an equilateral triangle. Define contractions  $f_j : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  by

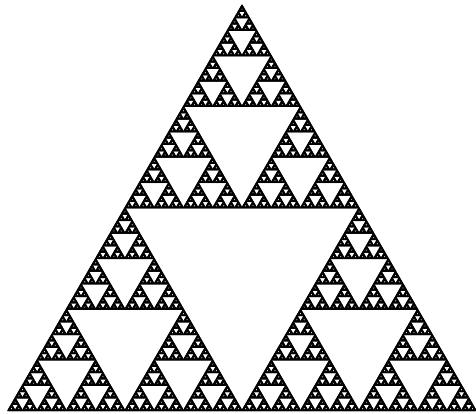
$$f_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{x}{2} \\ \frac{y}{2} \end{pmatrix}, \quad f_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{x+1}{2} \\ \frac{y}{2} \end{pmatrix}, \quad f_3 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{x+1/2}{2} \\ \frac{y+\sqrt{3}/2}{2} \end{pmatrix}.$$

Then if  $F = f_1 \cup f_2 \cup f_3$ ,  $F(K)$  is the union of the three equilateral triangles having vertices

$$(0, 0), (1/2, 0), (1/4, \sqrt{3}/4); \quad (1/2, 0), (1, 0), (3/4, \sqrt{3}/4); \quad \text{and}$$

$$(1/4, \sqrt{3}/4), (3/4, \sqrt{3}/4), (1/2, \sqrt{3}/2).$$

We have removed an interior triangle with vertices which are the mid-points of the sides of the triangle  $K$ , leaving a closed set which is the union of three equilateral triangles, and whose side lengths are one half the lengths of the original triangle. This completes the first stage in the construction of the *Sierpinski triangle*. The construction continues by performing the same operation on each of the three remaining triangles. Continuing to iterate  $K$  under  $F$ , gives rise to the Sierpinski triangle, as the unique fixed point of the map  $F$ .



The Sierpinski Triangle.

### Exercises 10.5

1. Show that the Cantor set is a fixed point of the map  $F$  defined in Example 10.5.9.
  
  
  
  
  
  
2. Show that the box counting dimension of the Sierpinski triangle is  $\log(3)/\log(2)$ .
  
  
  
  
  
  
3. Let  $f : I \rightarrow I$  where  $I$  is a non-empty subinterval of  $\mathbb{R}$ , and  $f$  has the property:

$$|f(x) - f(y)| < |x - y| \quad \text{for all } x, y \in I, \quad x \neq y$$

(note that  $f(x)$  is necessarily continuous on  $I$ ). Show:

- (a)  $f$  has at most one fixed point.
- (b) If  $I = [a, b]$ , ( $a < b$ ), then  $f(x)$  has at least one fixed point. (Hint: Consider  $\inf\{|f(x) - x| : x \in [a, b]\}$ . See also Theorem 17.3.1).
- (c)  $g(x) = x + 1/x$  has the above property on  $[1, \infty)$ , but has no fixed points.
- (d)  $h(x) = (x + \sin(x))/2$  has the above property on  $\mathbb{R}$ . Does  $h$  have a fixed point?

4. Let  $f(x) = x^2 - a$  with  $1 < a < 3$ , and let  $N_f$  be the corresponding Newton function. Show that  $N_f$  satisfies the hypothesis of the Contraction Mapping Theorem on  $(1, \infty)$ . What is the fixed point?
5. Brouwer's Fixed Point Theorem can be stated in the form: *Let  $f : \mathbb{D} \rightarrow \mathbb{D}$  be a continuous map, where  $\mathbb{D}$  is the closed unit ball in  $\mathbb{R}^n$ , then  $f$  has a fixed point in  $\mathbb{D}$  (e.g.  $\mathbb{D} = [0, 1] \subseteq \mathbb{R}$ , or  $\mathbb{D} = \{(x, y); x^2 + y^2 \leq 1\} \subseteq \mathbb{R}^2$ ).* We have seen this result for  $\mathbb{D} = [0, 1]$ .
- (a) Prove the following special case of Brouwer's Theorem: *If  $f : \mathbb{D} \rightarrow \mathbb{D}$ , where  $\mathbb{D} \subset \mathbb{R}^n$ , has the property that  $d(f(x), f(y)) \leq d(x, y)$  for all  $x, y \in \mathbb{D}$ , then  $f$  has a fixed point in  $\mathbb{D}$ .* We are using the standard metric on  $\mathbb{R}^n$ . (Hint: Consider  $g_k = (1 - 1/k)f$ , for  $k \in \mathbb{Z}^+$ ).
- (b) Give an example of a continuous, onto map,  $f : (0, 1) \rightarrow (0, 1)$  which does not have a fixed point.
6. (a) Let  $f : X \rightarrow X$  be a map on a complete metric space  $X$  for which  $f^k$  is a contraction for some  $k > 1$ . Show that  $f$  has a unique fixed point  $p$ , and for every  $x \in X$  we have  $f^n(x) \rightarrow p$  as  $n \rightarrow \infty$ . (Hint: Note that if  $p$  is a fixed point of  $f^k$ , then so is  $f(p)$ ).
- (b) Give an example of a map  $f : [0, 1] \rightarrow [0, 1]$  for which  $f^2$  is a contraction, but  $f$  is not continuous.
- (c) Set  $f(x) = e^{-x}$ ,  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Show that  $f$  is not a contraction, but  $f^2$  is a contraction with contraction constant  $\alpha < 1/e$ . (Hint: Use the Mean Value Theorem). (See [28]).
7. Find the distance between the sets  $A = \{0, 1/n, 2/n, \dots, (n-1)/n, 1\}$  and  $B = [0, 1]$ , in the Hausdorff metric. Deduce that the distance between an infinite set and a finite set can be arbitrarily small.

## CHAPTER 11

### Newton's Method for Real Quadratics and Cubics.

In Section 1.2 we introduced Newton's method for a differentiable function  $f : \mathbb{R} \rightarrow \mathbb{R}$  as a very efficient way of approximating the zeros of  $f$ . This is achieved by iterating the Newton function  $N_f$ , defined where  $f'(x) \neq 0$ , by

$$N_f(x) = x - \frac{f(x)}{f'(x)}.$$

Newton's method applied to quadratic polynomials is quite straightforward when the polynomial has two distinct real roots. We leave it as an exercise to show that for the quadratic polynomials

$$f(x) = ax^2 + bx + c, \quad \text{and} \quad g(x) = x^2 - \alpha, \quad a \neq 0,$$

(where  $a, b, c$  and  $\alpha$  are real),  $N_f$  and  $N_g$  are conjugate when  $\alpha = b^2 - 4ac$ . The conjugacy is given by the map  $h(x) = 2ax + b$ . It is then easy to see that for  $\alpha > 0$ , the basin of attraction of the fixed points  $\sqrt{\alpha}$  and  $-\sqrt{\alpha}$  of  $N_g$  are  $(0, \infty)$  and  $(-\infty, 0)$  respectively. When  $\alpha = 0$ , the only fixed point is 0, and its basin of attraction for  $N_g$  is all of  $\mathbb{R}$ . The conjugacy then shows that  $N_f$  has similar properties.

We now examine Newton's method for quadratics of the form  $f_c(x) = x^2 + c$  where  $c > 0$ , ( $f_c$  does not have any zeros in  $\mathbb{R}$ ). We start with a detailed look at the binary representation of real numbers. Subsequently, we will examine the situation for certain cubic polynomials. In Sections 11.1 and 11.2, we follow [26] and [101]. In Sections 11.3 and 11.4, we will follow [126] (see also [69]). This chapter is dependent on Chapters 1 and 2, but also uses the notions of conjugacy and linear conjugacy from Chapter 7, together with the notions of countability and Cantor set from Chapter 5. This chapter is essentially independent of the remaining chapters of this book, and may be omitted on a first reading.

#### 11.1 Binary Representation of Real Numbers.

Let  $b \in \mathbb{Z}^+$  be a given base. Any  $x \in \mathbb{R}$  may be represented in base  $b$  in the form

$$x = a_N a_{N-1} \dots a_1 a_0 \cdot c_1 c_2 c_3 \dots c_k c_{k+1} \dots$$

$$= a_N b^N + a_{N-1} b^{N-1} + \cdots + a_1 b + a_0 + \frac{c_1}{b} + \frac{c_2}{b^2} + \cdots + \frac{c_k}{b^k} + \cdots,$$

where  $a_i, c_j \in \{0, 1, \dots, b-1\}$  for all  $i$  and  $j$ .

We will restrict our attention to  $x \in [0, 1]$ , and the case  $b = 2$  (binary representation). However, the situation is much the same for any base  $b$ . When  $b = 2$ ,  $a_i, c_j \in \{0, 1\}$ .

**Definition 11.1.1** The *pre-period* in a binary expansion, is the number of terms after the *decimal point*, prior to the start of the periodic part of the expansion. The *period* is the minimal length of a repeating part of the binary expansion.

The following properties of binary expansions are well known (see for example [105]).

### 11.1.2 Properties of Binary Expansions.

Rational numbers either have finite binary expansions (dyadic rationals), or infinite periodic representations, which are in general, not unique. Irrationals have unique infinite non-periodic representations. We denote by  $\phi(n)$ , *Euler's function*, which gives the number of positive integers less than or equal to  $n$ , and are co-prime with  $n$ . We write  $(n, q) = 1$  if  $n$  and  $q$  are integers which are co-prime. More details for the case of rationals is given by the following properties:

- (a) *A rational  $r \in (0, 1)$  has a finite binary representation if and only if it is dyadic, i.e., it can be written as  $r = k/2^n$  for some  $k, n \in \mathbb{Z}^+$  where  $k$  is odd.*
- (b) *A rational  $r \in (0, 1)$  has an infinite repeating binary representation if and only if it can be written as  $r = t/q$  where  $q$  is odd. In this case, the pre-period is zero and the period of the repeating sequence does not exceed  $\phi(q)$  (when  $(t, q) = 1$ ), where  $\phi$  is Euler's function.*
- (c) *A rational  $r \in (0, 1)$  is of the form  $r = t/2^n q$ ,  $(t, 2^n q) = 1$ , where  $q$  is odd and  $n > 0$ , if and only if the binary expansion of  $r$  is eventually repeating, with repeating part having period  $\phi(q)$  and pre-period  $n$ .*

In cases (b) and (c) we get additional information:

- (b') *If  $r \in (0, 1)$  is of the form  $r = t/q$  ( $(t, q) = 1$ ,  $q$  odd), we may write  $r = s/(2^p - 1)$  for some (minimal), positive integers  $s$  and  $p$ . In this case,  $p$  gives the period of the binary representation.*

(c') If  $r \in (0, 1)$  is of the form  $r = t/2^n q$  ( $t, 2^n q = 1$ ,  $q$  odd), we may write  $r = s/2^n(2^p - 1)$  for some (minimal), positive integers  $n$ ,  $s$  and  $p$ . Again,  $p$  gives the length of the period of the binary representation and  $n$  is the pre-period.

- Examples 11.1.3**
1.  $r = 11/16 = 1/2 + 1/2^3 + 1/2^4 = \cdot 1011$  is a dyadic rational.
  2. If  $r = 1/5 = 0 \cdot \overline{0011}$  (where the overline indicates that this part is repeated indefinitely), then the period is  $\phi(5) = 4$ , whereas  $r' = 1/13 = 0 \cdot \overline{000100111011}$  has period  $\phi(13) = 12$ .
  3. We have  $r = 1/(2 \cdot 5) = 0.\overline{000011}$  with period 4 (same as that of  $1/5$ ), and pre-period 1 since  $2^1 = 2$ .
  4.  $1/5 = 3/(2^4 - 1) = 0 \cdot \overline{0011}$ ,  $1/13 = 5 \cdot 63/(2^{12} - 1) = 0 \cdot \overline{000100111011}$ .
  5.  $1/12 = 1/(2^2(2^2 - 1)) = 0 \cdot \overline{00\overline{01}}$ ,  $1/14 = 1/2(2^3 - 1) = 0 \cdot \overline{\overline{001}}$ .

In summary we have:

- (1) If  $r \notin \mathbb{Q} \cap [0, 1]$ , then  $r$  has a unique binary representation which is infinite and non-periodic.
- (2) If  $r \in \mathbb{Q} \cap [0, 1]$ , then either
  - (i) there exists  $k, n \in \mathbb{Z}^+$  such that  $r = k/2^n$ , so that  $r$  has a binary representation that terminates after  $n$  digits, or
  - (ii) there exists  $k, n \in \mathbb{Z}^+$ ,  $p \geq 0$ , such that  $r = k/(2^p(2^n - 1))$ , so that  $r$  has a unique binary representation with pre-period  $p$  and period  $n$ .

## 11.2 Newton's Method for Real Quadratic Polynomials.

We return to Newton's method for the functions of the form  $f_c(x) = x^2 + c$ ,  $c > 0$ . In particular, consider  $f_1(x) = x^2 + 1$ . The Newton's function is

$$N_1(x) = x - \frac{f_1(x)}{f'_1(x)} = \frac{x^2 - 1}{2x}.$$

The sequence  $N_1^n(x)$  cannot converge, since it would have to converge to a fixed point of  $N_1$ , and there are no fixed points since  $f_1$  is never zero. We set  $N_1(0) = \infty$  for convenience, and say that the orbit of  $x$  terminates if  $N_1^n(x) = 0$ , for some  $n$ . For example,  $N_1(1) = 0$  and  $N_1^2(1 + \sqrt{2}) = 0$ .

On the other hand,  $N_1(1/\sqrt{3}) = -1/\sqrt{3}$  and  $N_1(-1/\sqrt{3}) = 1/\sqrt{3}$ , giving a point of period 2, and since  $N_1(\sqrt{3}) = 1/\sqrt{3}$ , we have an eventually periodic point.

Notice that the recurrence relation

$$x_{n+1} = \frac{x_n^2 - 1}{2x_n}$$

is reminiscent of the trigonometric identity:

$$\cot(2\theta) = \frac{\cot^2(\theta) - 1}{2\cot(\theta)}, \quad \theta \in (0, \pi), \quad \theta \neq \pi/2.$$

The map  $h(\theta) = \cot(\pi\theta)$  is a homeomorphism  $h : (0, 1) \rightarrow \mathbb{R}$ , so any conjugation involving  $h$  will preserve orbits of  $N_1$  in the usual way. Thus, if  $x_0 = \cot(\pi r_0)$ , then  $N_1^n(x_0) = \cot(2^n\pi r_0)$  for each  $n$ , provided that  $2^n\pi r_0$  is not an integer multiple of  $\pi$ , i.e.,  $r_0$  is not of the form  $k/2^n$  for some  $k, n \in \mathbb{Z}^+$ . In other words,  $N_1 \circ h(r) = h \circ T(r)$  if  $r \neq 1/2$ , where  $T : (0, 1) \rightarrow (0, 1)$  is given by  $T(x) = 2x \pmod{1}$ , the doubling map restricted to the interval  $(0, 1)$ . It can be seen that  $N_1$  (on the set of numbers  $x_0$  whose orbits do not terminate at 0), is conjugate to the map

$$T' : (0, 1) \setminus \{\text{dyadic rationals}\} \rightarrow (0, 1) \setminus \{\text{dyadic rationals}\}, \quad T'(x) = 2x \pmod{1}.$$

Motivated by this discussion, we have:

**Proposition 11.2.1** (i) *If  $r_0 \in (0, 1)$ , then  $r_0 = k/2^n$  for some  $k, n \in \mathbb{Z}^+$ , with  $k$  odd, if and only if the orbit of  $x_0 = \cot(\pi r_0)$  under  $N_1$ , terminates at 0 after  $n$  iterations.*

(ii) *The orbit  $O(x_0) = \{N_1^n(x_0) : n \in \mathbb{Z}^+\}$  is either finite or infinite (periodic/eventually-periodic), if and only if  $r_0$  is rational.*

(iii) *If  $r_0 \notin \mathbb{Q} \cap (0, 1)$ , then the orbit of  $x_0 = \cot(\pi r_0)$  under  $N_1$  is infinite, and there are points  $x_0$  having a dense orbit (so that  $N_1 : \mathbb{R} \rightarrow \mathbb{R}$  is topologically transitive).*

**Proof.** (i) If  $r_0 = k/2^n$  with  $k$  odd, then

$$x_{n-1} = \cot(2^{n-1}\pi r_0) = \cot(\pi k/2) = 0,$$

so that  $x_m$  is not defined for  $m \geq n$ , and the orbit of  $x_0 = \cot(\pi r_0)$  terminates at the fixed point 0 after  $n$  iterations.

Conversely, if the orbit of  $x_0$  terminates at 0 after  $n$  iterations, then

$$N_1^n(x_0) = \cot(2^n\pi r_0) = 0,$$

so there exists  $m \in \mathbb{Z}$  with  $2^n\pi r_0 = m\pi + \pi/2$ , so  $r_0 = (2m + 1)/2^{n+1}$ .

(ii) If  $r_0$  gives rise to a periodic or eventually periodic orbit, then it cannot be dyadic: there must exist  $m$  and  $p$  with

$$N_1^{m+p}(x_0) = N^m(x_0), \quad \text{or} \quad \cot(2^{m+p}\pi r_0) = \cot(2^m\pi r_0),$$

and solving gives  $r_0 = k/2^m(2^p - 1)$ ,  $k, p, m \in \mathbb{Z}^+$ ,  $p \geq 2$ .

(iii) We have a conjugacy between  $N_1$  and  $T$ , the doubling map (at least if we exclude the dyadic rationals). Since the orbit of any irrational  $r_0$  under  $T$  will be infinite, the corresponding orbit will be infinite under  $N_1$ . In addition, if  $r_0$  has a dense orbit in  $(0, 1)$  under  $T$  (such orbits exist since  $T$  is transitive), then  $x_0 = \cot(\pi r_0)$  will have a dense orbit under  $N_1$ .  $\square$

**Examples 11.2.2** 1. If  $r_0 = 1/3 = 1/(2^2 - 1) = 0 \cdot \overline{01}$ , then  $m = 0, p = 2$  and  $x_0 = \cot(\pi/3) = 1/\sqrt{3}$ ,  $x_1 = \cot(2\pi/3) = -1/\sqrt{3}$ . We see that the orbit of  $x_0$  is periodic with period  $p = 2$ .

2. If  $r_0 = 1/7 = 1/(2^3 - 1) = 0 \cdot \overline{001}$ , then  $m = 0, p = 3$  and  $x_0 = \cot(\pi/7)$ , giving rise to the 3-cycle  $\{\cot(\pi/7), \cot(2\pi/7), \cot(4\pi/7)\}$ .

**Remark 11.2.3** In the case of  $f_c(x) = x^2 + c$  for  $c \in \mathbb{R}$  with  $c < 0$ , we see that  $f_c$  has two roots  $\pm\sqrt{-c}$ . The corresponding dynamics of  $N_c$  is trivial since the respective basins of attraction are  $(-\infty, 0)$  and  $(0, \infty)$ , with  $N_c(0)$  being undefined.

If  $c > 0$ , then a change of variable can be shown to reduce the situation to the case where  $c = 1$ . When  $c = 0$ ,  $N_0(x) = x/2$ , so the dynamics is trivial: the orbit of any point tends to the origin.

### 11.3 Newton's Method for Real Cubic Polynomials.

In this section we restrict ourselves to cubic polynomials. We follow the development in Walsh [126]. Let  $f(x)$  be a cubic polynomial and  $N_f(x) = x - f(x)/f'(x)$ , the corresponding Newton's function. We first show that it suffices to consider only *monic polynomials* (cubics for which the coefficient of  $x^3$  is 1). We then show that shifting the polynomial (for example, replacing  $x$  by  $x - h$ ), does not change the dynamics of Newton's method. In this way, the study of the dynamics of Newton's method for any cubic polynomial can be reduced to the study of a few special cases.

**Proposition 11.3.1** *Let  $f(x)$  be a cubic polynomial with corresponding Newton's function  $N_f(x)$ . Then*

- (i) If  $k \in \mathbb{R}$ ,  $k \neq 0$ , and  $g(x) = kf(x)$ , then  $N_f(x) = N_g(x)$ .
- (ii) Denote by  $A(x) = ax + b$ ,  $a, b \in \mathbb{R}$ ,  $a \neq 0$ , an affine transformation. If  $g(x) = f(A(x))$ , then  $AN_gA^{-1}(x) = N_f(x)$ , i.e.,  $N_f$  and  $N_g$  are conjugate via  $A$ .
- (iii) Let  $f_{a,c}(x) = (x - a)(x^2 + c)$ . There exists  $a, c \in \mathbb{R}$  such that  $N_f(x)$  is conjugate to  $N_{f_{a,c}}(x)$ .

**Proof.** (i) This is straightforward since  $g'(x) = kf'(x)$ .

(ii) We show that  $AN_g(x) = N_fA(x)$ . The right-hand side is

$$N_fA(x) = A(x) - \frac{f(A(x))}{f'(A(x))} = ax + b - \frac{f(A(x))}{f'(A(x))}.$$

The left-hand side is

$$AN_g(x) = A\left(x - \frac{g(x)}{g'(x)}\right) = A\left(x - \frac{f(A(x))}{af'(A(x))}\right) = a\left(x - \frac{f(A(x))}{af'(A(x))}\right) + b,$$

and these are equal.

(iii) From (i), we may assume that the polynomial is monic, say

$$f(x) = x^3 + kx^2 + mx + n.$$

Replacing  $x$  by  $x - h$  gives

$$\begin{aligned} g(x) &= f(x - h) = (x - h)^3 + k(x - h)^2 + m(x - h) + n \\ &= x^3 + (k - 3h)x^2 + (3h^2 - 2hk + m)x - h^3 + kh^2 - mh + n. \end{aligned}$$

It can be shown (for example, using a computer algebra system), that  $h$  can be chosen so that the other coefficients give the form:

$$g(x) = x^3 - ax^2 + cx - ac = (x - a)(x^2 + c).$$

Now use (ii) to see that  $N_f$  and  $N_g$  are conjugate, where  $g = f_{a,c}$ . □

**Examples 11.3.2** 1. If  $L_\mu(x) = \mu x(1 - x)$ ,  $\mu \neq 0$ , then Proposition 11.3.1 shows that the Newton functions  $N_{L_\mu}$  are independent of  $\mu$ . These logistic maps need not be conjugate, so that the Newton functions  $N_f$  and  $N_g$  may be conjugate, without  $f$  and  $g$  being conjugate. Conversely, conjugacy between  $f$  and  $g$  does not imply conjugacy between  $N_f$  and  $N_g$ .

2. Let  $f(x) = \sin(x)$  and  $g(x) = \cos(x)$ . Since  $g(x) = f(x + \pi/2)$ , Proposition 11.3.1 tells us that  $N_f$  and  $N_g$  are conjugate.

### 11.4 The Cubic Polynomials $f_c(x) = (x + 2)(x^2 + c)$ .

Proposition 11.3.1 tells us that for any cubic polynomial  $p(x)$ , there exist  $a, c \in \mathbb{R}$  such that the Newton map  $N_p$  is conjugate to  $N_{f_{a,c}}$ , where

$$f_{a,c}(x) = (x - a)(x^2 + c).$$

For this reason (following [126]), when we investigate Newton's method for cubic polynomials of the form  $f(x) = (x+2)(x^2+c)$ , there is not too much loss of generality.

We first look at the case where  $c < 0$ , (so  $f(x) = 0$  when  $x = -2$  and  $x = \pm\sqrt{-c}$ ). The case where  $c > 0$  is then considered, so  $f(x) = 0$  has a single real root at  $x = -2$ .

In each case, Newton's function is

$$N_c(x) = x - \frac{f_c(x)}{f'_c(x)} = \frac{2x^3 + 2x^2 - 2c}{3x^2 + 4x + c},$$

a one-parameter family of rational functions whose fixed points are the zeros of  $f_c(x)$ . Since  $N'_c(x) = f''_c(x)f_c(x)/(f'_c(x))^2$ , it has critical points at the zeros of  $f_c(x)$ , together with  $x = -2/3$ , the root of  $f''_c(x) = 0$ .

#### Case 1: $c < 0$ .

Here  $f_c(x) = 0$  when  $x = -2$  or  $x = \pm\sqrt{-c}$ , three distinct roots. Let us take  $c = -1$  and write  $f(x) = f_{-1}(x)$  and  $N(x) = N_{-1}(x)$ , so that the fixed points of  $N$  are  $-2$  and  $\pm 1$ . Denote the immediate basin of attraction of a fixed point  $p$  by  $W(p)$  (the largest interval containing  $p$ , contained in the basin of attraction of  $p$ ). If  $e_1$  and  $e_2$  are the critical points of  $f$ , then we make the following claim:

**Claim 1.**  $W(1) = (e_2, \infty)$ , and  $W(-2) = (-\infty, e_1)$ , where  $e_1 < e_2$ .

Note that

$$N(x) = \frac{2x^3 + 2x^2 + 2}{3x^2 + 4x - 1} = 1, \quad \text{when } x = 1,$$

so that  $N(x) > 1$  when

$$2x^3 - x^2 - 4x + 3 = (x - 1)^2(2x + 3) > 0.$$

It follows that for  $e_2 < x < 1$ ,  $N(x) > 1$ , and the same is true when  $x > 1$ . Now  $N(x) = x - f(x)/f'(x) < x$ , for  $x > 1$ , since both  $f(x)$  and  $f'(x)$  are positive for these values of  $x$ . In other words, if  $x \in (e_2, 1)$ , then  $N(x) \in (1, \infty)$ , and if  $x \in (1, \infty)$ ,  $N^n(x)$  is a decreasing sequence bounded below by 1, so must converge to a fixed point of  $N(x)$ , and therefore must converge to 1. This shows  $W(1) = (e_2, \infty)$ . A similar argument may be made to show  $W(-2) = (-\infty, e_1)$ .

Since  $x = -1$  is a (super) attracting fixed point for  $N$  ( $N'(-1) = 0$ ), and  $N$  is continuous on  $(e_1, e_2)$ ,  $W(-1) = (a, b)$  is an open interval for some  $a, b \in \mathbb{R}$  with  $e_1 < a < b < e_2$ .

As we have seen previously, we must have  $N(a, b) = (a, b)$ , and from the continuity, we have  $N(a) = b$  and  $N(b) = a$  (so  $\{a, b\}$  is a 2-cycle), since neither  $a$  nor  $b$  is a fixed point, or an eventually fixed point of  $N$ . We shall see that there are no other periodic points, and that  $\{a, b\}$  is a repelling 2-cycle.

**Claim 2.**  $N(e_1, a) = (b, \infty)$  in a one-to-one fashion, so that some points in  $(e_1, a)$  map onto  $W(1)$ , the immediate basin of attraction of 1, and are consequently in the basin of attraction of 1. In other words, there is a point  $e_3 \in (e_1, a)$  with  $N(e_3) = e_2$ , and  $N(e_1, e_3) = (e_2, \infty)$ , and  $N(e_3, a) = (b, e_2)$ .

To see this, notice that  $f$  has a point of inflection at  $x = -2/3$ , and  $N'(x) = f''(x)f(x)/(f'(x))^2 < 0$  on  $(e_1, -1) \cup (-2/3, e_2)$ , since on  $(e_1, -1)$ ,  $f(x) > 0$  with  $f$  concave down, and on  $(-2/3, e_2)$ ,  $f(x) < 0$  with  $f$  concave up. In particular,  $N$  is decreasing on  $(e_1, a)$ ,  $N(a) = b$ , and we can think of  $N(e_1)$  as being  $+\infty$ .

In a similar way, there is  $e_4 \in (b, e_2)$  with  $N(e_4) = e_1$ , and since  $N(e_3, a) = (b, e_2)$ , there is  $e_5 \in (e_3, a)$  with  $N(e_5) = e_4$ . Continuing in this way, we can find a sequence  $(e_n)_{n=1}^\infty$  with

$$e_1 < e_3 < e_5 < \dots < a, \quad \text{and} \quad e_2 > e_4 > e_6 > \dots > b,$$

and

$$N(e_{2n+1}) = e_{2n} \quad \text{and} \quad N(e_{2n+2}) = e_{2n-1}, \quad n \geq 1,$$

so that  $N(e_1, a) = (b, \infty)$ ,  $N(e_3, a) = (b, e_2)$  with  $(e_1, e_3), (e_5, e_7), \dots$  contained in the basin of attraction of  $x = 1$ , and  $(e_4, e_2), (e_6, e_5), \dots$ , contained in the basin of attraction of  $-2$ . The sequences  $(e_{2n})$  and  $(e_{2n-1})$ , must converge, and we can argue that they must converge to a 2-cycle of  $N$ , which has to be  $\{a, b\}$ . In other words,

$$e_{2n-1} \rightarrow a \quad \text{and} \quad e_{2n} \rightarrow b \quad \text{as} \quad n \rightarrow \infty.$$

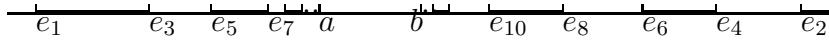
In addition, the basin of attraction of  $x = 1$  contains the disjoint union  $(e_1, e_3) \cup (e_5, e_7) \cup (e_6, e_4) \cup \dots$ , and the basin of attraction of  $x = -2$  contains the disjoint union  $(e_4, e_2) \cup (e_8, e_6) \cup (e_3, e_5) \cup \dots$

In summary,

$$B_N(1) = (e_2, \infty) \cup \bigcup_{n=1}^{\infty} (e_{4n-3}, e_{4n-1}) \cup \bigcup_{n=1}^{\infty} (e_{4n+2}, e_{4n}),$$

$$B_N(-2) = (-\infty, e_1) \cup \bigcup_{n=1}^{\infty} (e_{4n-1}, e_{4n+1}) \cup \bigcup_{n=1}^{\infty} (e_{4n}, e_{4n-2}),$$

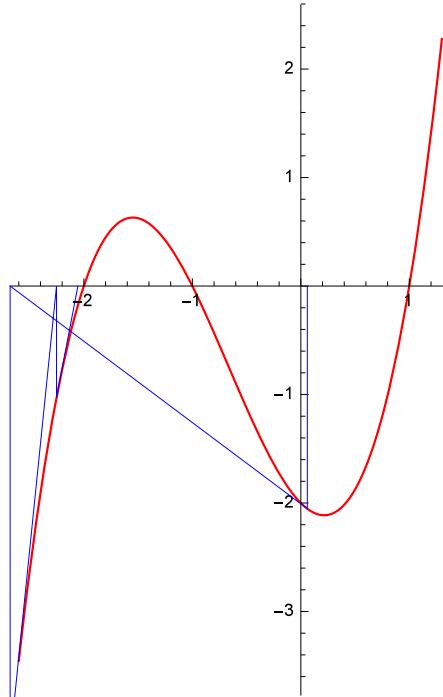
$$B_N(-1) = (a, b).$$



Double lines are the basin of attraction of 1, single lines give the basin of attraction of -2.

We observe that the set

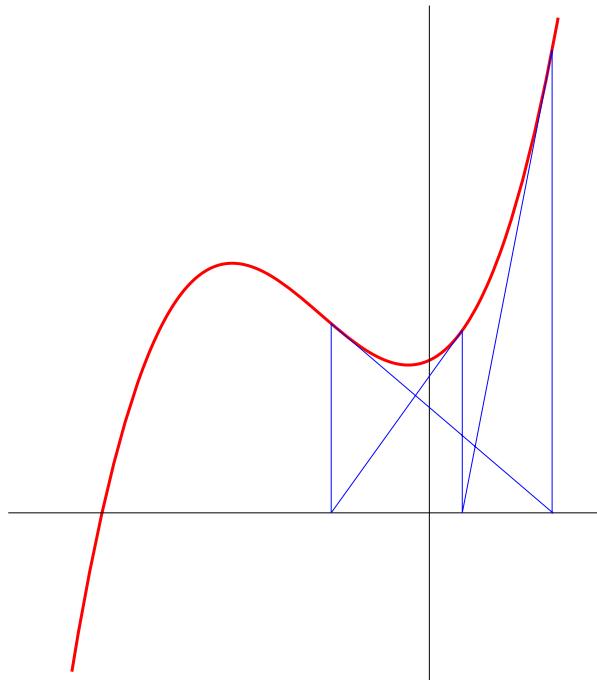
$E = \{x \in \mathbb{R} : \text{the sequence } N^n(x) \text{ does not converge to a root of } f(x) \text{ as } n \rightarrow \infty\}$ , consists of the 2-cycle  $\{a, b\}$ , and the sequence  $(e_n)_{n=1}^{\infty}$ . Of course, the critical points of  $f$  and the *eventual critical points* (which is where the sequence  $(e_n)_{n=1}^{\infty}$  arises), will always lie in  $E$ . Consequently, Newton's method gives convergence to a root of  $f(x)$  everywhere except on a countable set.



The graph of  $f(x) = (x + 2)(x^2 - 1)$  has critical points  $e_1 = .21525$  and  $e_2 = -1.54858$ . For  $x$  close to  $e_1$ ,  $x > e_1$ ,  $N_f^n(x)$  approaches  $x = 1$ , and if  $x < e_1$ ,  $N_f^n(x)$  approaches  $-2$ .

**Case 2:  $c > 0$ .** The situation is now more complicated. For example, we can see graphically that as  $c$  decreases from 1 to 0, a 3-cycle is born at around  $c = .466$ . Also, when  $c = 1/5$ ,  $\{1/10, -3/5\}$  (and also  $\{0.093561\dots, -0.634223\dots\}$ ) is a 2-cycle. Sharkovsky's Theorem does not apply in the usual way to  $N_f(x)$ , since in general  $N_f$  is not a continuous function, but other results can be used to show that when we have a 3-cycle, we have cycles of all other possible orders.

□



A 3-cycle appears for the Newton function of  $f(x) = (x + 2)(x^2 + c)$ , when  $c = .466\dots$

The situation becomes considerably more complicated for polynomials of degree 4 or more. The following result is due to Barna [9]:

**Theorem 11.4.1** *Let  $f$  be a  $n$ th degree polynomial with  $n > 3$  distinct real roots. The set  $E = \{x \in \mathbb{R} : \text{the sequence } N^n(x) \text{ does not converge to a root of } f(x) \text{ as } n \rightarrow \infty\}$ , is a Cantor set. In particular, it is uncountable, has measure zero, is closed, totally disconnected, and each point of  $E$  is a limit point of  $E$ .*

## Exercises 11.4

1. Prove that if  $r \in (0, 1)$  has a finite binary representation, then  $r = k/2^n$  for some  $k, n \in \mathbb{Z}^+$ , where  $k$  is odd. Now prove the converse.
2. Show that  $1/13 = 5 \cdot 63/(2^1 2 - 1) = 0 \cdot \overline{000100111011}$ .
3. Prove that  $r \in (0, 1)$  has an infinite repeating binary representation, if and only if  $r = t/q$  for some  $t, q \in \mathbb{Z}^+$ , where  $q$  is odd.
4. Show that for the quadratic polynomials

$$f(x) = ax^2 + bx + c, \quad \text{and} \quad g(x) = x^2 - \alpha, \quad a \neq 0,$$

the Newton's functions  $N_f$  and  $N_g$  are conjugate when  $\alpha = b^2 - 4ac$ . The conjugacy is given by  $h(x) = 2ax + b$ , so that  $h \circ N_f(x) = N_g \circ h(x)$  for all  $x$  in the domain of  $N_f$ .

Note that  $f$  and  $g$  need not be conjugate. If  $f(x) = (x+1)^2 = x^2 + 2x + 1$  and  $g(x) = x^2$ , explain why  $N_f$  and  $N_g$  are conjugate, but  $f$  and  $g$  cannot be conjugate.

5. Show that if  $f(x) = (x+2)(x^2 + 1/5)$ , the Newton function  $N_f$  has the 2-cycle  $\{1/10, -3/5\}$ .
6. If  $g(x) = x^2 - 1$ , show that the basins of attraction of the fixed points 1 and  $-1$  of  $N_g$  are  $(0, \infty)$  and  $(-\infty, 0)$  respectively. (Hint: Look at a web plot for  $N_g$ . Show that if  $x_0 \in (1, \infty)$ , then  $1 < N_g(x_0) < x_0$ , so that  $N_g^n(x_0)$  is a decreasing sequence bounded below, and has to converge to the fixed point  $c = 1$  (since  $N_g$  is continuous on  $(0, \infty)$ ). The other cases are treated similarly.)
7. If  $g(x) = x^2 - \alpha$  for  $\alpha > 0$ , generalize the argument of the last question to show that the basins of attraction of the fixed points  $\sqrt{\alpha}$  and  $-\sqrt{\alpha}$  of  $N_g$ , are  $(0, \infty)$  and  $(-\infty, 0)$  respectively. When  $\alpha = 0$ , show that the only fixed point of  $N_g$  is 0 with basin of attraction  $\mathbb{R}$ .

8. Prove that if  $f$  and  $g$  are linearly conjugate via  $h(x) = ax + b$ , and the Newton's functions  $N_f$  and  $N_g$  are also linearly conjugate via  $h$ , then  $b = 0$ .

## CHAPTER 12

### Coppel's Theorem and a Proof of Sharkovsky's Theorem.

In this chapter, we return to Sharkovsky's Theorem and a related result due to Coppel. Coppel's Theorem gives additional insight into continuous maps on an interval  $[a, b]$ , when the map has no points of period two. We also provide a proof of Sharkovsky's Theorem that is due to Bau-Sen Du (see [38], and also [39], [25], [121]). Although these results are elementary in that they depend on little more than the Intermediate Value Theorem, and other properties of continuous functions and infinite sequences, they are still quite subtle and somewhat technical, and may be omitted from a first course in dynamical systems. The proof of Sharkovsky's Theorem is independent of Coppel's Theorem. This chapter is dependent on the material from Chapters 1, 2 and 3.

#### 12.1 Coppel's Theorem.

Our aim in this section is to prove a result due to Coppel (1955, [29], see also [19]), that was a precursor of Sharkovsky's Theorem, although, it seems that Sharkovsky was unaware of Coppel's work. Prior to establishing his major results, in the early 1960's, Sharkovsky reproved Coppel's result, and went on to prove his famous theorem on the periodic points of one-dimensional maps. It is easy to see that a continuous function on a compact interval  $[a, b]$  has the property: if  $\lim_{n \rightarrow \infty} f^n(x)$  exists for every  $x \in [a, b]$ , then  $f$  has no points of period two. That the converse is also true, is the content of Coppel's Theorem.

**Theorem 12.1.1** *Let  $f : [a, b] \rightarrow [a, b]$  be a continuous map. Then  $\lim_{n \rightarrow \infty} f^n(x)$  exists for every  $x \in [a, b]$ , if and only if  $f(x)$  has no points of period 2.*

Theorem 12.1.1 immediately gives one of the consequences of Sharkovsky's Theorem:

**Corollary 12.1.2** *A continuous map  $f : [a, b] \rightarrow [a, b]$  must have a 2-cycle if it has any periodic points that are not fixed.*

**Proof.** If  $f$  has an  $m$ -cycle,  $m > 1$ , then it has points in  $[a, b]$  for which the sequence  $(f^n(x))$  does not converge. It follows by Theorem 12.1.1 that  $f$  has a point of period 2.  $\square$

Before proving Coppel's Theorem, we prove some preliminary results.

**Lemma 12.1.3** *Let  $f : [a, b] \rightarrow [a, b]$  be a continuous map. If there exists  $c \in [a, b]$  with*

$$f^2(c) < c < f(c), \quad \text{or} \quad f(c) < c < f^2(c),$$

*then  $f(x)$  has a period-2 point.*

**Proof.** Suppose that  $c \in [a, b]$  with  $f^2(c) < c < f(c)$ . Set  $g(x) = f^2(x) - x$ . Then  $g(a) = f^2(a) - a \geq 0$ , and  $g(c) = f^2(c) - c < 0$ , so by the Intermediate Value Theorem, there exists  $p \in [a, c]$  with  $g(p) = 0$ , or  $f^2(p) = p$ . We may assume that

$$p = \max\{x \in [a, c] : f^2(x) = x\}.$$

If  $f(p) \neq p$ , we are done. Suppose that  $f(p) = p$ . Now  $(p, f(c)) \subset f(p, c)$ , so  $c \in f(p, c)$ . Hence there exists  $q \in (p, c)$  with  $f(q) = c$ .

As before,  $g(c) = f^2(c) - c < 0$ , and

$$g(q) = f^2(q) - q = f(c) - q > f(c) - c > 0,$$

so  $f^2(r) = r$  for some  $r$  in  $[q, c]$ . This is a contradiction, since it gives a point  $r \in [a, c]$  with  $f^2(r) = r$  and  $r > p$ . Consequently,  $p$  must be a period-2 point. This completes the proof in this case.

The case where  $f(c) < c < f^2(c)$  is similar.  $\square$

The next Lemma generalizes Lemma 12.1.3 in the following sense: it tells us that if there exists  $c \in [a, b]$  and  $n \geq 2$  with  $f^n(c) < c < f(c)$ , then  $f$  must have a 2-cycle. On the other hand, we can give examples of a continuous functions  $f$  on  $[a, b]$  for which there exists  $c \in [a, b]$ , with  $c < f^2(c) < f(c)$ , but having no points of period 2.

Roughly speaking, Lemma 12.1.3 shows that you cannot have oscillation about a repelling fixed point unless there is a 2-cycle: if  $p$  is a fixed point of  $f$  which is attracting, we may have a point  $c$  in the domain of  $f$  with  $f(c) < p < f^2(c) < c$ , and this gives rise to a 2-cycle.

**Lemma 12.1.4** *Let  $f : [a, b] \rightarrow [a, b]$  be a continuous map with no points of period 2.*

(i) If  $f(c) > c$ , then  $f^n(c) > c$  for all  $n \in \mathbb{Z}^+$ .

(ii) If  $f(c) < c$ , then  $f^n(c) < c$  for all  $n \in \mathbb{Z}^+$ .

**Proof.** We use a proof by induction: Let  $x \in [a, b]$  and  $m \in \mathbb{Z}^+$  be fixed.

As our induction hypothesis we take:

$$f(x) < x \Rightarrow f^n(x) < x, \text{ and } f(x) > x \Rightarrow f^n(x) > x, \quad n = 1, 2, \dots, m.$$

**Step 1.** We first show that a consequence of the induction hypothesis is:

$$f^{m+1}(x) = x \Rightarrow f(x) = x \quad \text{for } x \in [a, b],$$

(this is clearly true for  $m = 1$ ). Suppose that  $f^{m+1}(c) = c$  where  $f(c) > c$ . Then  $d = f^m(c) > c$  (by the induction hypothesis).

Suppose that

$$d = f^{m-1}(f(c)) \geq f(c).$$

Then  $f(f(c)) \geq f(c)$  (since if not,  $f(f(c)) < f(c) \Rightarrow f^{m-1}(f(c)) < f(c)$  by the induction hypothesis).

Then  $f^m(f(c)) \geq f(c)$  for the same reason, and this says  $f^{m+1}(c) = c \geq f(c)$ , a contradiction. Hence  $d = f^m(c) < f(c)$ , where  $f(d) = f^{m+1}(c) = c$ .

It follows that there exists  $q \in [c, d]$  where  $f(q) = d$ . We may assume that  $q$  is the nearest point to  $c$  at which this happens.

Then

$$c < q < d, \quad f(q) = d, \quad \text{and} \quad f(x) > d > q \quad \text{for } c \leq x < q.$$

Hence  $f^2(q) = f(d) = c < q$ . But  $f^2(c) > c$  because  $f(c) > c$ .

It follows (using  $g(x) = f^2(x) - x$  as before), that at some point  $x$  between  $c$  and  $q$ , we have  $f^2(x) = x$  and hence  $f(x) = x$ . But this is impossible because  $f(x) > q$  for  $c \leq x < q$ .

Similarly the assumption that  $f(c) < c$  leads to a contradiction. So we must have  $f(c) = c$

**Step 2.** We use the ideas of the proof of Lemma 12.1.3. The induction hypothesis clearly holds for  $m = 1$  (and also for  $m = 2$  by Lemma 12.1.3). Suppose that  $f(x) > x \Rightarrow f^n(x) > x$ ,  $n = 1, 2, \dots, m$ ,  $x \in [a, b]$ , and let  $c \in [a, b]$  with  $f(c) > c$  and  $f^{m+1}(c) < c$ . We will show that this leads to a contradiction.

As in Lemma 12.1.3 (using  $g(x) = f^{m+1}(x) - x$  in place of  $f(x) - x$ ), there is a point  $p \in (a, c)$  with  $f^{m+1}(p) = p$ . So by the induction hypothesis and Step 1,  $f(p) = p$ . As before, we may assume that there is no other such fixed point in  $(p, c)$ .

Since  $f(x) \neq x$  for  $x \in (p, c)$  and  $f(c) > c$ , it follows from the continuity that

$$(1) \quad f(x) > x \quad \text{for all } x \in (p, c).$$

Again the induction hypothesis gives

$$(2) \quad f^m(x) > x > p \quad \text{for all } x \in (p, c).$$

Choose  $q \in (p, c)$  very close to  $p$ , so that  $p < f^m(q) < c$ . Then from (1) and (2),

$$f^{m+1}(q) = f(f^m(q)) > f^m(q) > q,$$

contradicting, the definition of  $p$  since it gives rise to another fixed point in  $(q, c)$ . This contradiction gives us  $f^{m+1}(c) > c$ .

In a similar way, we show that  $f(c) < c$  and  $f^{m+1}(c) > c$  leads to a contradiction, and the Lemma follows.  $\square$

### 12.1.5 Proof of Coppel's Theorem.

If  $f$  has a 2-cycle  $\{x_1, x_2\}$ ,  $x_1 \neq x_2$ ,  $f(x_1) = x_2$ ,  $f(x_2) = x_1$ , then  $\lim_{n \rightarrow \infty} f^n(x_1)$  does not exist as  $f(x)$  oscillates between  $x_1$  and  $x_2$ .

To prove the converse, we use Lemma 12.1.4. Let  $x \in [a, b]$  and  $x_n = f^n(x)$ . If  $x_{m+1} = x_m$  for some  $m$ , then  $x_n = x_m$  for all  $n > m$ , and the sequence converges. Thus we can assume that  $x_{n+1} > x_n$  infinitely often, and  $x_{n+1} < x_n$  infinitely often.

Fix  $x \in [a, b]$  and set

$$A = \{x_n : f(x_n) > x_n\} \quad \text{and} \quad B = \{x_n : f(x_n) < x_n\}.$$

Then  $A$  and  $B$  are disjoint sets and their union is the orbit of  $x$ .

Suppose that

$$A = \{x_{n_1}, x_{n_2}, x_{n_3}, \dots, x_{n_p}, \dots\} \quad \text{where} \quad n_1 < n_2 < \dots < n_p < \dots,$$

so that  $x_{n_p} = f^{n_p}(x)$ , then by the definition of  $A$ ,  $f(x_{n_p}) > x_{n_p}$  for each  $p = 1, 2, \dots$

It follows that

$$x_{n_2} = f^{n_2}(x) = f^{n_2 - n_1}(f^{n_1}(x)) = f^{n_2 - n_1}(x_{n_1}) > x_{n_1}$$

by Lemma 12.1.4, since  $f(x_{n_1}) > x_{n_1}$ . In a similar way,

$$x_{n_{p+1}} = f^{n_{p+1}}(x) = f^{n_{p+1} - n_p}(f^{n_p}(x)) = f^{n_{p+1} - n_p}(x_{n_p}) > x_{n_p},$$

and gives rise to an increasing sequence which is contained in  $[a, b]$  (a subsequence of  $(x_n)$ ).

It follows that  $r = \lim_{p \rightarrow \infty} x_{n_p}$  exists, and in a similar manner, the terms of the  $B$  give a decreasing sequence, with limit  $q$  say,  $q \leq r$ .

For infinitely many  $n \in \mathbb{Z}^+$ , there exists  $x_n \in A$  with  $f(x_n) \in B$  (since we are assuming  $A$  and  $B$  are infinite sets whose union is all of the orbit of  $x$ ). Take a subsequence of  $(x_{n_p})$  (also denoted by  $(x_{n_p})$ ), with the property that  $f(x_{n_p}) \in B$  for all  $p$ . Then

$$\lim_{p \rightarrow \infty} x_{n_p} = r \quad \text{and} \quad \lim_{p \rightarrow \infty} f(x_{n_p}) = f(r) = q,$$

(by continuity). A similar argument shows that  $f(q) = r$ , so that  $\{q, r\}$  is a 2-cycle, contradicting our hypothesis. We must have  $q = r$ , i.e., the sequence converges for every  $x \in [a, b]$ . □

**Example 12.1.6** The logistic map  $L_\mu : [0, 1] \rightarrow [0, 1]$ ,  $L_\mu(x) = \mu x(1 - x)$ , is continuous and for  $0 < \mu \leq 3$ , has no points of period 2. Consequently,  $\lim_{n \rightarrow \infty} L_\mu^n(x)$  exists for all  $x \in [0, 1]$ . It follows that this sequence must converge to a fixed point. For  $0 < \mu \leq 1$  the only fixed point is  $x = 0$ , and so the basin of attraction of 0 is all of  $[0, 1]$ . For  $1 < \mu \leq 3$ ,  $x = 0$  is a repelling fixed point and  $x = 1 - 1/\mu$  is attracting. The basin of attraction of  $1 - 1/\mu$  is all of  $(0, 1)$ . In this case, the fixed point 0 and its eventual fixed point  $x = 1$  are the only points not in the basin of attraction of  $1 - 1/\mu$ .

If  $3 < \mu \leq 4$ , then  $L_\mu$  has period-2 points and Coppel's Theorem is not applicable. If we look at web plots near a fixed point  $p$ , we can see that they oscillate around  $p$ .

## 12.2 The Proof of Sharkovsky's Theorem.

Recall that the Sharkovsky ordering of  $\mathbb{Z}^+$  is:

$$3 \triangleright 5 \triangleright 7 \triangleright \cdots \triangleright 2 \cdot 3 \triangleright 2 \cdot 5 \triangleright \cdots \triangleright 2^2 \cdot 3 \triangleright 2^2 \cdot 5 \triangleright \cdots \triangleright 2^n \cdot 3 \triangleright 2^n \cdot 5 \triangleright \cdots \triangleright 2^n \triangleright 2^{n-1} \triangleright \cdots 2^3 \triangleright 2^2 \triangleright 2 \triangleright 1.$$

In the next two sections we will give a proof of the following form of Sharkovsky's Theorem:

**Theorem 12.2.1 (Sharkovsky's Theorem)** *Let  $f : I \rightarrow I$  represent a continuous map on a compact interval  $I \subseteq \mathbb{R}$ .*

(i) *If  $f$  has a point of period  $k$ , then it has points of period  $r$  for all  $r \in \mathbb{Z}^+$  with  $k \triangleright r$ .*

- (ii) For every  $k \in \mathbb{Z}^+$ ,  $f$  may be chosen so that it has period- $k$  points, but has no period- $n$  points for any  $n$  with  $n \triangleright k$ .
- (iii)  $f$  may be chosen so that it has period- $2^n$  points for every  $n \in \mathbb{Z}^+$ , and has no periodic points of any other period.

Currently there are many different proofs of Sharkovsky's Theorem (i). In [38], Bau-Sen Du surveys a number of these, and we closely follow the most straightforward proof given (see also [39], [25] and [121]). The proof we present for Sharkovsky's Theorem (ii) and (iii) will use a nice idea from [4], concerning the *truncated tent map*.

A number of authors have pointed out equivalences in Theorem 12.2.1(i) of Sharkovsky's Theorem (see [123]). These help to explain why the Sharkovsky ordering is defined the way it is, and also make it easier to prove the general result.

**Theorem 12.2.2** *Sharkovsky's Theorem part (i) is equivalent to the following three statements:*

- (a) if  $f$  has a period- $m$  point with  $m \geq 3$ , then  $f$  has a period-2 point;
- (b) if  $f$  has a period- $m$  point with  $m \geq 3$  odd, then  $f$  has a period- $(m + 2)$  point;
- (c) if  $f$  has a period- $m$  point with  $m \geq 3$  odd, then  $f$  has a period-6 point and a period- $2m$  point.

The detailed proof of this result is given in the remainder of this section, and completed in Section 12.3 (also see the exercises). However, the idea is as follows:

It is clear that Sharkovsky's Theorem (i) implies each of the statements (a), (b) and (c), so let us suppose that the latter three statements hold.

Statement (b) can be seen to imply that  $3 \triangleright 5 \triangleright 7 \triangleright 9 \dots$ , then (c) gives  $3 \triangleright 5 \triangleright 7 \triangleright 9 \dots \triangleright 2 \cdot 3$ .

We then use Lemma 12.2.3 (from [19]), below to show that

$$3 \triangleright 5 \triangleright 7 \triangleright 9 \triangleright \dots \triangleright 2 \cdot 3 \triangleright 2 \cdot 5 \triangleright 2 \cdot 7 \triangleright 2 \cdot 9 \triangleright \dots \triangleright 2^2 \cdot 3,$$

and then inductively complete the proof, except for the powers of 2, which follows from (a).

**Lemma 12.2.3** *Let  $k, m, n$  and  $s$  be integers.*

- (a) If  $f^m(x_0) = x_0$ , then the period of  $x_0$  under  $f$  divides  $m$  (recall that the period  $k$  is the least positive integer with  $f^k(x_0) = x_0$ ).
- (b) If  $x_0$  is a periodic point of  $f$  with period  $m$ , then it is a periodic point of  $f^n$  with period  $m/d$ , where  $d = (m, n)$  is the greatest common divisor of  $m$  and  $n$ .
- (c) If  $x_0$  is a periodic point of  $f^n$  having period  $k$ , then it is a periodic point of  $f$  with period  $kn/s$ , where  $s$  divides  $n$  and is relatively prime to  $k$ .

**Proof.** (a) See the exercises.

(b) Let  $p$  denote the period of  $x_0$  under  $f^n$ . Then  $m$  divides  $np$  since  $x_0 = (f^n)^p(x_0) = f^{np}(x_0)$ .

It follows that  $m/d$  divides  $(n/d) \cdot p$ . Since  $m/d$  and  $n/d$  are coprime,  $m/d$  divides  $p$ . On the other hand,  $(f^n)^{(m/d)}(x_0) = (f^m)^{(n/d)}(x_0) = x_0$ , so  $p$  divides  $m/d$ . This shows that  $p = m/d$ .

(c) Since  $x_0 = (f^n)^k(x_0) = f^{kn}(x_0)$ , the period of  $x_0$  under  $f$  is  $kn/s$  for some  $s \in \mathbb{Z}^+$ . From (b),  $\frac{(kn/s)}{((kn)/s, n)} = k$ . So  $n/s = ((n/s)k, n) = ((n/s)k, (n/s)s) = (n/s)(k, s)$ ,  $s$  divides  $n$ , and  $(s, k) = 1$ .

□

**Example 12.2.4** Suppose that  $f^2$  has a period-6 point. Then Lemma 12.2.3(c) implies that  $f$  has a period-12 point: simply take  $n = 2$ ,  $k = 6$ , then  $s = 1$  or 2, but since  $(s, 6) = 1$ , we have  $s = 1$  and so  $kn/s = 12$ .

The following lemma is a slightly more detailed version of Lemma 12.1.3.

**Lemma 12.2.5** *Let  $\mathcal{P}$  be an  $m$ -cycle,  $m \geq 3$ , for  $f$ . Then there exists  $v: f(v) \in \mathcal{P}$  with  $f^2(v) < v < f(v)$ .*

**Proof.** Suppose  $\mathcal{P} = \{x_1, \dots, x_m\}$  where  $x_1 < x_2 < \dots < x_m$ , then write

$$\mathcal{A} = \{x \in \mathcal{P}: f(x) > x\}, \quad a = \max(\mathcal{A}) \text{ and } b = \min\{x \in \mathcal{P}: a < x\}.$$

Neither  $f(b) > a$  or  $f(a) < b$  can happen, so we must have

$$f(b) \leq a < b \leq f(a).$$

Now choose  $v \in [a, b]$  so that  $f(v) = b$ . This can be done since  $[a, b] \subset f[a, b]$ . Since  $b$  is a period- $m$  point  $b \neq v$ . (In fact  $v$  will not, in general, belong to  $\mathcal{P}$ , but it is

possible that  $v = a$ ). Then

$$x_1 \leq f^2(v) < v < f(v).$$

□

#### 12.2.6 Proof of statements (a), (b) and (c) in Theorem 12.2.2.

(a) Using the notation of Lemma 12.2.5 with  $\mathcal{P} = \{x_1, \dots, x_m\}$  an  $m$ -cycle for  $f$ ,  $m \geq 3$ , there exists  $v \in [a, b]$  with

$$x_1 \leq f^2(v) < v < f(v) = b \in \mathcal{P}.$$

If  $g(x) = f(x) - x$ , then  $g(v) = f(v) - v = b - v > 0$ , and  $g(b) = f(b) - b = f^2(v) - f(v) < 0$ . Thus,  $g$  has a zero  $z$  in  $[v, b]$ , which must be a fixed point of  $f$ .

In a similar way, since  $f^2(x_1) > x_1$  and  $f^2(v) < v$ , there are points  $x \in [x_1, v]$  with  $f^2(x) = x$ . It follows that

$$y = \max\{x : x_1 \leq x \leq v, f^2(x) = x\},$$

exists.

Note that  $f(x) > z$ ,  $x \in [y, v]$ , for if there is an  $x$  with  $f(x) < z$ , then

$$[z, f(v)] \subseteq [f(x), f(v)] \subseteq f[x, v] \subseteq f[y, v],$$

and then we can find  $p \in [y, v]$  with  $f(p) = z$ . If  $h(x) = f^2(x) - x$ , then  $h(p) = f^2(p) - p = z - p > 0$ , and  $h(v) < 0$ , consequently there must exist  $w \in (y, v)$  with  $f^2(w) = w$ . This contradicts the maximality of  $y$ .

Therefore,  $f(x) > z > x$  on  $[y, v]$  and  $f^2(x) < x$  on  $(y, v]$ ,  $y$  must be a period-2 point, and (a) follows.

(b) Continuing with the notation in part (a), with  $m \geq 3$  odd, we have shown that

$$f(x) > z > x > f^2(x), \quad \text{for } x \in (y, v].$$

Since  $f^{m+2}(y) = f(y) > y$  and  $f^{m+2}(v) = f^2(v) = f(b) < v$ , the point

$$p_{m+2} = \min\{x : y \leq x \leq v, f^{m+2}(x) = x\}$$

exists.

Let  $k$  denote the period of  $p_{m+2}$  with respect to  $f$ . Then  $k$  divides  $m+2$ , so  $k$  is odd. Furthermore,  $k > 1$  because  $f$  has no fixed points in  $(y, v)$ . If  $k < m+2$ , let  $x_k = p_{m+2}$ , then  $x_k$  is a solution of the equation  $f^k(x) = x$  in  $(y, v)$ . Since  $f^{k+2}(y) = f(y) > y$  and  $f^{k+2}(x_k) = f^2(f^k(x_k)) = f^2(x_k) < x_k$ , the equation  $f^{k+2}(x) = x$  has a solution

$x_{k+2}$  in  $(y, x_k)$ . Continuing inductively for each  $n \geq 1$ , the equation  $f^{k+2n}(x) = x$  has a solution  $x_{k+2n}$  such that

$$y < \cdots < x_{k+4} < x_{k+2} < x_k < v.$$

Consequently, the equation  $f^{m+2}(x) = x$  has a solution  $x_{m+2}$  satisfying  $y < x_{m+2} < x_k = p_{m+2}$ . This contradicts the minimality of  $p_{m+2}$ , so  $p_{m+2}$  is a period- $(m+2)$  point of  $f$ , and (b) follows.

(c) Recall that  $f : I \rightarrow I$  is a continuous map of the compact interval  $I$ . Let  $I = [\alpha, \beta]$  and set

$$z_0 = \min\{x : v \leq x \leq z, f^2(x) = x\}.$$

Then  $f^2(x) < x$  and  $f(x) > z$ , when  $y < x < z_0$ . If  $f^2(x) < z_0$  whenever  $\alpha \leq x < z_0$ , then  $f^2([\alpha, z_0]) \subset [\alpha, z_0]$ , contradicting  $(f^2)^{(m+1)/2}(v) = b > z_0$ .

It follows that the point

$$d = \max\{x : \alpha \leq x \leq y, f^2(x) = z_0\}$$

exists, and  $f(x) > z \geq z_0 > f^2(x)$  for all  $x \in (d, y)$ . Therefore,  $f(x) > z \geq z_0 > f^2(x)$  whenever  $d < x < z_0$ .

Let  $s = \min\{f^2(x) : d \leq x \leq z_0\}$ . If  $s \geq d$ , then  $f^2([d, z_0]) \subset [d, z_0]$ , which again contradicts the fact that  $(f^2)^{(m+1)/2}(v) = b > z_0$ . Thus  $s < d$ .

We have shown that  $u = \min\{x : d \leq x \leq z_0, f^2(x) = d\}$  exists. Since  $f^2(d) = z_0 > d$  and  $f^2(u) = d < u$ , the point  $c_2 = \min\{x : d \leq x \leq u, f^2(x) = x\} (\leq y)$ , exists and  $d < f^2(x) < z_0$ , on  $(d, c_2)$ .

Let  $w \in (d, c_2]$  be such that  $f^2(w) = u$ . Then since  $f^4(d) = z_0 > d$  and  $f^4(w) = d < w$ , the point

$$c_4 = \min\{x : d \leq x \leq w, f^4(x) = x\} (< c_2),$$

exists, and  $d < f^4(x) < z_0$  on  $(d, c_4]$ .

Inductively, for each  $n \geq 1$ , let

$$c_{2n+2} = \min\{x : d \leq x \leq c_{2n}, f^{2n+2}(x) = x\}.$$

Then

$$d < \cdots < c_{2n+2} < c_{2n} < \cdots < c_4 < c_2 \leq y,$$

and  $d < (f^2)^k(x) < z_0$  on  $(d, c_{2k}]$ , for all  $1 \leq k \leq n$ .

Since  $f(x) > z_0$  on  $(d, z_0)$ , we have  $f^i(c_{2n}) < z_0 < f^j(c_{2n})$  for all even  $i : 2 \leq i \leq 2n$ , and all odd  $j : 1 \leq j \leq 2n - 1$ . So each  $c_{2n}$  is a period- $(2n)$  point of  $f$ . Therefore,  $f$  has points of all even periods, and (c) follows.  $\square$

### 12.3 The Completion of the Proof of Sharkovsky's Theorem.

We now complete the details of the proof outlined in Section 12.2. This amounts to using (a), (b) and (c) in that theorem, together with Lemma 12.2.3 to show that Sharkovsky's Theorem, part (i) follows.

#### 12.3.1 (a), (b), and (c) in Theorem 12.2.2 Imply Sharkovsky's Theorem.

It suffices to prove that (a), (b), and (c) in Theorem 12.2.2 imply Sharkovsky's Theorem part (i). We continue to follow [38].

**Proof of Theorem 12.2.2.** If  $f$  has period- $m$  points with  $m \geq 3$  odd, then by (b)  $f$  has period- $(m + 2)$  points, and by (c)  $f$  has period- $(2 \cdot 3)$  points. Thus, we get all odd terms appearing in the Sharkovsky ordering, and in particular

$$3 \triangleright 5 \triangleright 7 \triangleright 9 \triangleright \dots \triangleright 2 \cdot 3.$$

If  $f$  has period- $(2 \cdot m)$  points with  $m \geq 3$  odd, then by Lemma 12.2.3(b),  $f^2$  has period- $m$  points. By (b),  $f^2$  has period- $(m + 2)$  points, which implies (by Lemma 12.2.3(c)), that  $f$  has either period- $(m + 2)$  points or period- $(2 \cdot (m + 2))$  points. If  $f$  has period- $(m + 2)$  points, then according to (c),  $f$  has period- $(2 \cdot (m + 2))$  points. In either case,  $f$  has period- $(2 \cdot (m + 2))$  points. This gives:

$$2 \cdot 3 \triangleright 2 \cdot 5 \triangleright 2 \cdot 7 \triangleright 2 \cdot 9 \triangleright \dots.$$

On the other hand, since  $f^2$  has period- $m$  points ( $m$  odd), by (c),  $f^2$  has period- $(2 \cdot 3)$  points. By Lemma 12.2.3(c),  $f$  has period- $(2^2 \cdot 3)$  points. We now continue the proof inductively: if  $f$  has period- $(2^k \cdot m)$  points, with  $m \geq 3$  odd, and  $k \geq 2$ , by Lemma 12.2.3(b),  $f^{2^{k-1}}$  has period- $(2 \cdot m)$  points.

It follows from what we have established above,  $f^{2^{k-1}}$  has period- $(2 \cdot (m+2))$  points and period- $(2^2 \cdot 3)$  points.

Therefore, by Lemma 12.2.3(c),  $f$  has period- $(2^k \cdot (m + 2))$  points and period- $(2^{k+1} \cdot 3)$  points. This shows that

$$2^k \cdot m \triangleright 2^k \cdot (m + 2) \triangleright 2^{k+1} \cdot 3.$$

Consequently, if  $f$  has period- $(2^i \cdot m)$  points, with  $m \geq 3$  odd,  $i \geq 0$ , then by Lemma 12.2.3(b),  $f^{2^i}$  has period- $m$  points.

For each  $\ell \geq i$ , by Lemma 12.2.3(b)  $f^{2^\ell} = (f^{2^i})^{2^{\ell-i}}$  has period- $m$  points. By (c),  $f^{2^\ell}$  has period-6 points. So  $f^{2^{\ell+1}}$  has period-3 points and hence has period-2 points. This implies that  $f$  has period- $2^{\ell+2}$  points for all  $\ell \geq i$ , and  $2^i \cdot m \triangleright 2^{\ell+2}$ .

Finally, if  $f$  has period- $2^k$  points for some  $k \geq 2$ , then  $f^{2^{k-2}}$  has period-4 points. By (a),  $f^{2^{k-2}}$  has period-2 points. Therefore, by Lemma 12.2.3(c),  $f$  has period- $2^{k-1}$  points. This proves

$$2^k \triangleright 2^{k-1} \triangleright 2^{k-2} \triangleright \cdots \triangleright 2^2 \triangleright 2 \triangleright 1$$

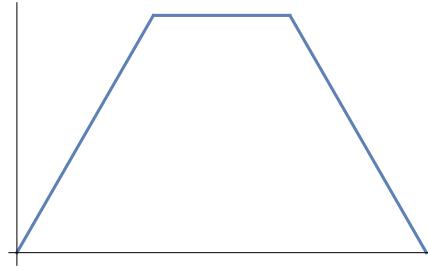
and completes the proof of Sharkovsky's Theorem part (i).  $\square$

**12.3.2 Proof of Sharkovsky's Theorem parts (ii) and (iii).** Theorem 12.2.1 parts (ii) and (iii) may be proved using a very elegant idea from [4]. See also [25]. The idea is to use a truncation of the standard tent map  $T : [0, 1] \rightarrow [0, 1]$ ,  $T(x) = 1 - |2x - 1|$ , which has periodic orbits of all periods. For each  $n \in \mathbb{N}$ ,  $T$  has finitely many periodic orbits of period  $n$ .

Let  $h \in [0, 1]$  and define the *truncated tent map*  $T_h : [0, 1] \rightarrow [0, 1]$  by

$$T_h(x) = \min\{h, T(x)\}.$$

To see how the method works recall that the equation  $T^n(x) = x$  has  $2^n$  solutions. For example, when  $n = 2$  there is a 2-cycle:  $\{2/5, 4/5\}$  together with the fixed points  $x = 0$  and  $x = 2/3$ . If we set  $h = 4/5$  (the maximum value of the unique 2-cycle), then  $T_h$  still has the same unique 2-cycle. However, we claim that  $T_h$  cannot have any cycles of period larger than 2 (see Exercises 3.2 # 14). In general, if  $\mathcal{O} = \{x_1, x_2, \dots, x_m\}$  is an  $m$ -cycle for  $T$  (in increasing order), then we set  $h_{\mathcal{O}} = x_m$ . Now write  $a = \min\{h_{\mathcal{O}}\}$ , where the minimum is taken over all possible  $m$ -cycles.



A truncated tent map.

**Claim 1.** The truncated tent map  $T_a$  has exactly one  $m$ -cycle, call it  $Q_m = \{z_1, z_2, \dots, z_m\}$  (in increasing order). Then  $a = z_m$ .

**Claim 2.**  $T(a) = z_1$ , and the interval  $[T(a), a]$  (the *convex hull* of the  $m$ -cycle), is invariant under  $T_a$ .

**Claim 3.** Every periodic orbit of  $T_a$  (except 0), eventually enters the interval  $[T(a), a]$  and stays there, so it has to be contained in  $[T(a), a]$ .

**Claim 4.** If  $k \triangleright m$  and  $T_a$  has a  $k$ -cycle, say  $R$ , then this orbit is contained in  $[T(a), a]$ . However, it follows from Sharkovsky's Theorem that since  $k \triangleright m$ ,  $T_a$  has an  $m$ -cycle whose convex hull is contained in the convex hull of  $R$ , so must be different to  $Q$ . This gives a contradiction, so  $R$  cannot exist, and completes the proof of Theorem 12.2.1 part (ii).

To show that there is a map having  $2^n$ -cycles for each  $n$ , but no other cycles, we look at the maps  $T_{a_{2^n}}$ , where  $a_{2^n}$  is chosen as above so that  $T_{a_{2^n}}$  has a unique  $2^n$ -cycle, but no  $k$ -cycles, with  $k \triangleright 2^n$ .

**Claim 5.** Set  $a = \lim_{n \rightarrow \infty} a_{2^n}$ , then  $T_a$  is the map with the required property.  $T_a$  has no other cycles besides  $2^n$ -cycles for  $n \in \mathbb{N}$ .

□

### Exercises 12.3

1. Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function having no 2-cycle. Show that if  $c \in [a, b]$  with  $f(c) > c$ , then  $f^2(c) > c$ . (Hint: Use the contrapositive of Lemma 12.1.3).
2. We know that if  $f : [0, 1] \rightarrow [0, 1]$  is continuous, then  $f$  has at least one fixed point. Show that if  $f$  is also onto, then  $f^2$  has at least two fixed points. Give an example of such an onto map having two fixed points but no period-2 points.
3. Use Coppel's Theorem to prove that if  $f : [a, b] \rightarrow [a, b]$  is a continuous function having a unique fixed point  $c$ , and no points of period 2, then  $\lim_{n \rightarrow \infty} f^n(x) = c$  for all  $x \in [a, b]$ .
4. If  $f^m(x_0) = x_0$ , show that the period of  $x_0$  under  $f$  divides  $m$  (recall that the period  $k$  is the least positive integer with  $f^k(x_0) = x_0$ ).

5. Use the argument of Lemma 12.1.3 to show that if a continuous function  $f$  has a 3-cycle  $\{a, b, c\}$ , with  $a < b < c$ ,  $f(a) = b$ ,  $f(b) = c$ , and  $f(c) = a$ , then  $f$  has a 2-cycle.
6. Give an example of a continuous real function on an interval  $[a, b]$  for which there exists  $c \in [a, b]$  with  $c < f^2(c) < f(c)$ , but no 2-cycle.
7. Use Lemma 12.2.3 to show the following:
- If  $f^2$  has period- $m$  points, then  $f$  has either period- $m$  points or period- $2m$  points.
  - If  $f$  has period- $m$  points,  $m$  odd, then  $f^{2^k}$  has period- $m$  points for  $k \geq 1$ .
  - If  $f$  has period- $(2^k \cdot m)$  points with  $m$  odd,  $k \geq 1$ , then  $f^{2^k}$  has period- $m$  points.
8. (a) Suppose that  $f$  is a  $C^1$ -function with a fixed point  $p$  for which  $f'(p) < -1$ . Show that  $f$  has a 2-cycle. (Hint: Show that Lemma 12.1.3 is applicable).
- (b) Use (a) to show that the logistic map  $L_\mu(x) = \mu x(1-x)$  has a 2-cycle for  $\mu > 3$ .
- (c) Do the same for  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^3 - \lambda x$ , by looking at the fixed point  $x = 0$ , and showing that a 2-cycle is created when  $\lambda = 1$ . Show that another 2-cycle is created when  $\lambda = 2$ , and find all of the fixed points. (Hint: It is helpful to investigate this using a computer algebra system).
9. In (b) and (c) of the last exercise, we have examples of what are called *period doubling bifurcations*. Roughly speaking, a family of maps  $f_\mu$  which vary according to a parameter  $\mu$ , has a period doubling bifurcation at  $\mu = \mu^*$  if
- $f_{\mu^*}(p) = p$ , so that  $f_{\mu^*}$  has a fixed point at  $x = p$ ,
  - $f'_{\mu^*}(p) = -1$ , so the fixed point is non-hyperbolic,
  - the graph of  $f_\mu^2$  crosses that line  $y = x$  when  $\mu < \mu^*$  (so that there is a single fixed point close to  $p$  for  $f_\mu$ ), and is tangent to  $y = x$  when  $\mu^* = \mu$ , and “snakes around” the line  $y = x$ , when  $\mu > \mu^*$  (so  $f_\mu^2$  now has three fixed points).

- (a) Let  $\lambda < 0$ . If  $E_\lambda(x) = \lambda e^x$ , show that  $x = -1$  is a fixed point when  $\lambda = -e$ , and that we have a period doubling bifurcation as  $\lambda$  decreases through  $-e$ .
- (b) If  $f_\mu(x) = \mu x e^{-x}$ , find  $\mu^*$  and  $p$  where we have a period doubling bifurcation.

10\*. Use the following steps to give alternate proofs of Lemmas 12.1.3 and 12.1.4 for a continuous function  $f$  on a compact interval  $I = [\alpha, \beta]$  (see [40]):

- (a) Prove that if there exists  $a, b \in [\alpha, \beta]$  with

$$f(b) < a < b \leq f(a),$$

then  $f$  has a periodic point  $z < b$  and a period-2 point  $y < z$ . This can be done as follows:

- (i) Find the fixed point  $z < b$  in the usual way, and show that there exists  $v \in [a, z]$  with  $f(v) = b$ .
- (ii) Set  $u = \alpha$  if  $f(x) > z$  for all  $x \in [\alpha, v]$ , otherwise set  $u = \max\{x \in [\alpha, v] : f(x) = z\}$ .
- (iii) Show that there exists  $y \in [u, v]$  with  $f^2(y) = y$ , and  $f(y) \neq y$ .

- (b) Suppose there exists  $c \in I$  and  $n \geq 2$  with

$$f^n(c) < c < f(c).$$

Set  $\mathcal{A} = \{f^k(c) : 0 \leq k \leq n-1\}$ ,  $a = \max\{x \in \mathcal{A} : c \leq x \text{ and } f(x) > x\}$ , and choose  $b \in \{x \in \mathcal{A} : a < x \leq f(a)\}$  with  $f(b) < a$ . Show that  $b$  exists and hence

$$f(b) < a < b \leq f(a).$$

- (c) Deduce that  $f$  has a fixed point  $z \in I$ ,  $z < b$ , and has a period-2 point  $y \in I$ ,  $y < z$ , and complete the proof of the two lemmas.

Using these ideas and arguments similar to those from the proof of Theorem 12.2.2, (b) and (c), the following can be shown: (i) if  $n \geq 3$  is odd and  $c \in [\alpha, \beta]$  with  $f^n(c) < c < f(c)$ , then  $f$  has a period- $n$  point, and (ii) if  $f$  is transitive, then  $f$  has points of all even periods (see [40]).

11. Prove Claims 1-5, concerning the truncated tent map in 12.3.2.

## CHAPTER 13

# Real Linear Transformations, the Hénon Map, and Hyperbolic Toral Automorphisms.

### 13.1 Linear Transformations.

In Chapter 1 we saw how continuous linear transformations  $f : \mathbb{R} \rightarrow \mathbb{R}$  behave under iteration. Such maps are of the form  $f(x) = ax$  for some  $a \in \mathbb{R}$ , so that when  $a \neq 1$ ,  $x = 0$  is the only fixed point. This fixed point is attracting if  $|a| < 1$ , repelling if  $|a| > 1$ , and *neutral* (stable, but neither attracting nor repelling), when  $|a| = 1$  (of course when  $a = 1$ , every point is fixed). Consequently, the dynamics of these maps is trivial in each case. The situation is much more complicated for maps on higher dimensional spaces, and we see phenomena that does not arise in the 1-dimensional case. We begin this chapter with a study of real linear transformations  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  for  $n > 1$ , and then look at various types of maps arising from these transformations. Just as the linear transformation  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = 2x$  induces a map  $T(x) = 2x \pmod{1}$ , on the interval  $[0, 1)$ , linear transformations on  $\mathbb{R}^2$  induce maps on  $[0, 1) \times [0, 1)$ . These considerations gives rise to hyperbolic toral automorphisms, that we study in Section 13.4 and 13.5. In Section 13.2, we give a brief introduction to non-linear maps on  $\mathbb{R}^2$ , with particular emphasis on the Hénon map. This chapter assumes familiarity with the material from Chapters 1 and 2, together with a basic knowledge of linear algebra and the properties of matrices. The results from Chapter 4 are also applicable, since  $\mathbb{R}^n$  is a metric space with its usual Euclidean distance. The notion of chaos from Chapter 6, and conjugacy from Chapter 7, are also briefly mentioned. None of the other chapters of this text are dependent on the material in this chapter.

**Definition 13.1.1** A (real) *linear transformation* is a function  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  having the property:

$$F(ax + by) = aF(x) + bF(y), \quad \text{for all } x, y \in \mathbb{R}^n \quad \text{and } a, b \in \mathbb{R}.$$

In this chapter it is useful to think of  $\mathbb{R}^n$  as the vector space of all  $n$ -by-1 *column vectors* (or  $n$ -by-1 *matrices*). Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a real linear transformation with  $n > 1$ . It is well known that such a map can be represented by matrix multiplication: there is a matrix  $A \in M_n(\mathbb{R})$  (the vector space of all  $n$ -by- $n$  matrices having real entries), for which  $F(x) = A \cdot x$ . We examine the case where  $n = 2$  in more detail. Here we have

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\} = \{(x_1, x_2)^t : x_1, x_2 \in \mathbb{R}\},$$

where  $A^t$  represents the transpose of the matrix  $A$ . Note that  $\mathbb{R}^2$  is also a metric space with the usual distance between  $(x_1, x_2)^t$  and  $(y_1, y_2)^t$  given by  $\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ . It readily follows that every linear function on  $\mathbb{R}^2$  is continuous everywhere.

Suppose that  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . Write  $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ , then

$$F(x) = A \cdot x = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix}.$$

The *eigenvalues* of the linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  are those  $\lambda \in \mathbb{C}$  for which there exists  $v \in \mathbb{R}^n$ ,  $v \neq 0$ , with  $F(v) = \lambda \cdot v$ . These can be found by solving the *characteristic equation*  $\det(A - \lambda I) = 0$ , where  $\det(A)$  is the *determinant* of  $A$  ( $\det(A) = ad - bc$  when  $n = 2$ ), and  $I$  is the  $n$ -by- $n$  identity matrix. The eigenvalues of  $A$  may be real or complex, and we shall see that they determine the dynamics of the corresponding dynamical system  $F$ . The vectors  $v$  with  $F(v) = \lambda v$ , are the *eigenvectors* of  $F$  corresponding to  $\lambda$ , and the set

$$E_\lambda = \{v \in \mathbb{R}^n : F(v) = \lambda v\},$$

is a vector subspace of  $\mathbb{R}^n$ , called the *eigenspace* of  $F$  corresponding to the eigenvalue  $\lambda$ .

Again  $x = 0 = (0, 0)^t$  is always a fixed point of  $F$ , and the dynamics of  $F$  will be seen to be fairly straightforward.

Suppose that  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a linear transformation having two distinct real eigenvalues  $\lambda_1, \lambda_2 \in \mathbb{R}$ . If  $F(x) = Ax$ , then the 2-by-2 matrix  $A$  is diagonalizable, i.e., there is an invertible matrix  $P \in M_2(\mathbb{R})$  such that  $A = P\Lambda P^{-1}$ , where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Notice that if  $A = P\Lambda P^{-1}$ , then  $AP = P\Lambda$ . Write  $P$  in terms of its columns as  $P = [v_1 \ v_2]$ , then

$$AP = A[v_1 \ v_2] = [Av_1 \ Av_2],$$

and

$$P\Lambda = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = [v_1 \ v_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = [\lambda_1 v_1 \ \lambda_2 v_2],$$

so that  $Av_1 = \lambda_1 v_1$  and  $Av_2 = \lambda_2 v_2$ . In other words, to find the matrix  $P$  we find (any linearly independent), eigenvectors of  $A$ , and use these as the columns of  $P$ . This argument is valid whatever the values of the eigenvalues  $\lambda_1$  and  $\lambda_2$  (real or complex, or with  $\lambda_1 = \lambda_2$ ), as long as the eigenvectors are linearly independent (as this ensures that  $P$  is invertible). Notice that we now have

$$A^n = P \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix} P^{-1}.$$

This will help us analyze the long term behavior of the map  $F(x) = Ax$ .

The diagonalization of  $A$  tells us that  $F$  is (linearly), conjugate to the map  $G(x) = \Lambda x$  (via a homeomorphism  $H : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $H(x) = P \cdot x$ ), so the dynamics of  $F$  and  $G$  are essentially the same.

Suppose the eigenvectors  $v_1$  and  $v_2$  of  $F$  corresponding to  $\lambda_1$  and  $\lambda_2$  respectively, are linearly independent. Then we can write any  $v \in \mathbb{R}^2$  as

$$v = c_1 v_1 + c_2 v_2, \quad \text{for some } c_1, c_2 \in \mathbb{R},$$

and

$$F(v) = c_1 F(v_1) + c_2 F(v_2) = c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2, \quad F^n(v) = c_1 \lambda_1^n v_1 + c_2 \lambda_2^n v_2.$$

If  $c_2 = 0$ , then  $F^n(v) = c_1 \lambda_1^n v_1$ . If  $|\lambda_1| < 1$ , all points in the eigenspace  $E_{\lambda_1}$  iterate toward 0. If  $|\lambda_1| > 1$ , then the iterates of the eigenspace are unbounded.

When neither eigenvalue has absolute value equal to 1 (the *hyperbolic case*), there are three cases to consider: (i)  $|\lambda_1|, |\lambda_2| < 1$ , (ii)  $|\lambda_1|, |\lambda_2| > 1$ , (iii)  $|\lambda_1| < 1, |\lambda_2| > 1$ . In case (i), all points iterate towards the origin (0 is an attracting fixed point of  $F$ ). For example, a circle centered on the origin will be contracted in the directions of the eigenspaces by a factor corresponding to the absolute values of the eigenvalues, resulting in an ellipse. In case (ii), 0 is a repelling fixed point, and we have an expansion of the circle in the direction of the eigenspaces, giving rise to a large ellipse.

In case (iii) we have attraction in the direction of the eigenspace  $E_{\lambda_1}$ , and repelling in the direction of  $E_{\lambda_2}$  (both are lines through the origin). The circle is now expanded in one direction and contracted in the other, giving rise to a long thin ellipse. Points

not in the eigenspaces move in hyperbolic trajectories, first toward the origin, and then away. Other types of behavior occur when the eigenvalues are not real, or have absolute value one. We have the following summary in the two dimensional setting. The general finite dimensional case is similar.

**Theorem 13.1.2 Stability of Linear Maps** *Let  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a linear transformation with  $F(x) = A \cdot x$  for some  $A \in M_2(\mathbb{R})$ , having distinct non-zero real eigenvalues  $\lambda_1$  and  $\lambda_2$ .*

- (a) *If  $|\lambda_1| < 1$  and  $|\lambda_2| < 1$ , then  $F^n(x) \rightarrow 0$  as  $n \rightarrow \infty$ . We say that 0 is an attractor of  $F$ , and the fixed point  $x = 0$  is asymptotically stable.*
- (b) *If  $|\lambda_1| > 1$  and  $|\lambda_2| > 1$ , then  $x = 0$  is a repelling fixed point.*
- (c) *If  $|\lambda_1| > 1 > |\lambda_2|$  or  $|\lambda_1| < 1 < |\lambda_2|$ , then  $x = 0$  is called a saddle. In this case, the fixed point  $x = 0$  is unstable because there are trajectories arbitrarily close to the origin that move away from the origin.*

**Definition 13.1.3** An invertible linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is said to be *hyperbolic*, if the corresponding matrix  $A$  has no eigenvalues of absolute value equal to one. Otherwise it is *non-hyperbolic*.

**Examples 13.1.4** 1. Let  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ ,  $F(x) = Ax$ , where

$$A = \begin{bmatrix} 19/2 & 45/2 \\ -3 & -7 \end{bmatrix}.$$

$\det(A - \lambda I) = \lambda^2 - 5\lambda/2 + 1 = (\lambda - 1/2)(\lambda - 2) = 0$  when  $\lambda_1 = 2$  and  $\lambda_2 = 1/2$ . We find the eigenvectors by solving  $Ax - \lambda_i x = 0$  for  $i = 1, 2$ . This gives  $v_1 = (3, -1)^t$  (the eigenspace is the line  $x + 3y = 0$ ), and  $v_2 = (-5, 2)^t$  (the line  $2x + 5y = 0$ ). The matrix  $P$  is given by  $P = [v_1 \ v_2] = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}$ , and  $\Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}$ , giving  $AP = P\Lambda$ .

Since  $|\lambda_1| > 1$ , we have expansion along  $x+3y = 0$ , and  $|\lambda_2| < 1$  gives a contraction along  $2x + 5y = 0$ , giving case (c), so  $x = 0$  is a saddle.

2. We ask what happens if the linear map arises from a matrix  $A$  having complex eigenvalues? For example, suppose  $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ , then the determinant of  $A - \lambda I$  is  $\lambda^2 + 1$ , and the eigenvalues are  $\pm i$ . Now

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix},$$

and we see that  $A$  acts as a rotation through  $90^\circ$ . In a similar way,  $\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$  gives rise to a rotation through  $45^\circ$ , with eigenvalues  $1/\sqrt{2} \pm i/\sqrt{2}$ . Notice that the eigenvalues appear in complex conjugate pairs. This is because the characteristic polynomial  $\det(A - \lambda I)$  has only real coefficients. Stability follows exactly as in 13.1.2, except that there may be some rotation involved, in addition to the expansion or contraction. Since conjugate eigenvalues have the same absolute value, case (c) does not happen. If  $|\lambda_1| = |\lambda_2| = 1$ , then  $F$  is purely a rotation.

**Proposition 13.1.5** *Let  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be an invertible linear transformation with  $F(x) = Ax$  for  $A \in M_2(\mathbb{R})$  having non-real eigenvalues  $\lambda = a + ib$  and  $\bar{\lambda} = a - ib$ , with  $b \neq 0$ . Then  $A$  can be written in the form*

$$A = [\operatorname{Re}(\alpha) \operatorname{Im}(\alpha)] \begin{bmatrix} a & b \\ -b & a \end{bmatrix} [\operatorname{Re}(\alpha) \operatorname{Im}(\alpha)]^{-1},$$

where  $\operatorname{Re}(\alpha)$  and  $\operatorname{Im}(\alpha)$  are the real and imaginary parts of an eigenvector  $\alpha$  of  $A$  corresponding to  $a + ib$ .  $F^n(x) \rightarrow 0$  as  $n \rightarrow \infty$  for  $|\lambda| = \sqrt{a^2 + b^2} < 1$ , and  $F^n(x)$  is unbounded for  $|\lambda| > 1$ .

**Proof.** The diagonalization of the matrix  $A$  when  $A$  has two distinct real eigenvalues remains valid even when the eigenvalues are non-real. Suppose the eigenvalues of  $A$  are  $a \pm ib$ , occurring in conjugate pairs. Then we can diagonalize  $A$  as

$$A = P \begin{bmatrix} a + ib & 0 \\ 0 & a - ib \end{bmatrix} P^{-1},$$

where  $P = [v_1 \ v_2]$  and  $v_1, v_2$  are the eigenvectors corresponding to the eigenvalues. Now if  $A\alpha = \lambda\alpha$ , then  $A\bar{\alpha} = \bar{\lambda}\bar{\alpha}$ , since  $A$  is real. It follows that we may assume  $P = [\alpha \ \bar{\alpha}]$ , for some eigenvector  $\alpha$  of  $A$ . Set  $Q = P \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ i/\sqrt{2} & -i/\sqrt{2} \end{bmatrix}^{-1}$ . Then a calculation shows that  $Q = [\operatorname{Re}(\alpha) \operatorname{Im}(\alpha)]$  and

$$A = \frac{1}{2}Q \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} a + ib & 0 \\ 0 & a - ib \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^{-1} Q^{-1} = Q \begin{bmatrix} a & b \\ -b & a \end{bmatrix} Q^{-1}.$$

This completes the first part of the proof.

Since  $A^n = P \begin{bmatrix} \lambda^n & 0 \\ 0 & \bar{\lambda}^n \end{bmatrix} P^{-1}$ , the last part now follows. □

**Remark 13.1.6** 1. In the case where  $A \in M_2(\mathbb{R})$  has an eigenvalue  $\lambda$  of multiplicity two (so the eigenspace is 1-dimensional), the matrix  $A$  is not diagonalizable; but there is an invertible matrix  $P \in M_2(\mathbb{R})$  with

$$A = P\Lambda P^{-1}, \quad \Lambda = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}.$$

We can then use  $A^n = P \begin{bmatrix} \lambda^n & n\lambda^{n-1} \\ 0 & \lambda^n \end{bmatrix} P^{-1}$ , to analyze the dynamics of  $F(x) = Ax$  (see the exercises).

2. For  $A \in M_2(\mathbb{R})$ , it is clear that we can write down a closed formula for  $A^n$  in each of the three possible cases. Using the conjugacy between the induced linear map  $F$  and the map  $G(x) = \Lambda x$ , the behavior of  $F$  under iteration is now readily determined in the hyperbolic case.

**Examples 13.1.7** 1. If  $A = \begin{bmatrix} 2 & -3 \\ 4 & -4 \end{bmatrix}$ , then  $\det(A - \lambda I) = \lambda^2 + 2\lambda + 4$ , and the eigenvalues are  $-1 \pm i\sqrt{3}$ . Both eigenvalues have absolute value equal to 2, so the fixed point  $x = 0$  is repelling.

2. Let  $A = \begin{bmatrix} -3/2 & 4 \\ -1 & 5/2 \end{bmatrix}$ . Then  $\det(A - \lambda I) = \lambda^2 - \lambda + 1/4 = (\lambda - 1/2)^2$ , so there is a single eigenvalue  $\lambda = 1/2$ . Solving  $(A - I/2)v = 0$  gives a 1-dimensional eigenspace spanned by  $v_1 = (1, 1)^t$ . Using the method of Exercises 13.1 # 5, we see that  $A = P\Lambda P^{-1}$ , where  $P = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ , and  $\Lambda = \begin{bmatrix} 1/2 & 1 \\ 0 & 1/2 \end{bmatrix}$ . In particular, the corresponding linear map is hyperbolic with  $x = 0$  an attracting fixed point.

**Remark 13.1.8** In the case where the matrix  $A$  has two distinct eigenvalues satisfying  $|\lambda_1| > 1 > |\lambda_2|$ , the eigenspace  $E_{\lambda_1}$  is the expanding direction and  $E_{\lambda_2}$  is the contracting direction. These are referred to as the *unstable manifold*  $W^u$ , and the *stable manifold*  $W^s$ . In the case of a suitably differentiable non-linear map of  $\mathbb{R}^2$ , having a hyperbolic fixed point of a saddle type, there is an open ball centered on the fixed point, and continuous curves defined in the open ball,  $W^u$  and  $W^s$  intersecting at the fixed point, giving the expanding and contracting directions respectively. This is the *Stable Manifold Theorem* [32].

## Exercises 13.1

1. Determine the dynamics of the dynamical systems whose matrices are:

$$(i) \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \quad (ii) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad (iii) \begin{bmatrix} .5 & 0 \\ 0 & .3 \end{bmatrix}, \quad (iv) \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix}.$$

2. Diagonalize each of the following matrices  $A$ , and find  $A^n$ :

$$(i) \begin{bmatrix} 1 & 1/2 \\ 3/4 & 5/4 \end{bmatrix}, \quad (ii) \begin{bmatrix} -5/3 & 3 \\ -2 & 10/3 \end{bmatrix}, \quad (iii) \begin{bmatrix} -5 & -7 \\ 7/2 & 11/2 \end{bmatrix}.$$

Deduce the dynamics of the induced linear transformation.

3. Write each of the following matrices in the form  $Q\Lambda Q^{-1}$  (where  $\Lambda = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ , for some  $a, b \in \mathbb{R}$ ), and deduce the dynamics of the corresponding linear map.

$$(i) A = \begin{bmatrix} -7 & 15 \\ -6 & 11 \end{bmatrix}, \quad (ii) B = \begin{bmatrix} 1/6 & 2/3 \\ -1/3 & 5/6 \end{bmatrix}.$$

4. (a) Diagonalize the matrix  $I_{a,b} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ , for  $a, b \in \mathbb{C}$ .

(b) Show that any two matrices of the form  $I_{a,b}$  must commute.

(c) Note that if  $a + ib \neq 0$  with  $a, b \in \mathbb{R}$ , then  $\det(I_{a,b}) \neq 0$ . Give an example to show that this need not be true when  $a, b \in \mathbb{C}$ . (Such matrices are examples of *normal matrices*: they have the property  $AA^* = A^*A$ , where  $A^*$  is the *conjugate transpose* of  $A$ ).

5. (a) If  $A \in M_2(\mathbb{R})$  has a single eigenvalue of multiplicity two (so the eigenspace is 1-dimensional), show that there is  $P \in M_2(\mathbb{R})$ , invertible, with

$$A = P\Lambda P^{-1}, \quad \text{where } \Lambda = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}.$$

(Hint: Let  $P = [v_1 \ v_2]$ , where  $v_1 \neq 0$  is an eigenvalue of  $A$ , and  $(A - \lambda I)v_2 = v_1$ ).

$$(b) \text{ Deduce that } A^n = P \begin{bmatrix} \lambda^n & n\lambda^{n-1} \\ 0 & \lambda^n \end{bmatrix} P^{-1}.$$

(c) If  $F(x) = Ax$ , where  $A$  is the matrix in (a), deduce the long term behavior of  $F$ .

6. Let  $A = \begin{bmatrix} 5 & -4 \\ 9 & -7 \end{bmatrix}$ . Show  $A$  has a single eigenvalue of multiplicity one, and find the matrices  $P$  and  $\Lambda$  defined in the previous exercise. (This gives the *Jordan form* of the matrix  $A$ ). Deduce the long term behavior of the induced map.

7. (a) If  $A \in M_2(\mathbb{R})$  is a matrix with characteristic equation

$$\lambda^2 - \tau\lambda + \Delta = 0,$$

show that  $\Delta =$  the determinant of  $A$  and  $\tau =$  trace of  $A$  (sum of the diagonal elements).

(b) Show that if the eigenvalues of  $A$  are  $\lambda_1$  and  $\lambda_2$ , then  $|\lambda_1| < 1$  and  $|\lambda_2| < 1$  if and only if  $(\tau, \Delta)$  falls inside the triangle

$$\{(\tau, \Delta) : \Delta < 1, \Delta > -\tau - 1, \Delta > \tau - 1\}.$$

(c) Show that we have  $\lambda = 1$ , for an eigenvalue on the edge of the triangle where  $\Delta = \tau - 1$ , and  $\lambda = -1$  where  $\Delta = -\tau - 1$ .

(d) Show that the eigenvalues are non-real in the subset of the triangle  $\{(\tau, \Delta) : \Delta < 1, 4\Delta < \tau^2\}$ . What do points on the boundary where  $\Delta = 1$  correspond to?

8. (a) If  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a linear map, show that  $F$  is continuous. (Hint: Consider why is it sufficient to show that  $F$  is continuous at zero). Deduce that if  $x_n \rightarrow 0$  as  $n \rightarrow \infty$  (in  $\mathbb{R}^2$ ), then  $F(x_n) \rightarrow 0$  as  $n \rightarrow \infty$ .

(b) Show that if  $F$  is diagonalizable with eigenvalues  $|\lambda_1| < 1, |\lambda_2| < 1$ , then  $F^n x \rightarrow 0$  as  $n \rightarrow \infty$ , for all  $(x_1, x_2)^t \in \mathbb{R}^2$ .

(c) If  $H(x) = P \cdot x$ , where  $P \in M_n(\mathbb{R}^n)$  is invertible, show that  $H$  is a homeomorphism.

### 13.2 The Hénon Map.

The Hénon map was first defined by Maurice Hénon in 1976 [68], as a discrete approximation to the 3-dimensional flow of Lorenz. It can also be considered as a 2-dimensional analog of the logistic family. We define a version of the Hénon map  $H : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  by

$$H \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f(x, y) \\ g(x, y) \end{pmatrix},$$

where

$$f(x, y) = 1 - ax^2 + y, \quad g(x, y) = bx, \quad \text{for } a \neq 0, \quad 0 < b < 1.$$

We focus on this map as it is one of the simplest possible 2-dimensional non-linear dynamical systems. However, it remains under intense study as it is still not fully understood: for certain values of the parameters  $a$  and  $b$ , iterates give rise to a geometric object known as a *strange attractor*, the set on which the chaotic behavior appears to take place.

We start by looking for the fixed points of  $H$  (see [89]). These occur when

$$x = 1 - ax^2 + y, \quad \text{and} \quad y = bx,$$

so that  $ax^2 + x(1 - b) - 1 = 0$ , and solving:

$$x = \frac{b - 1 \pm \sqrt{(1 - b)^2 + 4a}}{2a}.$$

In order for these roots (denoted by  $x^+$  and  $x^-$  with corresponding  $y$ -values  $y^+$  and  $y^-$ ), to be real, we require  $(1 - b)^2 + 4a \geq 0$ . There is a single fixed point when  $(1 - b)^2 + 4a = 0$ , and two fixed points when  $(1 - b)^2 + 4a > 0$ .

To examine the stability of these fixed points, we introduce the *Jacobian* of the map  $H$ .

**Definition 13.2.1** Let  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a  $C^1$  map (this means that the components of  $F$  have continuous first partial derivatives). The *Jacobian* of  $F$  is the matrix

$$JF \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix}, \quad \text{when} \quad F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f(x, y) \\ g(x, y) \end{pmatrix}.$$

The stability of the fixed points  $(x^+, y^+)$  and  $(x^-, y^-)$  (as is suggested by the linear case), is determined by the eigenvalues of the Jacobian, evaluated at the fixed points. The analog of the 1-dimensional result for the stability of fixed points, is that we require the eigenvalues to have absolute value less than one. If one of the eigenvalues has absolute value greater than one, the fixed point is unstable. If one

eigenvalue has absolute value greater than one, and other is less than one, we have a *saddle*.

For the Hénon map,  $JH \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} -2ax & 1 \\ b & 0 \end{bmatrix}$ . The eigenvalues are the solutions of  $\lambda^2 + 2ax\lambda - b = 0$ , or

$$\lambda = \frac{-2ax \pm \sqrt{4a^2x^2 + 4b}}{2} = -ax \pm \sqrt{a^2x^2 + b}.$$

Denote the eigenvalues by  $\lambda^+$  and  $\lambda^-$  when  $x = x^+$ . These are real since  $b > 0$ . In order for stability, we need  $|\lambda^\pm| < 1$ . Now  $|\lambda^+| < 1$  if and only if  $-1 < \lambda^+ < 1$ . First we consider the inequality  $\lambda^+ < 1$ . This gives

$$\sqrt{(ax^+)^2 + b} < 1 + ax^+.$$

We can square and retain the inequality provided  $1 + ax^+ > 0$ . However,

$$2(ax^+ + 1) = 2 + (b - 1) + \sqrt{(b - 1)^2 + 4a} = b + 1 + \sqrt{(b - 1)^2 + 4a} > 0.$$

Squaring gives

$$(ax^+)^2 + b < 1 + 2ax^+ + (ax^+)^2, \quad \text{or} \quad 2ax^+ + 1 - b > 0,$$

so we require  $b - 1 + \sqrt{(b - 1)^2 + 4a} + 1 - b > 0$ , or  $\sqrt{(b - 1)^2 + 4a} > 0$ .

Now consider the inequality  $\lambda^+ > -1$ , or equivalently  $\sqrt{(ax^+)^2 + b} > ax^+ - 1$ , clearly holding when  $ax^+ - 1 < 0$ . Assume  $ax^+ - 1 > 0$ , then we can check that

$$a < 3(b - 1)^2/4, \quad \text{so} \quad \lambda^+ > -1 \quad \text{for} \quad a < 3(b - 1)^2/4.$$

A similar analysis shows that  $|\lambda^-| < 1$  when  $a < 3(b - 1)^2/4$ . Putting all this together gives:

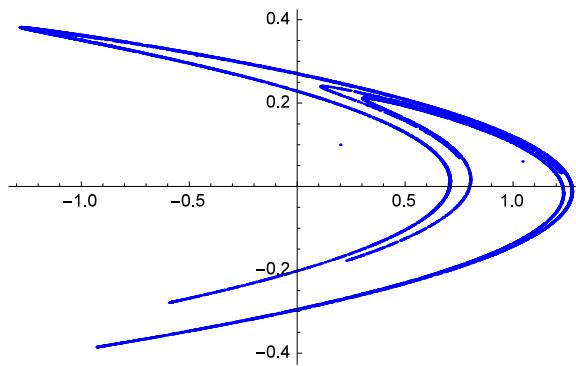
**Proposition 13.2.2** *The fixed point  $(x^+, y^+)$  of the Hénon map  $H : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , is stable when  $a < 3(b - 1)^2/4$ . It is unstable when  $a > 3(b - 1)^2/4$ .*

In Exercises 13.2 # 6, we see that a 2-cycle is created when the fixed point  $(x^+, y^+)$  ceases to be stable, i.e., when  $a = 3(1 - b)^2/4$ . The fixed point  $(x^-, y^-)$  is unstable.

**13.2.3 The Hénon Attractor.** The classical Hénon map studied in [68], uses the parameters  $a = 1.4$  and  $b = .3$ . Hénon showed that there is a quadrilateral region  $R$  in  $\mathbb{R}^2$ , for which points inside  $R$ , when iterated under  $H$ , map toward an invariant set called the Hénon attractor. This is a set invariant under  $H$  which seems to have the cross section of a Cantor set. This set has a very complicated structure of a fractal nature that is not fully understood. To quote David Ruelle [110]: “Ask your computer to plot the points  $H^n(0, 0)^t$ , and you will find that they accumulate, as

$n \rightarrow \infty$ , on a convoluted fractal set  $A$  known as the *Hénon attractor*. This set is prototypical of what one wants to call a *strange attractor*. Such objects often arise when a diffeomorphism  $f$  stretches and folds an open set  $U$ , and maps the closure  $f(\overline{U})$  inside  $U$ . The strange attractor is visualized when a computer plots the points  $x_n = f^n(x_0)$ , with almost any initial value  $x_0$  in  $U$ ."

Below, we show the result of plotting  $H^n(0, 0)^t$  for  $n = 10,000$ , ignoring the first 1,000 plots. To see that  $H$  can be thought of as a combination of stretching and folding (see Exercises 13.2 # 4(b)).



The Attractor of the Classical Hénon Map ( $a = 1.4$ ,  $b = 0.3$ ).

## Exercises 13.2

1. Find the fixed points and determine their stability for the following maps  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ :

$$(i) \quad F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix} \quad (ii) \quad F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 - y^2 + x - 1 \\ 2xy + y \end{pmatrix}.$$

2. Determine the values of  $a$  and  $b$  for which the map  $F$  has a fixed point, and determine the stability when  $a = 6$  and  $b = 1$ :

$$F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ ay - bx - y^3 \end{pmatrix}.$$

3. (a) Show that if  $b \neq 0$ , the Hénon map  $H$  is invertible. Find the inverse map  $H^{-1}$ .
- (b) Show that when  $b = 1$ ,  $H$  is conjugate to its inverse  $H^{-1}$ . (Hint: Show that  $K(x, y)^t = (-y, -x)^t$  defines a conjugation map).
4. (a) Show that the Hénon map contracts areas for  $0 < b < 1$ , and is area preserving for  $b = 1$ . (Hint: It suffices to show that the Jacobian of  $H$  has the property  $|JH(x, y)^t| < 1$  for all  $(x, y)$  when  $0 < b < 1$ , with equality when  $b = 1$ ).
- (b) Show that the Hénon map is a combination of contractions and foldings, by showing that it is a composition of the following three maps:  $H = H_1 \circ H_2 \circ H_3$ , where
- $$H_1(x, y)^t = (y, x)^t, \quad H_2(x, y)^t = (bx, y)^t, \quad H_3(x, y)^t = (x, 1 - ax^2 + y)^t.$$
- (Note that  $H_1$  reflects in the line  $y = x$ ,  $H_2$  is a contraction in the  $x$ -direction, and  $H_3$  is area preserving).
5. (a) Show that the fixed point  $(x^-, y^-)$  (where  $x^- = \frac{b-1-\sqrt{(1-b)^2+4a}}{2a}$ ), of the Hénon map, is unstable (assume  $(1-b)^2 + 4a > 0$ ).
- (b) Show that the fixed point in (a) is a saddle for  $-(1-b)^2/4 < a < 3(1-b)^2/4$ ,  $a \neq 0$ .
6. Show that a 2-cycle for the Hénon map results from the solution to the equation
- $$a^3x^4 - 2a^2x^2 + (1-b)^3x + (a - (1-b)^2) = 0.$$
- Find the 2-cycle using the fact that fixed points give rise to solutions of the above equation. Deduce that there is a 2-cycle when  $a > 3(1-b)^2/4$ . (Hint: You may wish to use a computer algebra system for this problem).
- ### 13.3 Circle Maps Induced by Linear Transformations on $\mathbb{R}$ .
- Consider the linear transformation  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = 2x$ . This map gives rise to a mapping defined on the unit circle  $\mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$ , which is determined

as follows: think of the unit circle as the interval  $[0, 1]$  bent around so that 0 and 1 become identified, with the resulting figure appropriately scaled.

With this identification we can write

$$\mathbb{S}^1 = \{e^{2\pi ix} : x \in [0, 1)\} = \{e^{i\theta} : \theta \in [0, 2\pi)\}.$$

More formally, we are identifying  $\mathbb{S}^1$  with the interval  $[0, 1)$  via the mapping  $H : [0, 1) \rightarrow \mathbb{S}^1$ ,  $H(x) = e^{2\pi ix}$ .

Although  $\mathbb{S}^1$  and  $[0, 1)$  are not homeomorphic as metric spaces, when we identify 0 and 1 in this way, we think of them as being different versions of the same space. This idea can be put on a more rigorous footing (see for example [32]).

If we define a mapping  $D : [0, 1) \rightarrow [0, 1)$  by

$$D(x) = 2x \pmod{1},$$

then  $D$  is the transformation defined on  $[0, 1)$  induced by the linear transformation  $f$ . We think of  $D$  as a circle map as it is conjugate to the map  $T : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $T(e^{ix}) = e^{2ix}$  via the map  $H$ , which we are now regarding as a conjugacy.  $T$  is just the squaring function (angle doubling map),  $T(z) = z^2$  defined on  $\mathbb{S}^1$ .

We ask the question: when does a linear transformation  $f(x) = ax$  on  $\mathbb{R}$  give rise to a well defined circle map?

In order to be well defined, we require the induced map  $T(e^{ix}) = e^{aix}$  to “wrap around” the circle in the appropriate manner. This means that

$$T(e^{i(x+2\pi)}) = T(e^{ix}) \quad \text{for } x \in [0, 1), \quad \text{or} \quad e^{ai(x+2\pi)} = e^{aix},$$

so that  $e^{2\pi ia} = 1$ . Thus  $a \in \mathbb{Z}$  is the required condition. It is now readily seen that  $a \in \mathbb{Z}$  is a necessary and sufficient condition for  $f(x) = ax$  to give rise to a circle map.

We have examined the dynamical properties of  $D(x) = 2x \pmod{1}$ , (the doubling map), and maps such as  $D_3(x) = 3x \pmod{1}$ , can be analyzed in a similar fashion. Such maps, when regarded as maps of the circle, are continuous and onto, but not one-to-one. They have very strong chaotic properties (for  $a \neq 1$ ).

### 13.4 Endomorphisms of the Torus.

Given a linear map  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ ,  $f \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix}$ , we can use it to define a map from  $X = [0, 1] \times [0, 1]$  to itself in an analogous way to the 1-dimensional

case. The corresponding map  $T'$  is now defined on the *torus*  $\mathbb{S}^1 \times \mathbb{S}^1$  by

$$T' \begin{pmatrix} e^{ix_1} \\ e^{ix_2} \end{pmatrix} = \begin{pmatrix} e^{i(ax_1+bx_2)} \\ e^{i(cx_1+dx_2)} \end{pmatrix}.$$

The torus can be thought of as being obtained from the unit square  $[0, 1] \times [0, 1]$  in the following way: identify opposite edges of the square to form a cylinder. Then bend around to join the remaining edges of the cylinder to form the donut shape called a torus. In this way, if we define  $T : [0, 1) \times [0, 1) \rightarrow [0, 1) \times [0, 1)$  by

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix} \pmod{1}.$$

Then (by examining the wrap around properties of  $T'$ ), we see that  $T$  is well defined if and only if  $a, b, c, d \in \mathbb{Z}$  (see the exercises). As before, it can be seen that  $T$  is a continuous map of the torus which is onto when the matrix  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  is non-singular ( $\det(A) \neq 0$ ), but is not in general one-to-one.

**Example 13.4.1** Consider  $T : [0, 1) \times [0, 1) \rightarrow [0, 1) \times [0, 1)$  defined by

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \pmod{1}.$$

In other words,

$$T(x_1, x_2)^t = \begin{cases} (2x_1, 2x_2)^t; & 0 \leq x_1 < 1/2, 0 \leq x_2 < 1/2, \\ (2x_1 - 1, 2x_2)^t; & 1/2 \leq x_1 < 1, 0 \leq x_2 < 1/2, \\ (2x_1, 2x_2 - 1)^t; & 0 \leq x_1 < 1/2, 1/2 \leq x_2 < 1, \\ (2x_1 - 1, 2x_2 - 1)^t; & 1/2 \leq x_1 < 1, 1/2 \leq x_2 < 1, \end{cases}$$

where  $(x_1, x_2)^t$  denotes the transpose of  $(x_1, x_2)$ .

The map  $T$  is not one-to-one, but is in fact everywhere four-to-one - if we partition  $[0, 1) \times [0, 1)$  into four equal sub-squares, we see that each gets mapped onto  $[0, 1) \times [0, 1)$ . Consequently,  $T$  is onto, and it can be shown to be area preserving. It is useful to use the binary representations of  $x_1$  and  $x_2$ , to establish the properties of this map. In Exercises 13.4 # 5, we see that  $T$  is chaotic.

Let us denote the  $n$ -dimensional torus by  $\mathbb{T}^n = [0, 1]^n = [0, 1) \times [0, 1) \times \cdots \times [0, 1)$  (the direct product of  $n$  copies of  $[0, 1)$ ). Notice that  $\mathbb{T}^n$  is a group when given the group operation of addition modulo one on each coordinate. Recall that an *endomorphism* of a group is a homomorphism that is also onto. If the homomorphism is also one-to-one, then it is a *group automorphism*. (A map  $\phi : G \rightarrow G$  from a group

$\langle G, * \rangle$  (where  $*$  is the group operation), is a *group homomorphism*, if  $\phi(x * y) = \phi(x) * \phi(y)$  for all  $x, y \in G$ ). It is known that a continuous endomorphism of the torus is transitive if and only if there are no eigenvalues which are roots of unity ([127], Theorem 1.11). We need the following characterization of endomorphisms of the torus:

**Proposition 13.4.2** (a) *Every endomorphism  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is of the form*

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \pmod{1},$$

for  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ , where  $a, b, c, d \in \mathbb{Z}$ .

(b) *The endomorphism  $T$  maps  $\mathbb{T}^2$  onto  $\mathbb{T}^2$  if and only if  $\det(A) \neq 0$ .*

(c) *The endomorphism  $T$  is an automorphism of  $\mathbb{T}^2$  if and only if  $\det(A) = \pm 1$ .*

**Proof.** We omit the proof (see for example Walters [127], where the proof is given for endomorphisms of  $\mathbb{T}^n$ ).

### Exercises 13.4

1. Prove that the maps  $T(x) = ax \pmod{1}$  on  $[0, 1]$ , and  $T'(e^{ix}) = e^{iax}$  on  $\mathbb{S}^1$  are conjugate for  $a \in \mathbb{Z}$ . (Here we are treating  $[0, 1]$  and  $S^1$  as homeomorphic spaces as described in Section 13.3).

2. Prove that the map  $T : [0, 1] \times [0, 1] \rightarrow [0, 1] \times [0, 1]$

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix} \pmod{1},$$

is well defined if and only if  $a, b, c, d \in \mathbb{Z}$ .

3. Prove that the maps  $T : [0, 1] \times [0, 1] \rightarrow [0, 1] \times [0, 1]$  and  $T' : \mathbb{S}^1 \times \mathbb{S}^1 \rightarrow \mathbb{S}^1 \times \mathbb{S}^1$ , defined by

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix} \pmod{1},$$

and

$$T' \begin{pmatrix} e^{ix_1} \\ e^{ix_2} \end{pmatrix} = \begin{pmatrix} e^{i(ax_1+bx_2)} \\ e^{i(cx_1+dx_2)} \end{pmatrix},$$

are conjugate when  $a, b, c, d \in \mathbb{Z}$ .

4. Prove, that the map  $T$  of the last exercise is onto when  $\det(A) \neq 0$ , and one-to-one if and only if  $\det(A) = \pm 1$ . In this case,  $T$  is an invertible area-preserving mapping of the torus. The torus can be given a group structure in the obvious way (with addition modulo one), and we now see that  $T$  will be an automorphism of this group. When the determinant is not  $\pm 1$ ,  $T$  is still a homomorphism (actually an endomorphism of the torus since it is onto).

5. Let  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  be the endomorphism of the torus induced by the matrix  $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ :  $T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \pmod{1}$ . Show that  $T$  is chaotic on  $\mathbb{T}^2$ .

6. The Baker's transformation  $B : [0, 1]^2 \rightarrow [0, 1]^2$  is defined by

$$B(x, y) = \begin{cases} (2x, y/2) & ; \quad 0 \leq x < 1/2 \\ (2x - 1, y/2 + 1/2) & ; \quad 1/2 \leq x \leq 1 \end{cases}$$

(a) Describe how  $B$  acts on the unit square  $[0, 1]^2$ .

(b) Express  $B$  in terms of what it does to  $(x, y)$  when  $x$  and  $y$  are written in terms of their binary expansions. Deduce that  $B$  is invertible.

(c) Find the fixed points and the period-2 points of  $B$ .

(d)\* Show that  $B$  is chaotic on  $[0, 1]^2$ .

7\*. If  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is an endomorphism of the torus induced by a matrix  $A$  (so  $\det(A) \neq 0$ ), having no eigenvalues that are roots of unity, show that  $T$  is transitive. Show that the converse is also true.

### 13.5 Hyperbolic Toral Automorphisms.

In this section, we study an important class of 2-dimensional chaotic maps which are induced by linear maps. We see a type of behavior that cannot happen in the

1-dimensional case. Although we restrict our attention to the 2-dimensional case, proofs generalize in a straight forward manner to higher dimensions. See [32] for more details. Let  $T$  be an endomorphism of the torus  $\mathbb{T}^2$  defined by:

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix} \bmod 1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \bmod 1,$$

where  $a, b, c, d \in \mathbb{Z}$ .

**Definition 13.5.1** If the matrix  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  has integer entries with  $|\det(A)| = 1$ , then the map  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  defined above, is called a *torus automorphism*. If neither of the eigenvalues has absolute value equal to one, then  $T$  is called a *hyperbolic torus automorphism*.

To say that  $T$  is an automorphism means that  $T$  is one-to-one, onto, and in particular, it is an automorphism of the group  $\mathbb{T}^2$  (with respect to addition modulo one).  $T$  is also a homeomorphism as it is continuous, with continuous inverse. Notice that since  $\det(A) = \lambda_1\lambda_2$ , where  $\lambda_1$  and  $\lambda_2$  are the eigenvalues of  $A$ ,  $|\lambda_1| = 1/|\lambda_2|$ . So one of the eigenvalues has absolute value greater than one, and one has absolute value less than one.

**Examples 13.5.2** (i) A classical example is the hyperbolic toral automorphism due to Arnold and Avez [5], called the *CAT-map*. It is induced by the matrix  $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ .  $\det(A) = 1$  and  $\lambda_1 = (3 + \sqrt{5})/2$ ,  $\lambda_2 = 1/\lambda_1$ . (C = continuous, A = automorphism, and T = torus, although CAT also refers to the effect the map has on the face of a cat contained in the unit square).

(ii) If  $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ , then  $\det(A) = 1$ , but  $\lambda = 1$  is the only eigenvalue, so  $A$  does not induce a hyperbolic toral automorphism.

**Proposition 13.5.3** *A hyperbolic toral automorphism  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ , is a homeomorphism.*

**Proof.** If  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ , then  $\det(A) = ad - bc = \pm 1$ , so  $A^{-1} = \pm \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ . It follows that the inverse of  $T$  is also a hyperbolic toral automorphism having the same eigenvalues.

It suffices to show that  $T$  is continuous. Denote the first coordinate of  $T \begin{pmatrix} x \\ y \end{pmatrix}$  by  $\left[ T \begin{pmatrix} x \\ y \end{pmatrix} \right]_1$ , and similarly for the second coordinate. Then  $|x_1 - x_2| < \epsilon$  and  $|y_1 - y_2| < \epsilon$  implies that

$$\left| \left[ T \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \right]_1 - \left[ T \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right]_1 \right| = |ax_1 + by_1 - ax_2 - by_2| \leq (|a| + |b|)\epsilon,$$

and

$$\left| \left[ T \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \right]_2 - \left[ T \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right]_2 \right| = |cx_1 + dy_1 - cx_2 - dy_2| \leq (|c| + |d|)\epsilon,$$

and continuity follows. □

Our aim now is to show that hyperbolic toral automorphisms are chaotic. Recall that for a continuous map on a metric space, it suffices to show the periodic points are dense, and the map is transitive.

**Theorem 13.5.4** *Let  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  be a hyperbolic toral automorphism induced by a matrix  $A$ .*

- (i) *The eigenvalues of  $A$  are real and irrational. The eigenspaces are straight lines through the origin having irrational slopes.*
- (ii) *The periodic points of  $T$  are dense in  $\mathbb{T}^2$ .*
- (iii)  *$T$  is transitive, so it is a chaotic map.*

**Proof.** (i) See the exercises for the proof that the eigenvalues are real. Let  $\lambda = m/n$  (where  $m, n \in \mathbb{Z}, n \neq 0$ ), be a solution to the characteristic equation  $\lambda^2 - \tau\lambda + \Delta$ , where  $m$  and  $n$  have no common factors,  $\Delta, \tau \in \mathbb{Z}$ . A standard argument leads to a contradiction:  $\Delta = \pm 1$ . We see that  $m$  must divide  $n^2$ , so must also divide  $n$ , contradicting our assumption. It now readily follows that the lines that define the eigenspaces have irrational slopes.

(ii) We now show that the periodic points of  $T$  are dense. Let  $x_1, x_2 \in \mathbb{T}^2$  be of the form  $x_1 = p_1/q, x_2 = p_2/q, p_1, p_2, q \in \mathbb{Z}^+$ . Then

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} (ap_1 + bp_2)/q \\ (cp_1 + dp_2)/q \end{pmatrix} \bmod 1.$$

The number of points in  $\mathbb{T}^2$  of the form  $(s/q, t/q)$  with  $0 \leq |s|, |t| < q$  is finite, so there exists integers  $m, n \geq 0$  with  $m < n$  such that  $T^n(x_1, x_2)^t = T^m(x_1, x_2)^t$ , so  $T^{n-m}(x_1, x_2)^t = (x_1, x_2)^t$ . In other words, every point of this form is periodic. Since such points are dense in  $[0, 1] \times [0, 1]$ , the periodic points of  $T$  are dense in  $\mathbb{T}^2$ .

(iii) To show that  $T$  is chaotic, it suffices to show that  $T$  is transitive. We give a sketch of the proof.

Since the matrix  $A$  has real irrational eigenvalues  $|\lambda_1| > 1$  and  $|\lambda_2| < 1$  (say), the unstable manifold  $W^u(0)$  and the stable manifold  $W^s(0)$  for the fixed point 0 (which are just the respective eigenspaces in  $\mathbb{R}^2$ ), have irrational slopes. When these are projected to the torus (also denoted by  $W^u$  and  $W^s$ ), the resulting unstable and stable manifolds for  $T$  wrap around  $\mathbb{T}^2$  to form dense sets (since the slopes are irrational).

### Exercises 13.5

1. Which of the following matrices induce hyperbolic toral automorphisms:

$$(i) \begin{bmatrix} 3 & 2 \\ 4 & 3 \end{bmatrix}, \quad (ii) \begin{bmatrix} 1 & -2 \\ 3 & -3 \end{bmatrix}, \quad (iii) \begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}, \quad (iv) \begin{bmatrix} 2 & 1/2 \\ 2 & 1 \end{bmatrix}, \quad (v) \begin{bmatrix} 3 & 1 \\ -7 & -2 \end{bmatrix}?$$

2. (a) Find the period-2 and period-3 points of the hyperbolic toral automorphism (the *cat map*), induced by the matrix  $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ .

(b) Show that  $A^n = \begin{bmatrix} F_{2n} & F_{2n-1} \\ F_{2n-1} & F_{2n-2} \end{bmatrix}$ , where  $F_n$  is the  $n$ th term in the Fibonacci sequence ( $F_0 = 1, F_1 = 1, F_2 = 2, \dots$ ).

(c) Use the diagonalization of  $A$ , to deduce that

$$F_{2n-1} = \frac{1}{\sqrt{5}} \left( \frac{3 + \sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left( \frac{3 - \sqrt{5}}{2} \right)^n,$$

and find a similar formula for  $F_{2n}$ .

3. Let  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  be the hyperbolic toral automorphism induced by  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . Find the eigenvalues, and show that they are real. Show that the eigenspaces have irrational slopes.

4. Let  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  be a torus automorphism. Let  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  be the corresponding matrix with eigenvalues  $\lambda_1$  and  $\lambda_2$ . Note that if  $T$  is not hyperbolic then the eigenvalues must be of absolute value one and one of the following holds: (i) the eigenvalues are not real, (ii)  $\lambda_1 = 1$  and  $\lambda_2 = -1$ , (iii)  $\lambda_1 = 1 = \lambda_2$ , or (iv)  $\lambda_1 = -1 = \lambda_2$  (see [81], where it is shown that in each of these cases,  $T$  cannot be chaotic).

(a) If  $\lambda_1 = 1$  and  $\lambda_2 = -1$ , show that  $a + d = 0$  and  $\det(A) = -1$ .

(b) From (a), show that  $A$  is of the form  $\begin{bmatrix} a & b \\ c & -a \end{bmatrix}$ , and deduce that  $T$  is not chaotic.

(c) Show that if  $\lambda_1$  and  $\lambda_2$  are non-real, then  $\det(A) = 1$ , and the eigenvalues are either (i)  $\pm i$ , (ii)  $(1 \pm i\sqrt{3})/2$ , or (iii)  $(-1 \pm \sqrt{3})/2$ .

(d) From (c), show that in each case  $T$  cannot be chaotic.

## CHAPTER 14

### Elementary Complex Dynamics.

In Chapters 1 and 2, we looked at the fixed points and periodic points of one-dimensional dynamical systems arising from real maps. We start this chapter by generalizing those results to the case of complex maps. This topic, called complex dynamics, is an important and deep subject worthy of study in its own right. Since many of the deeper results of this field are beyond the scope of this text, we will be content with giving a brief introduction, which it is hoped, will persuade the readers to continue their studies.

Much of the modern theory of dynamical systems arose from the study of Newton's method in the complex plane. The work of Schröder, Cayley, Fatou, and Julia is particularly noteworthy in this regard (see "A History of Complex Dynamics" by Daniel Alexander [1] for a detailed treatment of their work). It seems that Schröder was the first to seriously consider the idea of iterating a function, although, as he himself says, he initiated these studies for "no particular reason".

In order to keep the exposition as self contained as possible, we start with a review of some elementary complex analysis.

#### 14.1 The Complex Numbers.

We briefly remind the reader of some of the basic properties of the complex numbers

$$\mathbb{C} = \{a + ib : a, b \in \mathbb{R}, i^2 = -1\}.$$

The complex numbers are often defined as an algebraic structure consisting of the set of points in the plane  $(a, b) \in \mathbb{R}^2$ , together with the operations of addition and multiplication defined by

$$(a, b) + (c, d) = (a + c, b + d); \quad (a, b) \cdot (c, d) = (ac - bd, ad + bc).$$

It is then easily verified that with this addition and multiplication,  $\mathbb{R}^2$  is a field with identity  $(1, 0)$  and an element  $i = (0, 1)$  having the property  $i^2 = (-1, 0) = -(1, 0)$ . We identify  $(1, 0)$  with  $1 \in \mathbb{R}$  and write  $(a, b) = a(1, 0) + b(0, 1) = a + ib$  and denote the set obtained, by  $\mathbb{C}$ . Because the complex numbers form a field, all the

usual laws of arithmetic and algebra hold. For example, if  $(a, b) \neq (0, 0)$ , then the multiplicative inverse of  $(a, b)$  is  $(a, b)^{-1} = (a/(a^2 + b^2), -b/(a^2 + b^2))$ , so that if  $z = a + ib$ ,  $1/z = (a + ib)^{-1} = (a - ib)/(a^2 + b^2) = \bar{z}/|z|^2$ .

The *absolute value* or *modulus* of  $z \in \mathbb{C}$  is  $|z| = \sqrt{a^2 + b^2}$ , where  $z = a + ib$  and  $a, b \in \mathbb{R}$  are the *real* and *imaginary* parts of  $z$  respectively, written  $a = \operatorname{Re}(z)$  and  $b = \operatorname{Im}(z)$ . We can also write  $|z|^2 = z \cdot \bar{z}$ , where  $\bar{z} = a - ib$ , is the *complex conjugate* of  $z$ .

Points in  $\mathbb{C}$  may also be represented in *polar* form, or what is called *modulus-argument* form. If  $z = a + ib$ ,  $a, b \in \mathbb{R}$ , set  $r = |z|$ , and let  $\theta$  be the angle subtended by the  $x$ -axis and the ray going from  $(0, 0)$  to  $(a, b)$ .  $\theta$  is called the *argument* of  $z$  (written  $\arg(z)$ ), and is measured in the anti-clockwise direction. We see that  $a = r \cos(\theta)$ ,  $b = r \sin(\theta)$ , and

$$z = r(\cos(\theta) + i \sin(\theta)).$$

We do not define the argument when  $z = 0$  as this leads to ambiguities. The argument of  $z \in \mathbb{C}$  is not unique, since arguments differing by a multiple of  $2\pi$  give rise to the same complex number. The argument  $\theta$  of  $z$  for which  $-\pi < \theta \leq \pi$ , is called the *principal argument* of  $z$ , and is denoted by  $\operatorname{Arg}(z)$ .

Notice that if we formally manipulate the power series for sine and cosine, we obtain the identity

$$e^{i\theta} = \cos(\theta) + i \sin(\theta).$$

This equation, which we take as the definition of  $e^{i\theta}$  for  $\theta \in \mathbb{R}$ , is *Euler's formula*, and may be justified by showing that  $e^{i\theta}$  really behaves like an exponential. For example,  $(e^{i\theta})^n = e^{in\theta}$ , for  $n \in \mathbb{Z}$ , is a consequence of *DeMoivre's Theorem*:

$$(\cos(\theta) + i \sin(\theta))^n = \cos(n\theta) + i \sin(n\theta), \quad n \in \mathbb{Z}, \quad \theta \in \mathbb{R},$$

(clearly true when  $n = 1$ , then proved by induction for  $n > 1$ , and then checked for  $n \leq -1$ ). Care needs to be taken when  $n$  is a fraction such as  $n = 1/2$ , as the expression becomes multi-valued.

If we represent  $z, w \in \mathbb{C}$  as  $z = re^{i\theta}$  and  $w = se^{i\phi}$ , then we see that  $z \cdot w = rse^{i(\theta+\phi)}$  and  $z/w = r/se^{i(\theta-\phi)}$  (when  $s \neq 0$ ).

## 14.2 Analytic Functions in the Complex Plane.

In proving theorems about complex functions and limits of sequences, the various forms of the triangle inequality are often useful:

$$|z + w| \leq |z| + |w|, \quad \text{and} \quad ||z| - |w|| \leq |z - w|, \quad \text{for all } z, w \in \mathbb{C}.$$

As we saw in Chapter 4, the set  $\mathbb{C}$  together with the distance  $d(z, w) = |z - w|$  is a metric space. Using this metric, we can define limits and continuity as in an arbitrary metric space: open balls centered on  $z_0 \in \mathbb{C}$ , are sets of the form

$$B_\delta(z_0) = \{z \in \mathbb{C} : |z - z_0| < \delta\},$$

for  $\delta > 0$ .

In particular, for a function  $f(z)$  defined in an open ball centered on  $z_0 \in \mathbb{C}$  (but not necessarily defined at  $z_0$  itself),  $\lim_{z \rightarrow z_0} f(z) = L$ , if given any real number  $\epsilon > 0$ , there exists a real number  $\delta > 0$ , such that if  $|z - z_0| < \delta$ , then  $|f(z) - L| < \epsilon$ . The usual rules for limits are easily verified. Continuity can then be defined as follows:

**Definition 14.2.1** (i) Let  $f$  be defined on an open ball centered on  $z_0 \in \mathbb{C}$ . Then  $f$  is *continuous* at  $z_0$  if  $\lim_{z \rightarrow z_0} f(z) = f(z_0)$ .

(ii) If the domain of  $f$  is an open set  $D \subseteq \mathbb{C}$ , then  $f$  is continuous on  $D$  if it is continuous at every point of  $D$ .

All the results about continuity on metric spaces are applicable. For example if  $f$  is continuous on  $\mathbb{C}$  and the orbit of  $z_0$  under  $f$ ,  $O(z_0) = \{f^n(z_0) : n \in \mathbb{Z}^+\}$  has a unique limit point  $\alpha$ , then the continuity of  $f$  implies that  $\alpha$  is a fixed point of  $f$ .

**Proposition 14.2.2** Let  $f, g : \mathbb{C} \rightarrow \mathbb{C}$  be functions continuous at  $z = a$  and let  $\alpha \in \mathbb{C}$ . Then

- (i)  $\bar{f}$ ,  $f + g$ ,  $\alpha f$ ,  $f \cdot g$  are continuous at  $a$ . If  $g(a) \neq 0$ , then  $f/g$  is continuous at  $a$ .
- (ii)  $|f|$ ,  $\operatorname{Re}(f)$  and  $\operatorname{Im}(f)$  are continuous at  $a$ , as functions from  $\mathbb{C}$  to  $\mathbb{R}$  (where  $\operatorname{Re}(f)$  and  $\operatorname{Im}(f)$  denote the real part and imaginary part of  $f$  respectively).
- (iii) The composite of continuous functions is continuous (where defined).

**Proof of (ii).** Let  $\epsilon > 0$ . Since  $f$  is continuous at  $z = a$ , there exists  $\delta > 0$  such that if  $0 < |z - a| < \delta$ , then  $|f(z) - f(a)| < \epsilon$ . By the triangle inequality,

$$||f(z)| - |f(a)|| \leq |f(z) - f(a)| < \epsilon,$$

so that continuity of  $|f|$  is immediate.

The continuity of  $\operatorname{Re}(f)$  follows in a similar way, but using the inequality: for  $z = c + id$ ,  $c, d \in \mathbb{R}$ ,  $|\operatorname{Re}(z)| = |c| \leq \sqrt{c^2 + d^2} = |z|$ . The continuity of  $\operatorname{Im}(f)$  is similar.

□

The derivative of a complex function can now be defined in a natural way, and the usual rules of differentiation will hold:

**Definition 14.2.3** (i) The *derivative*  $f'(z_0)$ , of a function  $f : \mathbb{C} \rightarrow \mathbb{C}$  is defined by

$$f'(z_0) = \lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0},$$

provided that this limit exists.

(ii) The function  $f(z)$  is said to be *analytic* in an open ball  $B_\delta(z_0)$  if the derivative exists at every point in the ball. We talk about  $f(z)$  being analytic in a *neighborhood* of the point  $z_0$ , if there is an open ball containing  $z_0$  throughout which  $f(z)$  is analytic.  $f$  is analytic in an open set  $D$  if it is analytic inside every open ball contained in  $D$ .

The same limit rules and rules of algebra, as in the real case, show that for  $z \in \mathbb{C}$ , the derivative of  $f(z) = z^n$  is  $f'(z) = nz^{n-1}$  for  $n \in \mathbb{Z}^+$ . Thus  $f(z) = z^n$  is analytic on  $\mathbb{C}$ . The usual differentiation rules apply: the sum, product, quotient and chain rules.

It can be shown that functions analytic on some domain  $D$  (for our purposes we may assume that  $D$  is an open ball, or possibly all of  $\mathbb{C}$ ), have derivatives of all orders throughout that domain. In particular,  $f'(z)$  will be continuous. In fact, the property of being analytic on  $B_\delta(z_0)$ , is equivalent to  $f(z)$  having a power series expansion about  $z_0$ , valid throughout  $B_\delta(z_0)$ :

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n, \quad a_n \in \mathbb{C}.$$

One of the first results of iteration theory was Schröder's Theorem, (a variation of the real version from Chapter 1), concerning attracting fixed points. The first rigorous proof was given by Koenigs: see [1] and [78].

**Remarks 14.2.4** We need the following property of limits of functions, showing a behavior similar to the case of real functions:

- (i) If  $\lim_{z \rightarrow a} f(z) = L$  with  $|L| > 0$ , then there is a ball  $B_\delta(a)$  for which  $|f(z)| > 0$  for all  $z \in B_\delta(a)$ ,  $z \neq a$ .
- (ii) If  $\lim_{z \rightarrow a} f(z) = L$ , then  $\lim_{z \rightarrow a} |f(z)| = |L|$ . The converse is true when  $L = 0$ .

**Theorem 14.2.5** (Schröder) *Let  $f(z)$  be analytic in a neighborhood of  $\alpha \in \mathbb{C}$  with  $f(\alpha) = \alpha$  and  $|f'(\alpha)| < 1$ . Then for all  $z$  in some open ball centered on  $\alpha$ ,*

$$\lim_{n \rightarrow \infty} f^n(z) = \alpha.$$

**Proof.** Although different from the proof in the real case, the following proof will also work in that situation (see Theorem 1.4.4). By the definition of the derivative,

$$|f'(\alpha)| = \left| \lim_{z \rightarrow \alpha} \frac{f(z) - f(\alpha)}{z - \alpha} \right| = \lim_{z \rightarrow \alpha} \left| \frac{f(z) - f(\alpha)}{z - \alpha} \right| < 1.$$

From Remarks 14.2.4, there is an open ball  $D$  centered on  $\alpha$ , and a constant  $0 < \lambda < 1$  for which

$$\left| \frac{f(z) - \alpha}{z - \alpha} \right| < \lambda$$

on  $D$ . In other words,

$$|f(z) - \alpha| < \lambda |z - \alpha|.$$

Substituting  $f(z)$  (which must also lie in  $D$ ), for  $z$  in the last inequality, gives

$$|f^2(z) - \alpha| < \lambda |f(z) - \alpha| < \lambda^2 |z - \alpha|.$$

Continuing in this way, we obtain

$$|f^n(z) - \alpha| < \lambda^n |z - \alpha|.$$

Since  $\lambda^n \rightarrow 0$  as  $n \rightarrow \infty$ , the result follows. □

**Definition 14.2.6** Just as in the general case, we can define the notion of stable and asymptotically stable fixed points. For analytic functions with  $f(\alpha) = \alpha$ , a sufficient condition for  $\alpha$  being an asymptotically stable fixed point, is the requirement that  $|f'(\alpha)| < 1$ . If this is the case, we shall simply say that  $\alpha$  is an *attracting fixed point*, and we say  $\alpha$  is *repelling* if  $|f'(\alpha)| > 1$ . These are the hyperbolic cases. In the non-hyperbolic case (where  $|f'(\alpha)| = 1$ ),  $\alpha$  is sometimes said to be a *neutral fixed point*.

Following this convention, and in a way similar to the real case, if  $p$  is a point of period  $r$ , then  $p$  is said to be an *attracting periodic point* if

$$|(f^r)'(p)| = |f'(p)f'(f(p))f'(f^2(p)) \dots f'(f^{r-1}(p))| < 1,$$

and  $p$  is a *repelling periodic point* if

$$|(f^r)'(p)| = |f'(p)f'(f(p))f'(f^2(p)) \dots f'(f^{r-1}(p))| > 1.$$

Such periodic points are said to be *hyperbolic*, and we have the *non-hyperbolic* case when  $|f'(p)| = 1$ . Theorem 14.2.5 is now seen to generalize to the case of an attracting periodic point.

### Exercises 14.2

1. Prove the triangle inequality in  $\mathbb{C}$ :  $|z + w| \leq |z| + |w|$  for all  $z, w \in \mathbb{C}$ . (Hint: Use the fact that  $|z + w|^2 = (z + w)(\overline{z + w})$  and  $\operatorname{Re}(z) \leq |z|$ ). Deduce  $||z| - |w|| \leq |z - w|$  for all  $z, w \in \mathbb{C}$ .
  
2. Let  $f_c : \mathbb{C} \rightarrow \mathbb{C}$ ,  $f_c(z) = z^2 + c$ , for  $c \in \mathbb{C}$ . Recall that a period- $n$  point  $z_0$  is super-attracting if  $(f^n)'(z_0) = 0$ .
  - (a) If  $z_0$  and  $z_1$  are the fixed points of  $f_c$ , show that  $f'_c(z_0) + f'_c(z_1) = 2$ . Deduce that there can be at most one attracting fixed point. Give an example to show that  $f_c$  may not have any attracting fixed points.
  - (b) Show that if  $f_c$  has a super-attracting fixed point  $z_0$ , then  $z_0 = 0$  and  $c = 0$ .
  - (c) Find the value of  $c$  for which  $f_c$  has a super-attracting 2-cycle. Also find the 2-cycle.
  - (d) Why is it that  $z = 0$  is a point on the orbit of a cycle, if and only if the cycle is super-attracting?
  - (e) If  $f_c$  has a super-attracting 3-cycle, show that  $c$  satisfies the equation

$$c^3 + 2c^2 + c + 1 = 0.$$

(Hint: It is easiest to look at the orbit of 0 under  $f_c$ ).

- (f) Do all the solutions to the equation in part (e) give rise to a super-attracting 3-cycle? If not, what do they give? (The equation has solutions:  $c = -1.75488$ ,  $c = -0.122561 + 0.744862i$  and  $c = -0.122561 - 0.744862i$ ).

3. Prove that if  $f(z) = z^n$ ,  $n \in \mathbb{N}$ , then  $f'(z) = nz^{n-1}$ .

4. Let  $f$  be a continuous function on  $\mathbb{C}$ . Let  $z_0 \in \mathbb{C}$ . Show that if the orbit of  $z_0$  under  $f$  has a unique limit point  $\alpha$ , then  $\alpha$  is a fixed point of  $f$ . On the other hand, show that if the orbit has exactly  $p$  limit points  $\{\alpha_1, \alpha_2, \dots, \alpha_p\}$ , then there are cycles of period at most  $p$ .

5. Prove the statements in Remarks 14.2.4. (Hint: For (i) use the fact that  $|f| = \sqrt{u^2 + v^2}$  when  $f = u + iv$  and  $u$  and  $v$  are real functions).

### 14.3 The Dynamics of Polynomials and the Riemann Sphere.

We briefly introduce the reader to the dynamics of complex polynomials. A detailed treatment is beyond the scope of this book as it requires a more advanced knowledge of complex function theory (see for example Devaney [32] or Milnor [91]). The recent survey by Stankewitz and Rolf [120], is a useful resource.

In our first example, we look at the dynamics of  $f(z) = z^2$  on  $\mathbb{C}$ . Restricted to the unit circle  $\mathbb{S}^1$ ,  $f(z)$  is the angle doubling map. Points inside the circle iterate toward 0, and points outside go to  $\infty$ . We see that the complicated dynamics of  $f$  occurs on  $\mathbb{S}^1$ , a set on which  $f$  is chaotic. We call  $\mathbb{S}^1$  the *Julia set* of  $f$ . Generally, the Julia set of  $f : \mathbb{C} \rightarrow \mathbb{C}$  is defined as the closure of the repelling fixed points of  $f$ . We will not use this definition, but give a definition which is easier to work with when dealing with polynomials.

The notion of *Julia set* is named after Gaston Julia, a French mathematician whose prize winning work on the iteration of complex functions was published in 1918 [72]. Before we formally define Julia sets, let us consider some polynomials which will motivate our definition.

**Examples 14.3.1** 1. The quadratic map  $f : \mathbb{C} \rightarrow \mathbb{C}$ ,  $f(z) = z^2$  has two fixed points  $z = 0$  and  $z = 1$ . Since  $f'(z) = 2z$ ,  $z = 0$  is a super-attracting fixed point, whilst  $z = 1$  is repelling. In fact, since  $f^n(z) = z^{2^n}$ , if  $|z_0| < 1$ , then  $f^n(z_0) \rightarrow 0$  as  $n \rightarrow \infty$ , so the basin of attraction of  $z = 0$  is the interior of the unit circle. In addition,  $f$  maps the circle to itself.

To be more specific, if  $z_0 = re^{i\theta}$  where  $r < 1$ , then  $f(z_0) = z_0^2 = r^2 e^{2i\theta}$ ,  $r^2 < r$ . So the action of  $f$  on the interior of  $\mathbb{S}^1$  is a combination of a contraction and a rotation.

If  $|z_0| > 1$ , then  $r > 1$  and we see that the action of  $f$  is now a combination of a rotation and an expansion. In this case,  $|f^n(z_0)| \rightarrow \infty$  as  $n \rightarrow \infty$ , so it is

natural to regard  $\infty$  as being an attracting fixed point of  $f$  (we set  $f(\infty) = \infty$ , so  $f : \mathbb{C} \cup \{\infty\} \rightarrow \mathbb{C} \cup \{\infty\}$ ). The basin of attraction of  $\infty$  is therefore a set whose boundary is the unit circle:

$$B_f(\infty) = \{z \in \mathbb{C} : |z| > 1\}.$$

We saw in Section 6.1 that  $f$  has a countable dense collection of repelling periodic points on the unit circle  $\mathbb{S}^1$ , and  $f$  restricted to  $\mathbb{S}^1$  is chaotic. The closure of the repelling periodic points of  $f$  is  $\mathbb{S}^1$ , the Julia set of  $f$ , to be defined formally in the next section.

2. Let  $f(z) = z^2 - 1$ , a function having fixed points at  $z = (-1 \pm \sqrt{5})/2$ . We can again think of  $z = \infty$  as a fixed point of  $f$ . Set  $h(z) = 1/z$ , an analytic function on its domain, and write

$$g(z) = h^{-1} \circ f \circ h(z) = h^{-1}(f(1/z)) = h^{-1}(1/z^2 - 1) = \frac{1}{1/z^2 - 1} = \frac{z^2}{1 - z^2}.$$

The equation  $h \circ g = f \circ h$ , tells us that  $g$  and  $f$  are conjugate via  $h$ , where  $h$  maps fixed points to fixed points ( $h((1 \pm \sqrt{5})/2) = (-1 \pm \sqrt{5})/2$  and  $h(0) = \infty$ ,  $h(\infty) = 0$ ). We can think of  $h$  as a homeomorphism on the set  $\mathbb{C} \cup \{\infty\}$ , in a manner to be made more precise shortly.

The conjugation of  $f$  and  $g$  tells us that the behavior of  $f$  near  $\infty$  is the same as the behavior of  $g$  near 0. Consequently, we study the map  $f$  defined on the *extended complex plane*:

$$\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\},$$

by requiring  $f(\infty) = \infty$ , and treating  $\infty$  as just another attracting fixed point. This point of view was introduced by Koenigs [78] (see Alexander [1]). Using the function  $h : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ , the derivative of  $f(z)$  at infinity is defined to be the derivative of  $g(z)$  at zero. Now

$$g'(z) = \frac{2z}{(1 - z^2)^2}, \quad \text{so} \quad g'(0) = 0,$$

so  $z = 0$  is a super-attracting fixed point for  $g(z)$ . We conclude that  $z = \infty$  is a super-attracting fixed point for  $f(z)$ .

**Definition 14.3.2** If  $z = \infty$  is a fixed point of  $f : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ , and  $g(z) = h^{-1} \circ f \circ h(z)$  where  $h(z) = 1/z$ , we define the derivative of  $f(z)$  at  $z = \infty$  to be the derivative of  $g(z)$  at  $z = 0$ , if this derivative exists. If  $|g'(0)| < 1$ , then  $z = \infty$  is an attracting fixed

point of  $f$ . If  $|g'(0)| > 1$ ,  $z = \infty$  is a repelling fixed point of  $f$ , and if  $|g'(0)| = 1$ , we say it is *neutral*.

3. We can check that the Newton function  $N_f(z) = (z + 1/z)/2$  of  $f(z) = z^2 - 1$  is conjugate to  $M(z) = 2z/(z^2 + 1)$  via  $h(z) = 1/z$ . Consequently, the fixed point  $z = \infty$  of  $N_f$  has the same behavior as the fixed point  $z = 0$  of  $M$ . Since  $M'(0) = 2$ ,  $z = \infty$  is a repelling fixed point of  $N_f$ . In this case we have the situation where we have a fixed point of  $N_f$  ( $z = \infty$ ), which is not a zero of  $f$ .

4. If  $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$ , ( $a_n \neq 0$ ,  $n \geq 2$ ), is a polynomial, then we can extend  $p(z)$  to  $\widehat{\mathbb{C}}$  by setting  $p(\infty) = \infty$ , a fixed point. Set  $q(z) = h^{-1} \circ p \circ h(z) = 1/p(1/z)$ . Then

$$q(z) = \frac{z^n}{a_0 z^n + a_1 z^{n-1} + \cdots + a_n}.$$

We can check that  $\infty$  is an attracting fixed point for  $p(z)$  since

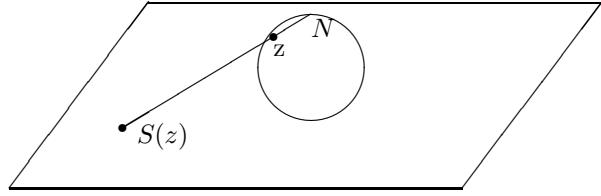
$$|p'(\infty)| = |q'(0)| < 1.$$

A metric can be defined on  $\widehat{\mathbb{C}}$  (see Exercises 14.3), which makes the set  $D = \{z \in \widehat{\mathbb{C}} : |z| > r\}$ , an open ball centered on  $\infty$ . In this case, if  $h(z) = 1/z$ , then  $h(D) = D^*$  where  $D^* = \{z \in \mathbb{C} : |z| < 1/r\}$ , is the corresponding open ball centered on 0. We can then check that  $p$  is a continuous function on  $\widehat{\mathbb{C}}$ , and that  $h : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ , is a homeomorphism.

### 14.3.3 The Riemann Sphere $\widehat{\mathbb{C}}$ .

The set  $\widehat{\mathbb{C}}$  can be identified with the so called *Riemann sphere*  $\mathbb{S}$ , which is defined as follows. Take a sphere  $\mathbb{S}$ , of radius  $1/2$ , sitting on the complex plane  $\mathbb{C}$ , with its south pole at the origin  $(0, 0, 0)$ , and its north pole at  $(0, 0, 1)$ . We are identifying  $z = a + ib \in \mathbb{C}$ ,  $a, b \in \mathbb{R}$ , with the point  $(a, b, 0)$  in 3-space.

From the north pole  $N$ , draw a straight line to the point  $z \in \mathbb{C}$ . Denote by  $S(z)$  the unique point where the line intersects the sphere. The map  $S : \mathbb{C} \rightarrow \mathbb{S} \setminus \{N\}$ , sending  $z$  to  $S(z)$  is a homeomorphism, which extends, on setting  $S(\infty) = N$ , to a homeomorphism of  $\widehat{\mathbb{C}}$  onto all of  $\mathbb{S}$ . The sphere  $\mathbb{S}$ , is called the *Riemann sphere*. Given a function such as  $f : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ ,  $f(z) = z^2$ , there is a corresponding function  $F : \mathbb{S} \rightarrow \mathbb{S}$ , defined by  $F(w) = S \circ f \circ S^{-1}(w)$  for  $w \in \mathbb{S}$ . However, we usually identify  $f$  and  $F$ , using whichever representation is more convenient.

The Riemann Sphere  $\mathbb{S}$ .

To make the above more precise, we should define a metric on both  $\widehat{\mathbb{C}}$  and  $\mathbb{S}$  which makes the map  $S$  a homeomorphism. We leave this to the exercises.

**Examples 14.3.4** 1. We studied the affine maps  $f(z) = az + b$ ,  $a \neq 0$ , as real functions in Chapter 1. In the complex case, we set  $f(\infty) = \infty$ , and then  $f$  becomes a homeomorphism of  $\widehat{\mathbb{C}}$ , with fixed points at  $z_0 = 1/(1-a)$  (when  $a \neq 1$ ), and  $z_1 = \infty$ . The fixed point  $z_0$  is attracting when  $|a| < 1$ , and repelling when  $|a| > 1$  (in the latter case  $\infty$  is attracting). When  $a = 1$ , every  $z \in \widehat{\mathbb{C}}$  is a fixed point. The dynamics is trivial in all cases.

2. Consider the *linear fractional transformation*

$$f(z) = \frac{az + b}{cz + d}, \quad \text{where } ad - bc \neq 0.$$

These maps (also called *Möbius transformations*), can be extended to all of  $\widehat{\mathbb{C}}$ , by setting  $f(\infty) = a/c$  and  $f(-d/c) = \infty$ , (when  $c \neq 0$ ). The inverse of  $f$  is

$$f^{-1}(z) = \frac{-dz + b}{cz - a},$$

again a linear fractional transformation (use the fact that  $ad - bc \neq 0$  to show that  $f$  is one-to-one). In this way, it can be seen that  $f$  is a homeomorphism of  $\widehat{\mathbb{C}}$ . These maps have an interesting, but uncomplicated dynamics which we shall study in the exercises.

### Exercises 14.3

- Show that if  $p(z)$  is a polynomial having degree at least 2, then  $p(\infty) = \infty$ . Use the definition of  $p'(\infty)$  to show that  $|p'(\infty)| < 1$ . Thus  $\infty$  is an attracting fixed point for  $p$  (in fact it is super-attracting). What happens if  $p(z) = az + b$  for some  $a, b \in \mathbb{C}$ ?

2. Let  $p(z) = z^2 - z$ . Show that  $p$  has no points of period 2. (It can be shown that if a polynomial  $q$  of degree at least 2 has no periodic points of period  $n$ , then  $n = 2$ , and  $q$  is conjugate to  $p(z) = z^2 - z$  - see [18]).
3. Let  $p(z)$  be a polynomial of degree  $d \geq 2$ . Show that  $z = \infty$  is a repelling fixed point for the Newton function  $N_p$ , with  $N'_p(\infty) = d/(d-1) > 1$ . What happens if  $p(z) = az + b$  for some  $a, b \in \mathbb{C}$ ?
4. Prove that if  $f(z) = \frac{az+b}{cz+d}$  is a linear fractional transformation with  $(a-d)^2 + 4bc = 0$ , then  $f(z)$  has a unique fixed point. Prove that in this case,  $f(z)$  is conjugate to a translation of the form  $g(z) = z + \alpha$ .
5. (a) Prove that if  $f(z) = \frac{az+b}{cz+d}$  has two fixed points, then  $f(z)$  is conjugate (via a linear fractional transformation), to a linear transformation of the form  $g(z) = \alpha z$ .  
(b) Use the result of (a) to determine a closed form for  $f^n(z)$ .
6. Show that the linear fractional transformation

$$T(z) = \frac{z-1}{z+1},$$

maps the imaginary axis in the complex plane onto the unit circle  $\mathbb{S}^1$ .

7. The Riemann sphere  $\mathbb{S}$  is a sphere of radius  $1/2$ , sitting on the complex plane  $\mathbb{C}$ , so that its south pole  $S$  at  $(0, 0, 0)$  is in contact with the origin  $z = 0$  in  $\mathbb{C}$ , and the North pole  $N$  is at  $(0, 0, 1)$ . If  $P$  is a point on the complex plane with coordinates  $(x, y)$  (so  $z = x + iy$ ), we obtain the corresponding point on the sphere by joining  $N$  to  $P$  with a straight line, and letting the point of intersection with the sphere be  $S(z)$ . Show that the coordinates of  $S(z)$  are

$$S(z) = \left( \frac{\operatorname{Re}(z)}{1 + |z|^2}, \frac{\operatorname{Im}(z)}{1 + |z|^2}, \frac{|z|^2}{1 + |z|^2} \right).$$

8. Define a distance  $\rho$  on the Riemann sphere by  $\rho(w_1, w_2) =$  the shortest distance between the points  $w_1$  and  $w_2$  on the sphere. Now define a distance  $d$  on the extended complex plane  $\widehat{\mathbb{C}}$  by  $d(z_1, z_2) = \rho(S(z_1), S(z_2))$ . Show that  $d$  is a metric on  $\widehat{\mathbb{C}}$  and the set  $D_r = \{z \in \mathbb{C} : |z| > r\}$  is open in  $\widehat{\mathbb{C}}$  for  $r > 0$ . It can be shown that

$$d(z_1, z_2) = \int_{z_1}^{z_2} \frac{1}{1 + |z|^2} dz.$$

9. With respect to the metric defined in the previous problem, show that  $h : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ ,  $h(z) = 1/z$ ,  $h(0) = \infty$ , and  $h(\infty) = 0$ , is a homeomorphism. Show that any polynomial  $p$  is continuous on  $\widehat{\mathbb{C}}$ .

#### 14.4 The Julia Set.

We have mentioned that the Julia set of a complex mapping  $f : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ , is the set on which its interesting dynamics takes place. It will be denoted by  $J(f)$ . In this section, we define  $J(f)$  and determine some of its properties.

For  $f(z) = z^2$ ,  $J(f) = \mathbb{S}^1$ , the unit circle. For  $f(z) = z^2 - 1$ , we will show that it is the interval  $[-2, 2]$ . It can be shown that if  $E(z) = e^z$ , then  $J(E) = \mathbb{C}$  (this was shown in 1981 by Misiurewicz [93], answering an open question of Julia).

We are mainly concerned with quadratic maps of the form:  $f_c(z) = z^2 + c$  for different values of the parameter  $c \in \mathbb{C}$ . Every quadratic polynomial is linear conjugate to a map of this form, so they are representative of all polynomials (see Exercises 7.3).

The Julia set of  $f_0(z) = z^2$  is  $J(f_0) = \mathbb{S}^1$ , a closed set invariant under  $f_0$ , containing all the repelling periodic points, on which  $f_0$  is chaotic. We shall see that this is typical for Julia sets.

Although  $f_0$  exhibits highly chaotic behavior on its Julia set  $\mathbb{S}$ , its study may lead one to believe that Julia sets are nice smooth curves. Nothing could be farther from the truth. A variation of the parameter  $c$  gives rise to quadratic maps whose Julia sets are fractals. In fact, D. Ruelle [110] has shown that for quadratic maps  $f_c(z) = z^2 + c$  with  $|c|$  small, the fractal dimension is approximately

$$d_c = 1 + \frac{|c|^2}{4 \log(2)},$$

and so  $J(f_c)$  is indeed a fractal in these cases. The question of what value the fractal dimension can take for other values of  $c$ , is still an open and difficult question.

An important property of the Julia set of a polynomial  $f$ , is that it is equal to the closure of its repelling periodic points. This fact is often taken as the definition of Julia set, but because we are mainly interested in polynomials, we give a definition that is easier to work with, and from which we can deduce this fact. For functions such as  $e^z$ , it has been shown that the closure of the repelling periodic points is all of  $\mathbb{C}$ .

We have seen that for polynomials of degree at least 2,  $\infty$  is always an attracting fixed point. We use this to define the Julia set of a polynomial:

**Definition 14.4.1** The *basin of attraction* of  $\infty$  for the polynomial  $p(z)$  having degree at least 2, is the open set

$$B_p(\infty) = \{z \in \mathbb{C} : p^n(z) \rightarrow \infty \text{ as } n \rightarrow \infty\}.$$

**Definition 14.4.2** (i) The *Julia set*  $J(p)$  of the polynomial  $p(z)$  having degree at least 2, is the boundary of the open set  $B_p(\infty)$ , i.e., the set  $\overline{B_p(\infty)} \setminus B_p(\infty)$  (see the exercises).

(ii) The *filled-in Julia set*  $K(p)$  is the set  $K(p) = \mathbb{C} \setminus B_p(\infty)$ , of all those points that do not converge to  $\infty$  under iteration by  $p$ . Therefore, the Julia set is also the boundary of the set  $K(p)$ .

(iii) The *Fatou set*  $F(p)$  is the complement of the Julia set:  $F(p) = \mathbb{C} \setminus J(p)$ .

#### 14.4.3 Properties of the Julia Set of a Polynomial.

The following are the main properties of Julia sets of polynomials (these also hold for certain more general functions). We shall prove some of these properties for polynomials  $p : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$  of degree at least 2.

1.  $J(p)$  is a non-empty, closed, bounded and uncountable set.
2. The Julia sets of  $p$  and  $p^r$  for  $r \in \mathbb{Z}^+$ , are identical.
3.  $J(p)$  is *completely invariant* under  $p$ , i.e.,  $p^{-1}(J(p)) = J(p)$ .
4. The repelling periodic points of  $p$  are dense in  $J(p)$ .  $J(p)$  has no isolated points.

5. The Julia set is either *path connected* (there is a continuous curve contained in  $J(p)$ , joining any two points of  $J(p)$ ), or  $J(p)$  is totally disconnected (a Cantor like set called *fractal dust*).
6. If  $\mathcal{A} \subset \widehat{\mathbb{C}}$  is the basin of attraction of an attracting periodic point, then its boundary  $\overline{\mathcal{A}} \setminus \mathcal{A}$  is equal to the entire Julia set.
7. If  $z_0 \in J(p)$ , then the set of iterated pre-images of  $z_0$ :

$$\bigcup_{n=0}^{\infty} p^{-n}(z_0) = \{z \in \mathbb{C} : p^n(z) = z_0, \text{ for some } n \in \mathbb{N}\},$$

is dense in  $J(p)$ .

In the next section, we shall show that both  $J(p)$  and  $K(p)$  are non-empty, closed and bounded sets which are invariant under  $p$ , when  $p$  is a quadratic polynomial. A consequence of property 6 is that any point on the Julia set must lie on the boundary of all basins of attraction for all attracting periodic points of  $p$ . This is hard to imagine when  $p$  has more than two attracting periodic points. In Section 14.6, we will outline a proof of this result for Newton maps arising from polynomials. Property 7 can be used to generate computer graphics of Julia sets: start with  $z_0 \in J(p)$  and compute all  $z_1$  with  $p(z_1) = z_0$  and plot these  $z_1$ 's. Now compute all possible  $z_2$ 's with  $p(z_2) = z_1$ , and plot these. Continue in this way to obtain an approximation of the Julia set. In Appendix B we explain why a complete metric space having no isolated points has to be uncountable. In Chapter 17, we will show that a closed, bounded subset of  $\mathbb{C}$ , being compact, can be regarded as a complete metric space. In this way, one shows that the Julia set  $J(p)$  is uncountable.

It is not hard to show that conjugate maps have homeomorphic Julia sets. In Chapter 17 we will show that homeomorphisms map closed, bounded subsets of  $\mathbb{C}$  (*compact sets*), to closed bounded subsets.

#### 14.4.4 The Quadratic Maps $f_c(z) = z^2 + c$ .

We consider the quadratic maps  $f_c(z) = z^2 + c$ , since these are more easily studied, and any quadratic map is conjugate to such a map. We saw in Section 7.3, that if  $c$  is real, there is a linear conjugacy between  $f_c$  and the logistic map  $L_\mu(z) = \mu z(1 - z)$ , for certain real  $\mu$ . In general we have:

**Proposition 14.4.5**  $f_c : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$  defined by  $f_c(z) = z^2 + c$ , is conjugate to the map  $L_\mu : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ ,  $L_\mu(z) = \mu z(1 - z)$  via the conjugacy  $h : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ ,  $h(z) = -z/\mu + 1/2$ , when  $c = (2\mu - \mu^2)/4$ .

**Proof.** It is straight forward to check that  $h \circ f_c = L_\mu \circ h$ , when  $c = (2\mu - \mu^2)/4$ .  $\square$

If  $c$  is real with  $c < -2$ , we see that  $f_c$  is conjugate to a logistic map  $L_\mu$  with  $\mu > 4$ . It follows from the analysis of Section 5.5 that for  $x \in \mathbb{R}$ ,  $L^n(x) \rightarrow -\infty$  for all  $x$  except for those  $x$  belonging to some Cantor set contained in  $[0, 1]$ . The same conjugacy holds when these functions are treated as functions of a complex variable, and the above discussion indicates that we have complicated behavior of  $f_c$  for  $|c|$  large. In particular, when  $c = -2$ ,  $\mu = 4$ , so  $f_{-2}$  is conjugate to  $L_4$ , a map which is chaotic as a real function on  $[0, 1]$ . We shall see that the Julia set of  $f_{-2}$  is fairly simple.  $J(f_{-2})$  is the interval  $[-2, 2]$  (the image of the interval  $[0, 1]$  under  $h^{-1}$ ).

Note that Proposition 14.4.5 tells us that the map  $L_2(z) = 2z(1 - z)$  is conjugate to  $f_0(z) = z^2$ , the latter map being chaotic on  $\mathbb{S}^1$ . But the dynamics of the real function  $L_2$  are relatively tame. In the real case, the conjugacy is between the dynamical systems  $L_2$  on  $[0, 1]$  and  $f_0$  on  $[-1, 1]$ . The complicated dynamics of  $L_2$  is on the image of the unit circle  $\mathbb{S}^1$  under  $h$  (an ellipse), and this will be the Julia set of  $L_2$ .

We continue with some more general results about polynomials of the form  $F_c(z) = z^d + c$ , where  $c \in \mathbb{C}$  and  $d \in \{2, 3, 4, \dots\}$  is fixed. It can be shown that any polynomial of degree  $d > 1$ , is linearly conjugate to one of the form

$$p(z) = z^d + a_{d-2}z^{d-2} + a_{d-3}z^{d-3} + \dots + a_1z + a_0,$$

(for example, a cubic polynomial is conjugate to a polynomial  $p(z) = z^3 + az + b$  for some  $a, b \in \mathbb{C}$ ). A polynomial of the form  $F_c(z) = z^d + c$  has exactly one critical point,  $z = 0$ ; whereas polynomials of degree greater than two, may have more than one critical point. We conclude that polynomials of the form  $F_c$  do not represent the most general polynomials of degree  $d$  for  $d > 2$ .

The following result is a type of *escape criterion*, to be used to show that the Julia sets of maps of the form  $F_c$ , are both closed, and bounded.

**Proposition 14.4.6** Let  $z \in \mathbb{C}$  with  $|z| > 2$  and  $|z| > |c|$ . If  $d \in \{2, 3, 4, \dots\}$  and  $F_c(z) = z^d + c$ , then  $F_c^n(z) \rightarrow \infty$  as  $n \rightarrow \infty$ .

**Proof.** Suppose the hypothesis of the proposition holds for  $z \in \mathbb{C}$ . Then by the triangle inequality, since  $|z| > |c|$ ,

$$|F_c(z)| = |z^d + c| \geq |z^d| - |c| = |z|^d - |c| > |z|^d - |z| = |z|(|z|^{d-1} - 1).$$

Since  $|z| > 2$ , if  $|z|^{d-1} - 1 = 1 + \alpha$ , then  $\alpha > 0$  and

$$|F_c(z)| > (1 + \alpha)|z|.$$

If we set  $w = F_c(z)$ , then  $|w| > |z|$ , and using the same argument,  $|F_c(w)| > (1 + \alpha')|w|$ , where  $\alpha' = |w|^{d-1} - 2$ , so that  $|F_c^2(z)| > (1 + \alpha')(1 + \alpha)|z| > (1 + \alpha)^2|z|$  (where it is easily checked that  $\alpha' > \alpha$ ). An induction argument now shows that  $|F_c^n(z)| > (1 + \alpha)^n|z|$ , for all  $n \in \mathbb{Z}^+$ , so that  $F_c^n(z) \rightarrow \infty$  as  $n \rightarrow \infty$ .

□

**Theorem 14.4.7** *The filled-in Julia set  $K(F_c) = \mathbb{C} \setminus B_{F_c}(\infty)$ , of  $F_c(z) = z^d + c$  ( $d \in \{2, 3, 4, \dots\}$ ), is a non-empty, closed, bounded and  $F_c$ -invariant set.*

**Proof.** Set  $A = \{z \in \mathbb{C} : |z| \leq r_c\}$ , where  $r_c = \max\{|c|, 2\}$ . Then  $A$  is a closed, bounded set ( $A = \bar{B}_{r_c}(0)$ , the closed ball of radius  $r_c$  centered on the origin). We can check that  $F_c^{-1}A \subseteq A$ , for if not, there exists  $w \in F_c^{-1}A \setminus A$ , so  $F_c(w) \in A$  with  $w \notin A$ . Thus  $|w| > r_c$ , and by the proof of Proposition 14.4.6,

$$|F_c(w)| > |w| > r_c,$$

contradicting the fact that  $F_c(w) \in A$ . Continuing this line of argument inductively, we see that

$$A \supseteq F_c^{-1}A \supseteq F_c^{-2}A \supseteq \cdots \supseteq F_c^{-n}A \cdots .$$

We now claim that the filled-in Julia set is

$$K(F_c) = \bigcap_{n=0}^{\infty} F_c^{-n}A.$$

Let  $z \in K(F_c)$ . Then if there exists  $n \in \mathbb{Z}^+$  with  $z \notin F_c^{-n}A$ , then  $F_c^n(z) \notin A$ . Therefore  $F_c^k(F_c^n(z)) \rightarrow \infty$  as  $k \rightarrow \infty$ , contradicting  $z$  being in  $K(F_c)$ .

On the other hand, suppose that  $z \in \bigcap_{n=0}^{\infty} F_c^{-n}A$ . Then  $F_c^n(z) \in A$  for all  $n \in \mathbb{Z}^+$ , so  $|F_c^n(z)| < r_c$  for all  $n \in \mathbb{Z}^+$ , and  $z \in K(F_c)$ . This proves the result.

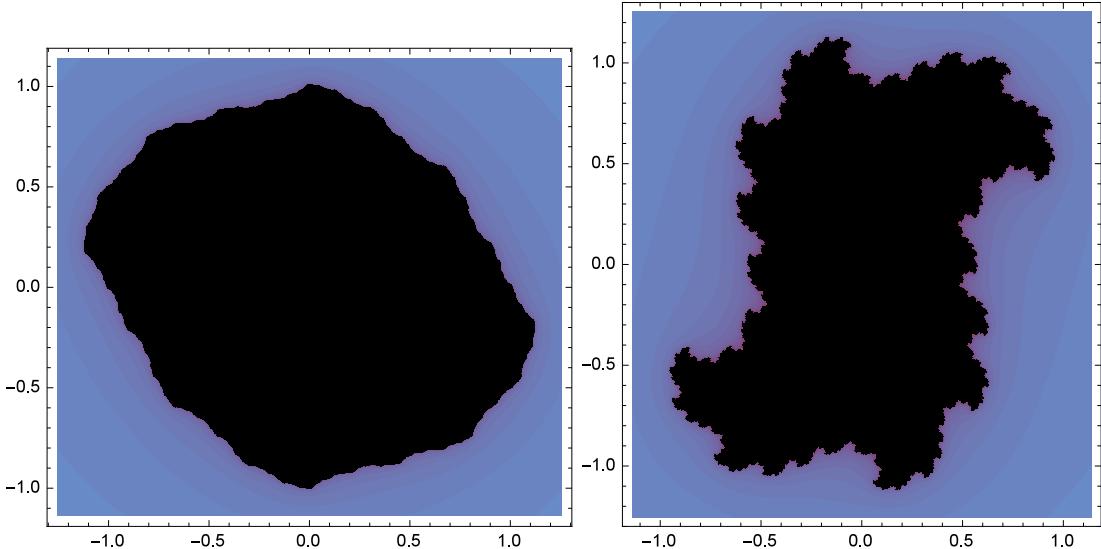
Now  $K(F_c)$  is non-empty since it contains the fixed points of  $F_c$  (these are the solutions of the equation  $z^d + c = z$ , which by the Fundamental Theorem of Algebra always exist). The iterates of these points under  $F_c$  never change, so they must lie in  $K(F_c)$ .

The continuity of  $F_c$  implies that each of the sets  $F_c^{-n}A$  is closed, and non-empty, so their intersection is closed. Clearly  $K(F_c) \subseteq A$ , so  $K(F_c)$  is a bounded set. Finally,

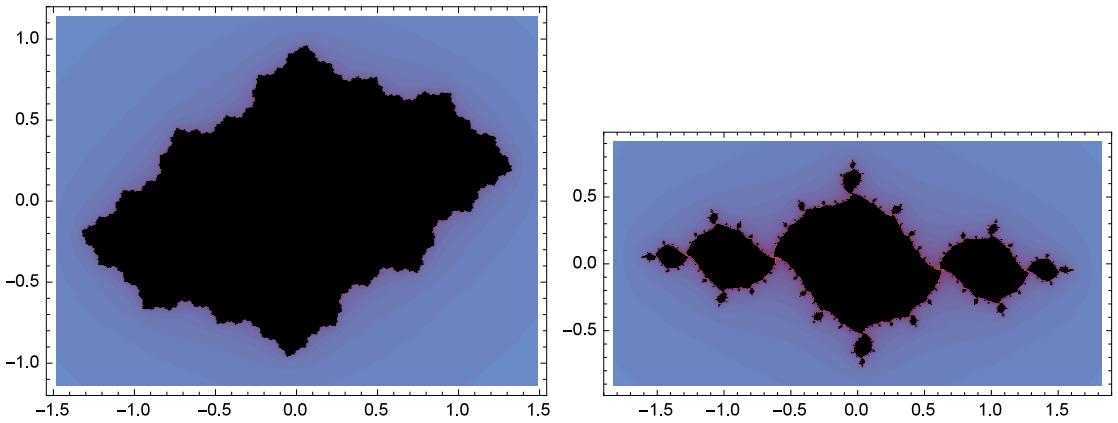
$$K(F_c) = A \cap F_c^{-1}A \cap F_c^{-2} \cap \dots, \quad \text{so} \quad F_c^{-1}K(F_c) = F_c^{-1}A \cap F_c^{-2} \cap \dots \supseteq K(F_c),$$

but since  $F_c^{-1}A \subseteq A$ , we must have equality, i.e.,  $F_c^{-1}(K(F_c)) = K(F_c)$ .  $\square$

**Remark 14.4.8** A sequence of sets with the property:  $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$ , is said to be a *nested sequence*. We will show in the Chapter 17 that a nested sequence of non-empty, closed and bounded sets in a complete metric space, has a non-empty intersection.



Julia sets when  $c = -.1 + .2i$  and  $c = .3 - .3i$ .



Julia sets when  $c = .4 - .3i$  and  $c = -1 + .1i$ .

**Corollary 14.4.9** *The Julia set  $J(F_c)$  of  $F_c(z) = z^d + c$  is a non-empty, closed, bounded, and  $F_c$ -invariant set.*

**Proof.**  $K(F_c)$  is a closed set, the basin of attraction of  $\infty$  is open, and clearly invariant under  $F_c$ . It follows that  $J(F_c) = K(F_c) \cap \overline{B}_{F_c}(\infty)$  is a closed set, being the intersection of closed sets.  $J(F_c)$  is bounded and non-empty, since any non-empty open set in  $\mathbb{C}$  will have non-empty boundary. Furthermore,

$$F_c^{-1}J(F_c) = F_c^{-1}K(F_c) \cap F_c^{-1}(\overline{B}_{F_c}(\infty)) = J(F_c),$$

because the basin of attraction of  $\infty$  is an invariant set.  $\square$

**Example 14.4.10** Consider the quadratic map  $f_{-2}(z) = z^2 - 2$ . The orbit of 0 is the set  $O(0) = \{0, -2, 2\}$  ( $z = 2$  is a fixed point which is repelling). This is a bounded set, so 0, -2 and 2 lie in the filled-in Julia set. In addition to  $f_0$ ,  $f_{-2}$  is one of the few maps of the form  $f_c$  whose Julia set can be explicitly determined in a simple way. We will show that

$$K(f_{-2}) = J(f_{-2}) = [-2, 2].$$

Set  $D^* = \{z \in \mathbb{C} : |z| > 1\} = B_{f_0}(\infty)$ , the basin of attraction of infinity for  $f_0$ . Define

$$h : D^* \rightarrow \mathbb{C}, \quad \text{by} \quad h(z) = z + \frac{1}{z}.$$

Then we claim that  $h$  is a homeomorphism onto the set  $\mathbb{C} \setminus [-2, 2]$ .

**Claim 1.**  $h$  is one-to-one, because if  $h(z) = h(w)$ , then

$$z + \frac{1}{z} = w + \frac{1}{w} \Rightarrow zw(z - w) = z - w,$$

so if  $z \neq w$ , then  $zw = 1$ . If  $|z| > 1$ , then  $|w| = 1/|z| < 1$ , contradicting  $z, w \in D^*$ , therefore  $h$  is one-to-one on  $D^*$ .

**Claim 2.**  $h(D^*) = \mathbb{C} \setminus [-2, 2]$ . To see this, suppose  $w \in \mathbb{C}$  with  $h(z) = w$ , then

$$z^2 - zw + 1 = 0, \quad \text{so} \quad z = \frac{w \pm \sqrt{w^2 - 4}}{2}.$$

If  $z_1$  and  $z_2$  are the two solutions,  $z_1 z_2 = 1$ , so either  $z_1$  lies in  $D^*$ , or  $z_2$  lies in  $D^*$ , or they both belong to  $\mathbb{S}^1$ .

In the latter case,  $h(z_1) = h(z_2) \in [-2, 2]$  (since  $z + 1/z = z + \bar{z} = 2\operatorname{Re}(z)$  for  $z \in \mathbb{S}^1$ ). In the former case, there exists  $z \in D^*$  with  $f(z) = w$ . It follows that  $h$  is onto.

It is easy to see that  $h : D^* \rightarrow \mathbb{C} \setminus [-2, 2]$  is also continuous, and its inverse is continuous, so  $h$  is a homeomorphism.

We can now check that  $h \circ f_0(z) = f_{-2} \circ h(z)$  for all  $z \in \widehat{\mathbb{C}}$ . It follows that  $f_0$  and  $f_{-2}$  are conjugate on their respective domains.

Note that the map

$$h : \mathbb{S}^1 \rightarrow [-2, 2], \quad h(z) = z + \frac{1}{z},$$

is a two-to-one onto map (except at  $\pm 1 \in \mathbb{S}^1$ ), so that  $f_{-2}$  restricted to  $[-2, 2]$  is a factor of  $f_0$  restricted to  $\mathbb{S}^1$  (see Section 7.1). Now,  $f_0$  being chaotic on  $\mathbb{S}^1$  implies that  $f_{-2}$  is chaotic on  $[-2, 2]$ , with  $f_{-2}^n(z)$  bounded for  $z \in [-2, 2]$ .

**Claim 3.**  $K(f_{-2}) = J(f_{-2}) = [-2, 2]$ . We know that the basin of attraction of  $\infty$  for  $f_0$  is  $D^*$ , so putting the above information together, the basin of attraction of  $\infty$  for  $f_{-2}$  must be  $\mathbb{C} \setminus [-2, 2]$ . It follows that the Julia set of  $f_{-2}$  is  $[-2, 2]$ .

#### 14.4.11 Maps for Which the Sequence $(F_c^n(0))$ is Unbounded.

Suppose that  $|c| > 2$ . Then  $F_c^n(0) \rightarrow \infty$  as  $n \rightarrow \infty$  (see Exercises 14.5). We consider the structure of the Julia set in this case. If  $A = \overline{B}_{r_c}(0)$ , where  $r_c = \max\{2, |c|\} = |c|$ , then the proof of Theorem 14.4.7 shows that the filled-in Julia set is

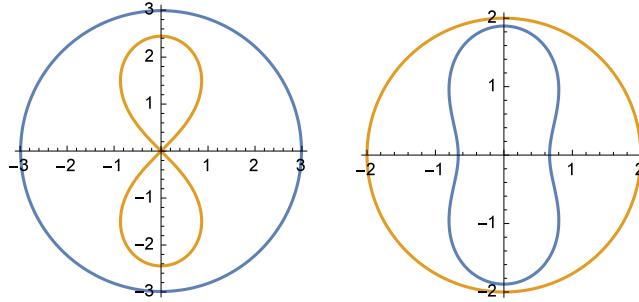
$$K(F_c) = \bigcap_{j=0}^{\infty} F_c^{-j}(A),$$

and

$$A \supseteq F_c^{-1}A \supseteq F_c^{-2}A \supseteq \cdots \supseteq F_c^{-n}A \cdots .$$

Set  $A_0 = A$  and  $A_i = F_c^{-i}(A)$ . Then just as in the construction of the Cantor set, the sets  $A_0, A_1, A_2, \dots$  give a nested sequence of sets which converge (in the Hausdorff metric), to the filled-in Julia set of  $F_c$ . The first iteration  $A_1 = F_c^{-1}A$  is a bounded set inside  $\overline{B}_{r_c}(0)$ , containing 0 (since  $F_c(0) = c$ , and  $c$  lies on the boundary of  $B_{r_c}(0)$ ). The boundary of  $A_1$  is a figure eight curve which self intersects at the origin. If  $I_0$  and  $I_1$  are the two “lobes” of this curve, it can be shown that  $A_2$  has boundary consisting of two figure eight curves, one contained in  $I_0$  and the other contained in  $I_1$ . Continuing in this way, the resulting filled-in Julia set (which is the same as the Julia set of  $F_c$ ), consists of the intersection of these figure-eights, a totally disconnected set

- a type of Cantor set. The above discussion motivates the following results, whose proofs are omitted.



$$F_c^{-1}(A) \text{ for } c > 2 \text{ and } c < 2.$$

**Theorem 14.4.12** (a) *If  $0 \notin K(F_c)$ , then the Julia set of  $F_c$  is totally disconnected, fractal dust.*

(b) *If  $0 \in K(F_c)$ , then the Julia set of  $F_c$  is pathwise connected.*

(c)  *$F_c$  restricted to its Julia set is chaotic.*

The orbits of critical points play an important role in the study of the iteration of complex maps. In the next section, we will examine the importance of the boundedness of the orbit of the critical point  $z = 0$ , for the quadratic maps  $f_c(z) = z^2 + c$ . First, we show that the Julia set contains all of the repelling periodic points of  $f_c$ . In fact, as we previously mentioned, the Julia set of  $f$  is the closure of its set of repelling periodic points.

We will need the following theorem (a form of Cauchy's Theorem), which is a standard result in a first course in complex analysis. We state it for a polynomial  $p(z)$ , although it holds more generally. It tells us that the value of the derivative  $p'(z_0)$  is controlled by the values of  $p$  in a disk centered at  $z_0$ . We will omit the proof.

**Theorem 14.4.13** *Let  $p(z)$  be a complex polynomial with  $|p(z)| \leq M$  for all  $z$  in the closed ball  $\overline{B}_r(z_0) = \{z \in \mathbb{C} : |z - z_0| \leq r\}$ . Then*

$$|p'(z_0)| < \frac{M}{r}.$$

We use this to prove:

**Theorem 14.4.14** Let  $z_0$  be a repelling periodic point for  $f_c(z) = z^2 + c$ . Then  $z_0 \in J(f_c)$ .

**Proof.** Suppose that  $z_0 \notin J(f_c)$ , where  $z_0$  is a repelling periodic point having period  $n$ . Then we have

$$|(f_c^n)'(z_0)| = \lambda > 1.$$

Since  $z_0$  is a periodic point, the orbit of  $z_0$  is finite, so  $z_0 \in K(f_c) \setminus J(f_c)$ , i.e., it belongs to the interior of  $K(f_c)$ . It follows that there is an open ball

$$B_r(z_0) \subseteq K(f_c),$$

and no point within this ball goes to  $\infty$  under iteration by  $f_c$ .

For each  $k \in \mathbb{Z}^+$ ,  $f_c^{kn}$  is a polynomial, and for each  $z \in \overline{B}_r(z_0)$ , it follows from Proposition 14.4.6 that we have

$$|f_c^{kn}(z)| \leq \max\{|c|, 2\}.$$

Let  $M = \max\{|c|, 2\}$ . Then by Theorem 14.4.12, we must have for each  $k \in \mathbb{Z}^+$ ,

$$|(f_c^{kn})'(z_0)| < \frac{M}{r}.$$

But (see Exercises 14.4 # 4),

$$|(f_c^{kn})'(z_0)| = \lambda^k \rightarrow \infty \quad \text{as } k \rightarrow \infty,$$

a contradiction. □

### Exercises 14.4

1. Let  $K_c$  be the filled-in Julia set for  $f_c(z) = z^2 + c$ . Show that  $2 \in K_{-6}$ , and find another point in  $K_{-6}$ . Find some points in  $K_{-1+3i}$ .
2. Let  $f_i(z) = z^2 + i$ . Find the period two points of  $f_i$ , and determine their stability. Prove that 0 is an eventually periodic point of  $f_i$ , and describe the filled-in Julia set of  $f_i$ .
3. Show that  $z = 0$  is eventually periodic for  $f_i(z) = z^2 + i$ . Show that  $z = -i$  is a repelling periodic point for  $f_i$ , and hence belongs to the Julia set  $J(f_i)$ .

4. If  $z_0$  is a period  $n$ -point for  $f$ , with  $z_l = f^l(z_0)$ ,  $l = 0, 1, \dots, n-1$ , show that  $(f^n)'(z_l)$  is independent of  $l$ . Deduce that if  $|(f^n)'(z_0)| = \lambda$ , then  $|(f^{kn})'(z_0)| = \lambda^k$  for  $k \in \mathbb{Z}$ .
5. Show that  $f_{-2}$  is a factor of  $f_0$  on their Julia sets, via  $h(z) = z + 1/z$ , i.e.,  $h : \mathbb{S}^1 \rightarrow [-2, 2]$  is onto and  $h \circ f_0 = f_{-2} \circ h$ . Is  $h$  one-to-one on these sets? (Assume the Julia sets of  $f_0$  and  $f_{-2}$  are  $\mathbb{S}^1$  and  $[-2, 2]$  respectively).
6. Show that  $K(f_c) \neq J(f_c)$  whenever  $f_c$  has a bounded attracting periodic orbit. Is the converse true? (Hint: For the converse, consider an  $n$ -periodic point  $p$  with  $|(f^n)'(p)| = 1$ ).
7. Let  $f : X \rightarrow X$  be a map with  $C \subset X$ .
- Prove that if both  $f(C) \subseteq C$  and  $f^{-1}(C) \subseteq C$ , then  $f^{-1}(C) = C$ , i.e.,  $C$  is completely invariant under  $f$ .
  - Show that if  $f$  is an onto map with  $f^{-1}(C) = C$ , then  $f(C) = C$ .
8. (a) We have seen that any quadratic polynomial  $p(z)$  is linearly conjugate to one of the form  $q(z) = z^2 + c$ . Use this to show that  $p(z) = z^2 - z$  is conjugate to  $q(z) = z^2 - 3/4$ .
- (b) Let  $p(z) = z^2 + c$ . Explain why  $p(z) - z$  divides  $p^2(z) - z$ . Use this to show that if  $p$  has no points of period 2, then  $p(z) = z^2 - 3/4$ .
- (c) Prove that any cubic polynomial  $p(z)$  is linearly conjugate to one of the form  $q(z) = z^3 + az + b$ .
9. If we define the Julia set of a polynomial to be the closure of the repelling periodic points, what are the possibilities for the Julia set of a degree one polynomial?

10. If  $f, g : \mathbb{C} \rightarrow \mathbb{C}$  are conjugate polynomials (via a conjugacy  $h$ ), and  $J(f)$  is the Julia set of  $f$ , show that  $h(J(f))$  is the Julia set of  $g$ . (Hint: Use the fact that the Julia set of  $f$  is the closure of the repelling periodic points of  $f$ ).
11. Show that if  $p(z)$  is a polynomial with degree  $d > 1$ , then  $J(p^k) = J(p)$ , for  $k \in \mathbb{Z}^+$ .
12. Let  $f_c(z) = z^2 + c$  with  $c > 2$  real and  $A = \overline{B}_{r_c}(0)$ , where  $r_c = \max\{2, |c|\} = c$ . Show that the boundary of the region  $f_c^{-1}(A)$  has polar equation  $r^2 = -2c \cos(2\theta)$ . It is instructive to use a computer algebra system to graph this region for various values of  $c$ .
13. Let  $d \in \{2, 3, \dots\}$  and  $c \in \mathbb{C}$ . Show that if  $F_c(z) = z^d + c$ , then the basin of attraction of  $\infty$ ,  $B_c(\infty) = \{z \in \mathbb{C} : F_c^n(z) \rightarrow \infty, \text{ as } n \rightarrow \infty\}$  is an open set. (Hint: Let  $z_0 \in B_c(\infty)$ , so that  $|F_c^k(z_0)| > \max\{|c|, 2\} + 1$  for some  $k > 0$ . Use the continuity of  $F_c^k$  to show that there exists  $\delta > 0$  for which  $|F_c^k(z) - F_c^k(z_0)| < 1$  when  $|z - z_0| < \delta$ . Now use the triangle inequality and the escape criterion to deduce that  $F_c^n(z) \rightarrow \infty$  for all  $z \in B_\delta(z_0)$ ).
14. It can be shown that the Fatou set of a rational function has either 0, 1, 2 or  $\infty$  connected components.
- Deduce that the Fatou set of  $f_{-1}(z) = z^2 - 1$  has infinitely many connected components. (Hint: Consider the attracting fixed points  $\{0, -1, \infty\}$  of  $f_{-1}^2$ ).
  - If  $F_0$  and  $F_{-1}$  are the components of the Fatou set of  $f_{-1}$ , containing 0 and  $-1$  respectively, show that  $f_{-1}(F_0) = F_{-1}$  and  $f_{-1}(F_{-1}) = F_0$ .
15. Show that if  $p$  is a polynomial, then the basin of attraction of infinity,  $B_p(\infty)$ , is a non-empty open set, invariant under  $p$ . Show that the boundary of an open set in  $\mathbb{C}$  is non-empty, and deduce that the Julia set of  $p$  is non-empty.

### 14.5 The Mandelbrot Set $\mathcal{M}$ .

The *Mandelbrot set*  $\mathcal{M}$ , is a compact subset of the complex plane having a distinctive fractal type boundary. The first crude pictures of  $\mathcal{M}$ , were given by Robert Brooks and Peter Matelski in 1978. In 1980, Benoit Mandelbrot graphed this set and studied its fractal nature. It was Mandelbrot who is responsible for making the Mandelbrot set  $\mathcal{M}$  widely popular. The first major mathematical study of  $\mathcal{M}$ , showing that  $\mathcal{M}$  is a connected set, was given by Douady and Hubbard in 1982, and they named the set for Mandelbrot. In 1994, Shishikura showed that the boundary of the Mandelbrot set is a fractal having fractal dimension equal to 2. This is the largest value the fractal dimension of a curve can possibly take, reinforcing the idea that the Mandelbrot set is a highly complex mathematical object.

Computer graphics for the Julia set of quadratic maps  $f_c(z) = z^2 + c$ , and those for the Mandelbrot set, are obtained by iteration of a similar type, but with a different emphasis.

If  $z_0$  belongs to the basin of attraction of  $\infty$  of  $f_c$ , iterating  $z_0$  will give  $f_c^n(z_0) \rightarrow \infty$ , as  $n \rightarrow \infty$ . On the other hand, if we look at  $f_c^{-n}(z_0)$  and plot these points for large positive values of  $n$ , an approximation to the Julia set will appear (the Julia set being the boundary of the basin of attraction of  $\infty$ ). This idea can be used to graph Julia sets with a computer algebra system, but it takes many iterations to fill the entire Julia set, as points tend to cluster around certain regions. By varying  $c$ , the types of Julia sets arising vary considerably, some being in one piece (connected), others consisting of many disjoint (totally disconnected) sets - a type of Cantor set referred to as *fractal dust*.

On the other hand, the Mandelbrot set is obtained by fixing  $z_0 = 0$  and looking at the behavior of  $f_c^n(0)$  as  $n \rightarrow \infty$  for different values of  $c$ . For example, suppose that  $c = 1$  ( $f_1(z) = z^2 + 1$ ), then

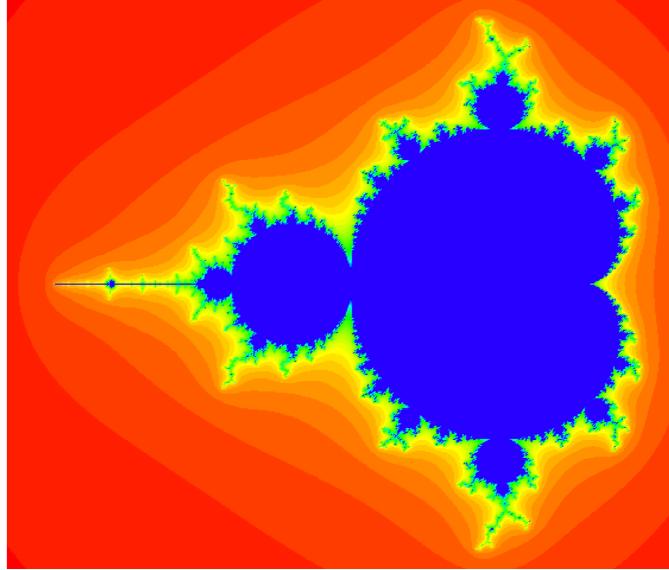
$$\{f_1^n(0) : n \in \mathbb{Z}\} = \{0, 1, 2, 5, 26, \dots\},$$

an unbounded set. If we color those points  $c \in \mathbb{C}$  white for which  $f_c^n(0) \rightarrow \infty$  and those points  $c$  black for which the set  $\{f_c^n(0) : n \in \mathbb{Z}\}$  remains bounded, then the resulting (black), set is the Mandelbrot set  $\mathcal{M}$ . It follows that  $1 \notin \mathcal{M}$ . If  $c = -1$  ( $f_{-1}(z) = z^2 - 1$ ), then

$$\{f_{-1}^n(0) : n \in \mathbb{Z}\} = \{0, -1, 0, -1, \dots\} = \{-1, 0\},$$

a bounded set, so  $-1 \in \mathcal{M}$ .

We shall define the Mandelbrot set  $\mathcal{M}$  to be the set of all complex numbers  $c$  for which the orbit of 0 is bounded under iteration by  $f_c$ . It turns out that this is the same set as those  $c$ 's for which  $K(f_c)$  (the filled in Julia set), is a connected set.



The Mandelbrot Set

Although the Mandelbrot set does not have the same type of linear self similarity as the Koch snowflake, it has a boundary that is fractal with fractal dimension equal to 2. The Mandelbrot set  $\mathcal{M}$  is a set with many incredible properties. It is a connected set which contains an infinite number of small copies of itself. If  $c$  lies in the interior of the main body of  $M$ , the corresponding Julia set  $J(f_c)$  is a fractally deformed circle surrounding a unique attracting fixed point. If  $c$  lies in the interior of one of the buds, the Julia set of  $f_c$  consists of infinitely many fractally deformed circles connected to each other, each surrounding an attracting periodic point. Other possibilities arise by taking  $c$  on the boundary of  $\mathcal{M}$ . In particular, as  $c$  passes through the boundary to the outside of the Mandelbrot set there is a dramatic change in the corresponding Julia sets for  $f_c$ . They decompose into a cloud of infinitely many points (the fractal dust). The proofs of many of these facts are beyond the scope of this text, but we shall prove some of the more important properties. More information can be found in [32] and [91].

Related to Proposition 14.4.12, we state without proof the following important properties of the maps  $f_c$ .

**14.5.1 The Fundamental Dichotomy.** Let  $f_c(z) = z^2 + c$ . Then either:

- (i) The orbit of the point  $z = 0$  goes to  $\infty$  under iteration by  $f_c$ . In this case  $K(f_c)$  consists of infinitely many disjoint components, or
- (ii) The orbit of 0 remains bounded ( $0 \in K(f_c)$ ). In this case  $K(f_c)$  is a connected set.

**Definition 14.5.2** The Mandelbrot set  $\mathcal{M}$  is defined to be

$$\mathcal{M} = \{c \in \mathbb{C} : \text{the orbit of } 0 \text{ is bounded under iteration by } f_c\}$$

From the Fundamental Dichotomy,  $\mathcal{M}$  is the set of those  $c \in \mathbb{C}$  for which the filled-in Julia set of  $f_c$  is a connected set.

### 14.5.3 Properties of the Mandelbrot Set.

Let  $f_c(z) = z^2 + c$ , then the Mandelbrot set can be defined as:

$$\mathcal{M} = \{c \in \mathbb{C} : \exists r > 0, \forall n \in \mathbb{N}, |f_c^n(0)| \leq r\}.$$

If  $f_c(z) = z$ , then  $z^2 - z + c = 0$ , so the fixed points are

$$z_1 = \frac{1 - \sqrt{1 - 4c}}{2}, \quad \text{and} \quad z_2 = \frac{1 + \sqrt{1 - 4c}}{2}.$$

Denote by  $z$  either  $z_1$  or  $z_2$ . Then in order for  $z$  to be attracting, we require  $|f'_c(z)| = |2z| < 1$ . It follows from a complex analog of Singer's Theorem due to P. Fatou ([43]), that if  $z_0$  is an attracting periodic point of a complex map  $f$ , then there is at least one critical point in the basin of attraction of  $z_0$  (in other words, there is a critical point which becomes arbitrarily close to the points of the attracting cycle, under iteration). Since the maps  $f_c$  have exactly one critical point,  $z = 0$ , they can have at most one attracting cycle. We deduce that at most one of  $z_1$  or  $z_2$  is an attracting fixed point. We state without proof, Fatou's result in the case of complex polynomials (see [91] where the theorem is given for rational maps).

**Theorem 14.5.4** (P. Fatou, 1918.) *The immediate basin of attraction of an attracting periodic point of a polynomial  $p(z)$ , contains a critical point (a point  $z_0$  where  $p'(z_0) = 0$ ).*

**Proposition 14.5.5** (a)  $\mathcal{M} \subseteq \overline{B}_2(0) = \{z \in \mathbb{C} : |z| \leq 2\}$ .

(b)  $c \in \mathcal{M}$  if and only if  $|f_c^n(0)| \leq 2$  for all  $n \in \mathbb{N}$ .

(c)  $\mathcal{M} \cap \mathbb{R} = [-2, 1/4]$ .

(d)  $\mathcal{M}$  is a connected, closed and bounded subset of  $\mathbb{C}$ .

**Proof.** (a) This is part of Exercises 14.5, 1(b).

(b) If  $|f_c^n(0)| \leq 2$  for all  $n \in \mathbb{N}$ , then the orbit of 0 is bounded under  $f_c$ , so  $c \in \mathcal{M}$ .  
For the converse, use Exercise 14.5, 1(a).

(c) We look at various cases.

**Case 1,  $c < -2$ .** We use induction to show that

$$f_c^n(0) \geq 2 + n|c + 2|, \quad \text{for } n \geq 2.$$

$f_c^2(0) = c^2 + c$ , and  $c^2 + c - 2 = (c+2)(c-1) = |c-1||c+2| = (-c+1)|c+2| \geq 2|c+2|$ , so the inequality holds for  $n = 2$ . Now  $f_c^{n+1}(0) = [f_c^n(0)]^2 + c \geq [2 + n|c + 2|]^2 + c$ , by the induction hypothesis. We need to show that the latter expression is at least  $2 + (n+1)|c + 2|$ . We have

$$[2 + n|c + 2|]^2 + c - [2 + (n+1)|c + 2|] = -(c+2)(n^2|c+2| + 3n - 2),$$

which is non-negative for  $c \leq -2$  and  $n \geq 2$ . It is now clear that  $f_c^n(0) \rightarrow \infty$  as  $n \rightarrow \infty$ .

**Case 2,  $-2 \leq c < 0$ .**  $z_1$  and  $z_2$  are the fixed points of  $f_c$ . We can check that for  $-2 \leq c < 0$ , we have  $z_2 > f_c^2(0) = c^2 + c$ . In this case,  $0 \in [z_1, z_2]$  and  $f_c^2[z_1, z_2] = f_c[c, z_2] \subseteq [c, z_2]$ . It follows that the orbit of 0 is bounded, so  $c \in \mathcal{M}$ .

**Case 3,  $0 \leq c \leq 1/4$ .**  $f_c[0, z_2] = [c, z_2] \subseteq [0, z_2]$ , so the orbit of 0 is always bounded, and  $c \in \mathcal{M}$ .

**Case 4,  $c > 1/4$ .** In this case we use induction to show that  $f_c^n(0) \geq n(c - 1/4)$  for  $n \geq 0$ . This is clearly true when  $n = 0$ . Suppose the statement is true for  $n$ . Then

$$f_c^{n+1}(0) = f_c(f_c^n(0)) = [f_c^n(0)]^2 + c \geq [n(c - 1/4)]^2 + c,$$

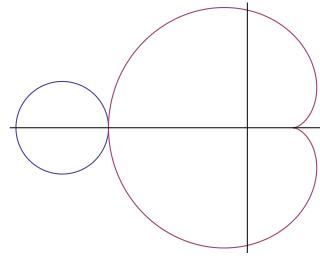
and we need to check that this is at least  $(n+1)(c - 1/4)$ . However, it can be seen that

$$\begin{aligned} [n(c - 1/4)]^2 + c - (n+1)(c - 1/4) &= n^2c^2 - c(n^2/2 + n) + n^2/16 + n/4 + 1/4 \\ &= (nc - n/4 - 1/2)^2 \geq 0, \end{aligned}$$

and the inequality follows. In particular,  $f_c^n(0) \rightarrow \infty$  as  $n \rightarrow \infty$ , so  $c \notin \mathcal{M}$  for  $c > 1/4$ .

(d) The proof of the connectedness is a famous result due to Douady and Hubbard [37], which is beyond the scope of this text. The boundedness of  $\mathcal{M}$  follows from part (a). We omit the proof that  $\mathcal{M}$  is closed.  $\square$

If we look at the computer graphic of  $\mathcal{M}$ , we see that there is a *primary bulb* whose boundary is a cardioid that lies between  $-3/4$  and  $1/4$ . If  $c$  belongs to the interior of this cardioid, then the Julia set of  $f_c$  will be a deformed circle (when  $c = 0$  the Julia set is a circle). The Julia set  $J(f_c)$  becomes more complicated as  $c$  approaches the boundary of the cardioid. Attached to the cardioid is a disc, whose boundary is a circle with center  $(-1, 0)$ , radius  $1/4$ . We verify this as follows.



The Primary Cardioid and the Circle  $|z + 1| = 1/4$ .

**Proposition 14.5.6** (a) *The primary cardioid of  $\mathcal{M}$  is the boundary of the set*

$$\mathcal{A} = \{c \in \mathbb{C} : f_c \text{ has an attracting fixed point}\}.$$

*This set is contained in  $\mathcal{M}$ , and its boundary is the parameterized curve  $c = e^{it}/2 - e^{2it}/4$ , or equivalently,  $x = \frac{1}{2} \cos t - \frac{1}{4} \cos 2t$ ,  $y = \frac{1}{2} \sin t - \frac{1}{4} \sin 2t$ ,  $t \in [0, 2\pi]$ , where  $c = x + iy$ .*

(b) *The interior of the disc attached to the main cardioid is the set*

$$\{c \in \mathbb{C} : f_c \text{ has an attracting cycle of period 2}\}.$$

*The boundary of this disc has equation  $|z + 1| = 1/4$ , the circle with center  $(-1, 0)$  and radius  $1/4$ .*

**Proof.** (a) First note that if  $c \in \mathcal{A}$ , then  $f_c$  has an attracting fixed point  $z_0$ , and so by Fatou's Theorem (14.5.4),  $z_0$  must attract the critical point  $z = 0$ . It follows that  $c \in \mathcal{M}$ . Since  $\mathcal{M}$  is a closed set, it must also contain the boundary of  $\mathcal{A}$ .

Without loss of generality, we may assume that  $z_1$  is an attracting fixed point of  $f_c$ . Then  $z_1$  satisfies the equation  $z^2 - z + c = 0$ , and  $|f_c(z_1)| < 1$ , so  $|z_1| < 1/2$ . We see that the set of those  $c$ 's for which  $f_c$  has an attracting fixed point is

$$\{c \in \mathbb{C} : z^2 - z + c = 0, \text{ and } |z| < 1/2\} = \{c \in \mathbb{C} : \left| \frac{1 - \sqrt{1 - 4c}}{2} \right| < 1/2\}.$$

This is an open set whose boundary is the set of  $c$ 's for which

$$1 - \sqrt{1 - 4c} = e^{it},$$

for some  $t \in [0, 2\pi]$ . Solving for  $c$  gives

$$c = e^{it}/2 - e^{2it}/4,$$

a parameterization of the boundary curve. Rewriting in real form, we have

$$x = 1/2 \cos t - 1/4 \cos 2t, \quad y = 1/2 \sin t - 1/4 \sin 2t, \quad \text{where } c = x + iy, \quad t \in [0, 2\pi],$$

a cardioid, which intersects the real axis when  $z = 1/4$ , and when  $z = -3/4$ , as expected.

(b) If  $f^2(z) = z$ , then  $z^4 + 2cz^2 - z + c^2 + c = 0$ . Dividing by  $z^2 - z + c$  in order to disregard the fixed points, gives  $z^2 + z + c + 1 = 0$ . Denoting the roots by  $z_1$  and  $z_2$ , the requirement for them to be attracting is

$$|f'(z_1)f'(z_2)| = |2z_12z_2| = |(-1 + \sqrt{-3 - 4c})(-1 + \sqrt{-3 - 4c})| = |4 + 4c| < 1.$$

Therefore  $|c + 1| < 1/4$ , which has as boundary, the circle with center  $(-1, 0)$ , and radius  $1/4$ .

□

It can be shown that each bulb of the Mandelbrot set corresponds to  $c$  values for which  $f_c$  has an attracting  $k$ -cycle for some  $k \in \mathbb{Z}^+$ . The primary bulb corresponds to those  $c$ 's for which  $f_c$  has an attracting fixed point. The interior of the circle  $|z + 1| = 1/4$  corresponds to those  $c$ 's for which  $f_c$  has an attracting 2-cycle. Each bulb has infinitely many smaller bulbs attached, with each bulb representing those  $c$ 's with a particular attracting  $k$ -cycle. Since any attracting cycle must attract a critical point, and  $f_c$  has only one critical point, for any value of  $c$  there is at most one attracting cycle. See [41] for more details.

Points on the boundaries of the bulbs do not give rise to attracting cycles. For example,  $c = 1/4$  is on the boundary of the primary bulb, and there is a unique fixed point  $z = 1/2$  for  $f_{1/4}$  with  $f'_{1/4}(1/2) = 1$ , so we have a neutral fixed point.

### 14.5.7 Computer Graphics for the Mandelbrot Set.

The ideas of this section, together with a computer algebra system, can be used to graph the Mandelbrot set. Let  $f_c(z) = z^2 + c$ , then we look at its critical point  $z = 0$  under iteration by  $f_c$ . We know that  $\mathcal{M}$  is contained in a circle centered on the origin and of radius 2. Fix  $c \in [-2, 2] \times [-2, 2]$ , (the square centered on the origin with side length 4). Compute the iterates

$$f_c(0), f_c^2(0), \dots, f_c^n(0),$$

for some fixed, predetermined, suitably large value of  $n$ .

If the iterates eventually leave the square, color  $c$  white, if they do not leave the square, color  $c$  black. Continue this process for a suitably fine selection of points in the square (this will determine the resolution of  $\mathcal{M}$ ). The resulting black shaded set is the Mandelbrot set.

### Exercises 14.5

1. We have seen that if  $|z| > 2$  and  $|z| \geq |c|$ , then  $f_c^n(z) \rightarrow \infty$  as  $n \rightarrow \infty$ , where  $f_c(z) = z^2 + c$ .

(a) Suppose  $|c| > 2$ . Show that  $f_c^n(0) \rightarrow \infty$  as  $n \rightarrow \infty$ .

(b) Deduce that the Mandelbrot set  $\mathcal{M}$ , has the property that

$$\mathcal{M} \subseteq \{z \in \mathbb{C} : |z| \leq 2\}.$$

(Note: We have seen that  $-2 \in M$  because the sequence  $(f_{-2}^n(0))$  is bounded).

(c) Let  $F_c(z) = z^3 + c$ . Prove that if  $|z| > \max\{|c|, \sqrt[3]{2}\}$ , then  $F_c^n(z) \rightarrow \infty$  as  $n \rightarrow \infty$   
 (Hint: Modify the proof of Theorem 14.5.6).

2. Consider the orbit of 0 under  $f_c(z) = z^2 + c$ , so  $f^2(0) = c^2 + c$ . If  $c^2 + c$  lies on a 3-cycle, show that  $c$  satisfies the equation

$$c^3(c+2)(c^3 + 2c^2 + c + 1)^2(c^6 + 2c^5 + 2c^4 + 2c^3 + c^2 + 1) = 0,$$

and explain the significance of the zero's of the different factors. Why must all these  $c$ -values belong to the Mandelbrot set?

3. Show that the Mandelbrot set is symmetric with respect to the real axis.
4. Show that the points  $c \in \mathbb{C}$  that lie on the boundary of the primary bulb (cardioid), give rise to those maps  $f_c(z) = z^2 + c$  having a neutral fixed point (a fixed point  $z_0$  with  $|f'(z_0)| = 1$ ). (Hint: Examine the proof of Proposition 14.5.6).

## 14.6 Newton's Method in the Complex Plane for Quadratics and Cubics.

Given a complex function  $f : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$  with derivative  $f'(z)$ , the Newton function  $N_f(z) = z - f(z)/f'(z)$  defines a new function on  $\widehat{\mathbb{C}}$  whose iterations give rise to the zeros of  $f$ . In Chapter 11, we looked at Newton's method in detail for real quadratic polynomials, and also for certain real cubic polynomials. We saw that the case of cubics is quite a bit more complicated than that for quadratics. In this section, we examine the quadratic case in detail for complex functions, and we see that the cubic case is considerably more complicated.

One important difference for the complex case as compared to the real case, is that for  $f(z)$  a polynomial of degree  $n$ ,  $f$  has exactly  $n$  (possibly repeated) roots, so in particular,  $f$  always has a root. Also, because we are looking at  $f$  as a function on the Riemann sphere, it may happen that  $\infty$  is a fixed point of  $N_f$ , without it being a zero of  $f$ .

We start this section by showing that for a quadratic polynomial having two distinct roots, the basins of attraction of the corresponding fixed points of the Newton function are given by two open half planes, divided by the perpendicular bisector of the line joining the two roots. This is not particularly surprising - what is surprising is the nature of the basins of attraction for the corresponding situation for the roots of cubic polynomials. We first prove a result about the linear conjugacy of Newton's functions for quadratic maps having distinct roots. A real version of this result was mentioned in Chapter 11.

**Theorem 14.6.1** *Let  $f(z) = az^2 + bz + c$  be a quadratic polynomial having two distinct roots. The Newton functions  $N_f$  and  $N_q$  are linearly conjugate, when  $q(z) = z^2 - \alpha$ , and  $\alpha = b^2 - 4ac$ .*

**Proof.** Since  $f(z)$  has two distinct roots,  $b^2 - 4ac \neq 0$ . A calculation shows that

$$N_f(z) = \frac{az^2 - c}{2az + b} \quad \text{and} \quad N_q(z) = \frac{z}{2} + \frac{\alpha}{2z}.$$

Set  $h(z) = 2az + b$ . Then we can check that  $h \circ N_f(z) = N_q \circ h(z)$ , using  $\alpha = b^2 - 4ac$ .

□

We now prove the Schröder-Cayley Theorem concerning the basins of attraction of a quadratic having distinct roots. Another proof using Halley's method, a generalization of Newton's method, is given in Section 14.7. The latter proof is more complicated, but of some historical interest.

Theorem 14.6.1 tells us that it suffices to consider polynomials of the form  $q(z) = z^2 - \alpha$ , where  $\alpha \neq 0$ .

**Theorem 14.6.2** (Schröder [113], 1872, Cayley [27], 1882). *Let  $f(z) = az^2 + bz + c$  be a complex quadratic function having two distinct roots  $\alpha_1$  and  $\alpha_2$ . Join  $\alpha_1$  and  $\alpha_2$  by a straight line, and denote the perpendicular bisector of this line by  $L$ . The basin of attraction of the fixed points  $\alpha_1$  and  $\alpha_2$  of the Newton function  $N_f$  consists of all those points in the same open half plane determined by the line  $L$ .  $N_f$  is chaotic on  $L$ .*

**Proof.** We may assume that  $f$  is a polynomial of the form  $f_a(z) = z^2 - a^2$ , ( $a \neq 0$ ), with Newton function  $N_a(z) = (z + a^2/z)/2$ .

It is straightforward to check that  $N_a$  is conjugate to the map  $f_0(z) = z^2$  on the extended complex plane  $\widehat{\mathbb{C}}$ , via a conjugacy  $T$ :

$$T \circ N_a = f_0 \circ T, \quad T(z) = \frac{z - a}{z + a}.$$

$T$  is the linear fractional transformation which maps the attracting fixed points  $\{-a, a\}$  of  $N_a$  to the attracting fixed points  $\{\infty, 0\}$  of  $f_0$ , and the repelling fixed point  $\infty$  of  $N_a$  to the repelling fixed point  $1$  of  $f_0$ . In other words,  $T^{-1}(0) = a$  and  $T^{-1}(\infty) = -a$ . It suffices to show that  $T^{-1}$  maps the open unit disc to one half-plane, and maps the open set exterior to the unit circle to the other half-plane. The unit circle is mapped onto the perpendicular bisector between the roots.

It is easier to do this in the case that  $a = 1$ , as the perpendicular bisector between the roots is simply the imaginary axis. We will restrict the proof to this case (which gives the basic idea).

When  $a = 1$ , we have  $T^{-1}(z) = (1 + z)/(1 - z)$ , and  $T^{-1}(0) = 1$ . It suffices to show that the basin of attraction of  $z = 0$  for  $f_0$  (the interior of the unit disc), is mapped onto the open right half-plane, and similarly for the basin of attraction of  $\infty$  for  $f_0$  being mapped to the open left half-plane.

Set  $z = re^{i\theta}$ , where  $r \geq 0$ . We can check that

$$T^{-1}(z) = \frac{1 - r^2 + 2ir \sin \theta}{1 + r^2 - 2r \cos \theta}.$$

If  $r = 1$ , this is purely imaginary ( $T^{-1}$  can be seen to map  $\mathbb{S}^1$  onto the imaginary axis). If  $r < 1$ ,  $\operatorname{Re}(T^{-1}(z)) > 0$ , so the interior of the unit disc is mapped onto the the open half-plane  $\{z \in \mathbb{C} : \operatorname{Re}(z) > 0\}$ . The basin of attraction of  $\infty$  for  $f_0$  is dealt with in a similar way.

Finally,  $N_f$  is chaotic on the imaginary axis since  $f_0(z) = z^2$  is chaotic on  $S^1$ , and the conjugacy  $T^{-1}$  sends chaotic sets to chaotic sets.

□

**Example 14.6.3** 1. The above proof shows that if  $f(z) = z^2 - 1$ , and we iterate  $z_0$  under  $N_f$  for points with  $\operatorname{Re}(z_0) > 0$ , the iterates go to 1, whilst those with  $\operatorname{Re}(z_0) < 0$  go to  $-1$ . The point is, the dynamics of  $N_f$  are trivial for these values of  $z_0$ . The interesting dynamics of the map occurs on the imaginary axis, where the map is chaotic. The imaginary axis is the *Julia set* of the map. We leave as an exercise, that for a quadratic polynomial having a single repeated root  $\alpha$ , the basin of attraction of the root under the Newton function is all of  $\mathbb{C}$ .

#### 14.6.4 Cubic Polynomials.

Cayley and his contemporaries were perplexed by the dynamics of Newton's method for cubic polynomials in the complex setting. We have seen in Chapter 11 that Newton's method for real cubics can give rise to extremely complicated dynamical behavior. Fatou and Julia were able to gain some understanding of the basins of attraction of the fixed points of the Newton map, but their results indicated that the boundaries between the different basins had very strange properties. This was not surprising since we will see from computer graphics, that the basins of attraction of the fixed points of  $N_p$  for  $p$  a cubic having three distinct roots, is unimaginably complex.

Consider possibly the simplest cubic polynomial having three distinct roots:

$$p(z) = z^3 - 1.$$

Solving  $p(z) = 0$  gives  $(z - 1)(z^2 + z + 1) = 0$ , so  $z = 1$ , or  $z = (-1 \pm i\sqrt{3})/2$ . Set  $\omega = (-1 + i\sqrt{3})/2$ , then  $\bar{\omega} = \omega^2 = (-1 - i\sqrt{3})/2$ , giving  $\{1, \omega, \omega^2\}$ , (where  $\omega = e^{2\pi i/3}$ ), known as the *cube roots of unity*. These roots lie on the unit circle  $\mathbb{S}^1$ , at the vertices of an equilateral triangle.

The Newton function is

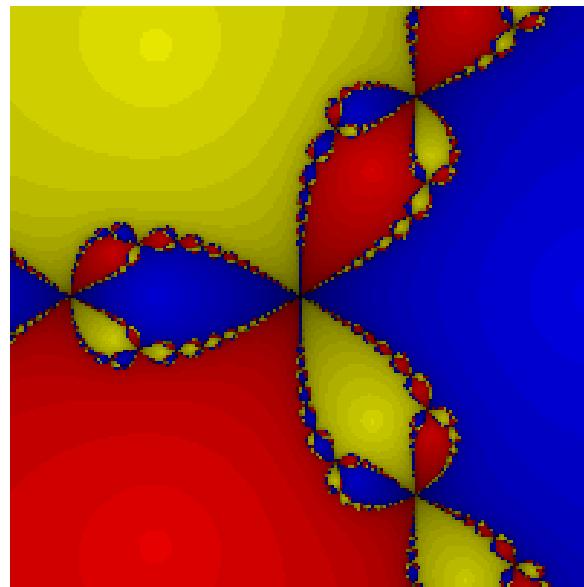
$$N_p(z) = \frac{2z^3 + 1}{3z^2}, \quad \text{and} \quad N_p(1) = 1, \quad N_p(\omega) = \omega, \quad N_p(\omega^2) = \omega^2.$$

Our theory tells us that each of the roots of  $p(z)$  is a super-attracting fixed point of multiplicity two, for  $N_p$  (as is easily checked), and there is an open ball surrounding these points, contained in their basin of attraction. The point at  $\infty$  may be regarded as a repelling fixed point with  $N_p^{-1}(\infty) = \{0, \infty\}$ .

The basins of attraction of the fixed points of  $N_p$  have a type of rotational symmetry: rotating  $B_{N_p}(1)$  through 120 degrees, gives  $B_{N_p}(\omega)$ , and a further rotation gives  $B_{N_p}(\omega^2)$  (see Exercise 14.6 # 5). These basins of attraction are also symmetric with respect to the real axis, and  $B_{N_p}(1)$  contains the ray  $\theta = 0$  (the positive real axis, excluding  $z = 0$ ), and similarly  $B_{N_p}(\omega)$  contains the ray  $\theta = 2\pi/3$ , and  $B_{N_p}(\omega^2)$  contains  $\theta = 4\pi/3$ .

The Newton map  $N_p$ , can be regarded as a continuous function on  $\widehat{\mathbb{C}}$ , so these basins of attraction are open sets, to be denoted by  $\mathcal{A}_i$ ,  $i = 1, 2, 3$ . The complement,  $\widehat{\mathbb{C}} \setminus \cup_{i=1}^3 \mathcal{A}_i$  is a closed set, which we regard as the Julia set  $J(N_p)$ , of  $N_p$ . We see that  $0, \infty \in J(N_p)$  (see the Exercises 14.6 # 2 to see that  $\infty$  is a repelling fixed point of  $N_p$ ). In particular, since  $J(N_p)$  is an invariant set, all points of the form  $N_p^{-n}(0)$ ,  $n \in \mathbb{Z}^+$  lie in  $J(N_p)$ . It can be shown that the set of all such points is dense in  $J(N_p)$ , and we shall use this fact to describe some surprising properties of this Julia set. Note that these arguments can be generalized to Newton functions arising from  $p(z) = z^n - 1$ , for  $n \in \mathbb{Z}^+, n > 2$ .

This discussion may lead one to think that the different basins of attraction are bounded by the rays  $\theta = \pm\pi/3$  and  $\theta = \pi$ , (the medians of the equilateral triangle). When we use computer graphics to plot the basins of attraction, we see that the truth is far stranger. We now outline an explanation for the different behavior of Newton's algorithm in the cases of quadratics and cubics.



The basins of attraction using Newton's method on  $f(z) = z^3 - 1$ .

#### 14.6.5 The Basin of Attraction of the Newton Function of a Polynomial Having at Least Three Distinct Roots.

In this section, we try to explain how we know that the Newton function of a cubic has a very complicated basin of attraction. Let  $p(z)$  be a polynomial having  $n \geq 3$ , distinct roots,  $z_1, z_2, \dots, z_n$ . Each of these roots is a super-attracting fixed point of  $N_p$ . If  $\mathcal{A}_i$ ,  $i = 1, 2, \dots, n$ , are the corresponding basins of attractions, there is an open ball centered on  $z_j$ , contained in  $\mathcal{A}_j$ , for  $j = 1, 2, \dots, n$ . The following result indicates the complexity of these basins of attraction: it is impossible to picture them precisely, and they cannot be connected sets. These basins of attraction have to be open sets, so their union,  $\cup_{k=1}^n \mathcal{A}_k$  will be open, with closed complement. This complement constitutes the Julia set  $J(N_p)$  of  $N_p$ , being the boundary between the different regions  $\mathcal{A}_k$ . Theorem 14.6.6 tells us that any point  $z_0 \in J(N_p)$  is arbitrarily close to points from  $\mathcal{A}_k$ , for  $k = 1, 2, \dots, n$ . In other words, any point in the Julia set of  $N_p$  lies on the boundary of all three sets  $\mathcal{A}_k$ , for  $k = 1, 2, \dots, n$ , a situation that is hard to imagine. This fact was known to both Fatou and Julia.

**Theorem 14.6.6** Suppose that  $z_0$  lies in the Julia set  $J(N_p)$  of  $N_p$ . If  $\epsilon > 0$ , the ball  $B_\epsilon(z_0)$  contains points of  $\mathcal{A}_k$ , for  $k = 1, 2, \dots, n$ .

**Outline of Proof.** We give the proof for the polynomial  $p(z) = z^3 - 1$ . There are three basins of attraction:  $\mathcal{A}_i$ ,  $i = 1, 2, 3$ .  $z_0 \in J(N_p)$ . Since the set  $\{N_p^{-n}(0) : n \in \mathbb{Z}^+\}$  is dense in  $J(N_p)$ , some iterate of  $B_\epsilon(z_0)$ , say  $N_p^n(B_\epsilon(z_0))$ , contains 0, and hence intersects the rays  $\theta = 0, \theta = 2\pi/3$  and  $\theta = 4\pi/3$ . In particular, this set contains points of  $\mathcal{A}_i$ ,  $i = 1, 2, 3$ . It follows that the ball  $B_\epsilon(z_0)$  contains points of  $\mathcal{A}_i$ ,  $i = 1, 2, 3$ .  $\square$

#### 14.6.7 The cubic $p_\alpha(z) = z(z - 1)(z - \alpha)$ .

It can be shown (see the exercises and [17]), that if  $p(z)$  is a cubic polynomial having three distinct roots, then  $N_p$  is linearly conjugate to  $N_{p_\alpha}$ , where  $p_\alpha(z) = z(z - 1)(z - \alpha)$ , for some  $\alpha$  belonging to the set

$$A = \{z \in \mathbb{C} : \operatorname{Im}(z) \geq 0, |z| \leq 1, |z - 1| \leq 0\}.$$

For the cubic  $p(z)$ , we have seen that  $N'_p(z) = p(z)p''(z)/[p'(z)]^2$ , so there are (potentially), four critical points for  $N_p$ : the three corresponding to the roots of  $p(z) = 0$ , and one coming from the (inflection) point, where  $p''(z) = 0$ . Fatou's Theorem (which is applicable to rational functions on  $\widehat{\mathbb{C}}$ ), tells us that for each attracting periodic orbit, there is a critical point which is attracted to it. In particular, there are at most four attracting periodic orbits for  $N_p$ . It is natural to ask, if  $N_p$  can have exactly four such orbits, three corresponding to the roots of  $p(z)$ , and one corresponding to the inflection point. This is clearly not the case for the Newton function arising from  $f(z) = z^3 - 1$ , since  $N_f$  only has 3 critical points. However, it can indeed happen for cubics of the form  $p_\alpha(z) = z(z - 1)(z - \alpha)$ . In contrast to the real case, for such a cubic,  $N_{p_\alpha}$  can have an attracting period point of period greater than one, with basin of attraction an open set in  $\mathbb{C}$  (so has non-zero two-dimensional measure). Consequently, there is a set of points of positive measure, which when iterated under  $N_{p_\alpha}$ , do not converge to a root of  $p_\alpha$ . For a general polynomial  $p(z)$ , the Fatou set of  $N_p$  (the complement of the Julia set), may contain basins of attraction that are non-empty open sets, and which do not correspond to one of the roots.

For  $p_\alpha(z) = z(z - 1)(z - \alpha)$ , Blanchard ([17]), has investigated the set of parameter values

$$\{\alpha \in \mathbb{C} : N_{p_\alpha}^n((\alpha + 1)/3)) \text{ does not converge to one of the roots }\},$$

$((1 + \alpha)/3$  is a critical point of  $N_{p_\alpha}$ , see Exercise 14.6 # 6(c)), and it is remarkable to see that computer graphics of this set gives rise to images of the Mandelbrot set.

### Exercises 14.6

1. Complete the proof of Theorem 14.6.1, by showing that if  $f(z) = az^2 + bz + c$  and  $q(z) = z^2 - \alpha$ , then the Newton functions  $N_f$  and  $N_q$  are conjugate when  $\alpha = b^2 - 4ac$ .
  
  
  
2. Let  $p(z)$  be a polynomial of degree  $d > 1$  with Newton function
$$N_p(z) = z - \frac{p(z)}{p'(z)}.$$

(a) Show that  $N'_p(z) = \frac{p(z)p''(z)}{(p'(z))^2}$ .

(b) Suppose that  $p(z) = z^m q(z)$  where  $q(0) \neq 0$ . Check that  $N_p(0) = 0$ , so that  $z = 0$  is a fixed point. Show that  $N'_p(0) = (m-1)/m$ , so  $z = 0$  is an attracting fixed point which is super-attracting only when  $m = 1$ .

(c) Let  $p(z)$  be a polynomial of degree  $d > 1$  with Newton function  $N_p$ . Show that  $z = \infty$  is a repelling fixed point for  $N_p$  with  $N'_p(\infty) = d/(d-1) > 1$ .

(d) Check that (b) and (c) above holds for  $N_p$ , where  $p(z) = z^3 - z^2$ .
  
  
  
3. (a) Show that the maps  $f_0(z) = z^2$ , and the Newton function  $N_\alpha$  where  $f_1(z) = z^2 - \alpha^2$ , are conjugate (Hint: Use the linear fractional transformation  $T(z) = \frac{z-\alpha}{z+\alpha}$ ).

(b) In the case that  $\alpha = i$ , deduce that the Julia set of  $N_{f_1}$  is the real axis, and the basin of attraction of the fixed point  $i$ , is the upper half plane, and that of  $-i$  is the lower half plane. Now set  $\alpha = 1$ , and confirm that the Julia set will be the imaginary axis.

  
  
  
4. In this question we outline the *relaxed Newton's method* (see [55]), which improves the speed of convergence when a function has a multiple root. Suppose that  $g(z) = (z - \alpha)^m h(z)$ , where  $h(\alpha) \neq 0$  and  $m \in \mathbb{Z}^+, m > 1$ .

- (a) Apply Newton's method to  $(g(z))^{1/m}$ , and show that the Newton function obtained is

$$N_m(z) = z - \frac{mg(z)}{g'(z)}.$$

- (b) Show that  $z = \alpha$  is a super-attracting fixed point of  $N_m$  (i.e.,  $N_m(\alpha) = \alpha$  and  $N'_m(\alpha) = 0$ ).

- (c) Show that the relaxed Newton method  $N_2$ , for the cubic with double root  $g(z) = (z - a)^2(z - b)$ , is conjugate to the quadratic  $p(z) = z^2 - 3/4$ . (Hint: Show that  $p \circ h = h \circ N_2$ , where  $h(z) = \frac{3z+a-4b}{2(z-a)}$ .)

5. Let  $N_p$  be the Newton function of the polynomial  $p(z) = z^3 - 1$ .

- (a) Show that the basins of attraction of the fixed points of  $N_p$ , are invariant under rotation through 120 degrees:

$$B_{N_p}(\omega) = \omega B_{N_p}(1), \quad \text{and} \quad B_{N_p}(\omega^2) = \omega^2 B_{N_p}(1).$$

- (b) Show that the basins of attraction are invariant under reflection in the real axis.

- (c) Show that the ray  $\mathbb{R}^+ = \{z \in \mathbb{C} : \operatorname{Im}(z) = 0, \operatorname{Re}(z) > 0\}$ , is entirely contained in  $B_{N_p}(1)$ . Use (a) to deduce corresponding results for the other basins of attraction. (Hint: For the first part, it suffices to show that the real function  $p(x) = x^3 - 1$ ,  $x \in \mathbb{R}$ , with the Newton function  $N_p$  has the fixed point  $x = 1$ , having basin of attraction containing  $\mathbb{R}^+$ ).

- (d) Note that the above results hold for  $p(z) = z^n - 1$ ,  $n \geq 2$ , with obvious modifications.

6. Let  $p(z)$  be a polynomial of degree  $d > 1$  with Newton function  $N_p(z) = z - \frac{p(z)}{p'(z)}$ .

- (a) If  $p(\alpha) = 0$  and  $p'(\alpha) \neq 0$ , show that  $\alpha$  is a fixed point of multiplicity two for  $N_p$  (there is a rational function  $k(z) = m(z)/n(z)$  with  $n(\alpha) \neq 0$  and  $N_p(z) - \alpha = (z - \alpha)^2 k(z)$ ).

- (b) If  $p(\alpha) = 0$ ,  $p'(\alpha) \neq 0$ , and  $p''(\alpha) = 0$ , show that  $\alpha$  is a fixed point of multiplicity three for  $N_p$ .
- (c) Show that for  $p_\alpha(z) = z(z-1)(z-\alpha)$ ,  $N_{p_\alpha}$  has a critical point where  $z = (\alpha+1)/3$ .
- (d) For what values of  $\alpha$  does  $p_\alpha$  satisfy the requirements of (b) above?

7\*. Show that if  $p(z)$  is a cubic polynomial having three distinct roots, then  $N_p$  is linearly conjugate to  $N_{p_\alpha}$ , where  $p_\alpha(z) = z(z-1)(z-\alpha)$ , for some  $\alpha$  belonging to the set

$$A = \{z \in \mathbb{C} : \operatorname{Im}(z) \geq 0, |z| \leq 1, |z-1| \leq 0\}.$$

## 14.7 Important Complex Functions.

In this section, we will define some of the elementary complex functions  $f : \mathbb{C} \rightarrow \mathbb{C}$  needed for Schröder's proof of the Schröder-Cayley Theorem. This section may be omitted, as it includes topics not usually included in a first course in dynamical systems. We shall not pursue the dynamics of transcendental functions (see [32], or [91] for more information about the dynamics of the exponential and trigonometric functions).

**14.7.1 The Exponential Function:  $e^z$ .** Let  $z \in \mathbb{C}$ ,  $z = x + iy$ , where  $x$  and  $y$  are real. The *complex exponential function* is defined to be:

$$e^z = e^x(\cos(y) + i \sin(y)).$$

**Proposition 14.7.2** *The complex exponential function has the following properties for  $z, w \in \mathbb{C}$ :*

- (i)  $e^{z+w} = e^z e^w$ ,
- (ii)  $(e^z)^n = e^{nz}$  if  $n$  is a positive integer,
- (iii)  $e^{2k\pi i} = 1$ , for all  $k \in \mathbb{Z}$ .

Equations such as  $(e^z)^w = e^{zw}$  are not generally true for arbitrary  $z, w \in \mathbb{C}$ .

**Definition 14.7.3** The complex *sine* and *cosine* functions  $\sin(z)$  and  $\cos(z)$  are defined for  $z \in \mathbb{C}$  by:

$$\sin(z) = \frac{1}{2i}(e^{iz} - e^{-iz}) \quad \text{and} \quad \cos(z) = \frac{1}{2}(e^{iz} + e^{-iz}).$$

The *tangent* function  $\tan(z)$  is defined to be

$$\tan(z) = \frac{\sin(z)}{\cos(z)}, \quad z \neq \pi/2 + k\pi, \quad k \in \mathbb{Z}.$$

The exponential function, and the functions sine and cosine are said to be *entire functions*, because it can be shown that they are analytic throughout the complex plane  $\mathbb{C}$ .

**Proposition 14.7.4** *The trigonometric functions have the following properties for  $z, w \in \mathbb{C}$ :*

(i)  $\sin(z + w) = \sin(z)\cos(w) + \cos(z)\sin(w)$ ,

(ii)  $\cos(z + w) = \cos(z)\cos(w) - \sin(z)\sin(w)$

(iii)  $\tan(z + w) = \frac{\tan(z) + \tan(w)}{1 - \tan(z)\tan(w)}$ ,  $\tan(2z) = \frac{2\tan(z)}{1 - \tan^2(z)}$ .

(iv)  $\sin(z)$  and  $\cos(z)$  are periodic of period  $2\pi$ , and  $\tan(z)$  is periodic of period  $\pi$  (so for example  $\tan(z + \pi) = \tan(z)$ , for all  $z \neq \pi/2 + k\pi$ ,  $k \in \mathbb{Z}$ ).

#### 14.7.5 The Complex Logarithm Function: $\text{Log}(z)$ .

The *complex logarithm* is usually regarded as a multi-valued function. For our purposes, we look at the logarithm function as a bijective function when the domain and range are restricted suitably. This is what is often called the *principal branch of the logarithm function*. We start with the exponential function  $f : \mathbb{C} \rightarrow \mathbb{C}$ ,  $f(z) = e^z$ , and we see that it is never zero, and its range is  $\mathbb{C} \setminus \{0\}$ .  $f$  is also not one-to-one since  $e^{z+2k\pi i} = e^z$  for all  $k \in \mathbb{Z}$ , and all  $z \in \mathbb{C}$ , so the inverse of  $f$  is not defined. However, if we restrict  $f$  to the infinite horizontal strip, where  $-\pi < \text{Im}(z) \leq \pi$ , then  $f$  will be one-to-one, and the range remains unchanged. In particular, the function (also denoted by  $f$ ):

$$f : S \rightarrow \mathbb{C} \setminus \{0\}, \quad f(z) = e^z,$$

is both one-to-one and onto, so has an inverse function  $g : \mathbb{C} \setminus \{0\} \rightarrow S$ , where

$$S = \{z \in \mathbb{C} : z = x + iy, x, y \in \mathbb{R}, -\pi < y \leq \pi\}.$$

Our aim is to define and give some basic properties of the principal branch of the logarithm function.

**Definition 14.7.6** The function  $g : \mathbb{C} \setminus \{0\} \rightarrow S$ , which is the inverse of  $f(z) = e^z$ , is written  $g(z) = \text{Log}(z)$ , and called the (*principal branch of the*) *logarithm function*.

For  $z \in S$  and  $w \in \mathbb{C} \setminus \{0\}$ , we have:

$$w = e^z \text{ if and only if } z = \text{Log}(w).$$

In addition, since  $f$  and  $g$  are inverse to each other,  $g(f(z)) = z$  for all  $z \in S$ , and  $f(g(z)) = z$  for all  $z \in \mathbb{C} \setminus \{0\}$ , so:

$$\text{Log}(e^z) = z, \forall z \in S, \text{ and } e^{\text{Log}(z)} = z, \forall z \in \mathbb{C} \setminus \{0\}.$$

Let  $w = e^z$ , where  $z = x + iy \in S$  and  $w = \rho e^{i\phi} \neq 0$ ,  $\rho > 0$ ,  $-\pi < \phi \leq \pi$ . Then

$$\rho e^{i\phi} = e^x e^{iy}, \text{ so } \rho = e^x, \text{ and } \phi = y.$$

It follows that

$$z = x + iy = \ln(\rho) + i\phi = \ln|w| + i\text{Arg}(w).$$

We have shown:

**Theorem 14.7.7** Let  $S = \{z \in \mathbb{C} : z = x + iy, x, y \in \mathbb{R}, -\pi < y \leq \pi\}$ . The logarithm function  $g : \mathbb{C} \setminus \{0\} \rightarrow S$ ,  $g(z) = \text{Log}(z)$  satisfies

$$\text{Log}(z) = \ln|z| + i\text{Arg}(z), \quad z \in \mathbb{C} \setminus \{0\},$$

where  $\text{Arg}(z)$  is the principal value of the argument of  $z$  ( $-\pi < \text{Arg}(z) \leq \pi$ ).

**Remarks 14.7.8** 1.  $g(z) = \text{Log}(z)$  is a bijective function, continuous, and analytic everywhere except on the negative real axis.

2. In general,  $\text{Log}(z_1 z_2) \neq \text{Log}(z_1) + \text{Log}(z_2)$  because the sum of the arguments may not lie in  $(-\pi, \pi]$ . This equation holds when it is treated as a set identity for the multivalued function  $\text{log}(z)$ , where  $\text{log}(z) = \{\text{Log}(z) + 2n\pi i : n \in \mathbb{Z}\}$ . For example,  $\text{Log}(i) = \ln|i| + i\pi/2 = i\pi/2$ ,  $\text{Log}(-1) = \ln|-1| + i\pi = i\pi$ , and  $\text{Log}(-i) = -i\pi/2$ , whereas  $\text{Log}(-1) + \text{Log}(i) = 3i\pi/2$ .

#### 14.7.9 The Complex Arctangent Function: $\text{Arctan}(z)$ .

To define the *arctangent function*, we start with the function  $f(z) = \tan(z)$ , and give an argument analogous to the one we gave for the logarithm function. The tangent function is not defined when  $\cos(z) = 0$ , or  $e^{2iz} + 1 = 0$ ,  $z = \pi/2 + k\pi$ ,

$k \in \mathbb{Z}$ . Also, the equation  $w = \tan(z)$  has no solution  $z$  when  $w = \pm i$ , since solving  $w = \tan(z)$  gives

$$\frac{e^{2iz} - 1}{e^{2iz} + 1} = iw, \quad \text{or} \quad e^{2iz} = \frac{1 + iw}{1 - iw}.$$

In addition,  $f(z) = \tan(z)$  is not one-to-one, for if  $\tan(z_1) = \tan(z_2)$ , then  $z_1 - z_2 = k\pi$  for some  $k \in \mathbb{Z}$ . With suitable restrictions, we are able to define the arctangent function as an inverse:

**Definition 14.7.10** If we restrict  $f(z) = \tan(z)$  to the infinite vertical strip

$$T = \{z \in \mathbb{C} : z = x + iy, x, y \in \mathbb{R}, -\pi/2 < x < \pi/2\},$$

we see that  $f : T \rightarrow \mathbb{C} \setminus \{\pm i\}$  is both one-to-one and onto, so has an inverse  $g : \mathbb{C} \setminus \{\pm i\} \rightarrow T$  which is written  $g(z) = \text{Arctan}(z)$  (called the *principal branch of the arctangent function*).

Again we see that  $g$  is a bijective function with the properties:

$$\text{Arctan}(\tan(z)) = z, \quad \forall z \in T, \quad \text{and} \quad \tan(\text{Arctan}(z)) = z, \quad \forall z \in \mathbb{C} \setminus \{\pm i\}.$$

From the equation  $e^{2iz} = \frac{1 + iw}{1 - iw}$ ,  $w \neq \pm i$ , derived above, if we interchange  $z$  and  $w$  and using:  $z \in S$  and  $w \in \mathbb{C} \setminus \{0\}$ ,  $w = e^z$  if and only if  $z = \text{Log}(w)$ . Then for  $w \in T$  and  $(1 + iz)/(1 - iz) \in \mathbb{C} \setminus \{0\}$ ,

$$e^{2iw} = \frac{1 + iz}{1 - iz}, \quad \text{if and only if} \quad 2iw = \text{Log}\left(\frac{1 + iz}{1 - iz}\right).$$

This proves:

**Theorem 14.7.11** If  $g : \mathbb{C} \setminus \{\pm i\} \rightarrow T$ ,  $g(z) = \text{Arctan}(z)$  is the principal branch of the arctangent function, then

$$\text{Arctan}(z) = \frac{1}{2i} \text{Log}\left(\frac{1 + iz}{1 - iz}\right), \quad z \in \mathbb{C} \setminus \{\pm i\},$$

where we have used the principal branch of the logarithm function.

**14.7.12 Halley's Method and the Schröder-Cayley Theorem.** In this section we give a second proof of the Schröder-Cayley Theorem which is mainly of historical interest, and may be omitted, as it is somewhat more complicated than our earlier proof. We give the original proof due to Ernst Schröder (as outlined by Alexander [1]), whose study of Newton's Method for quadratic functions preceded that of Cayley by 10 years. Although Schröder's proof is a little more involved than the proof we

gave earlier, it is interesting both from an historical point of view, and from its use of properties of certain complex trigonometric functions. Schröder was a German mathematician of the 19th century who mainly worked in logic. We first remind the reader of Halley's method, a generalization of Newton's method that is sometimes useful. Sir Edmond Halley, after whom the comet is named, was a contemporary of Sir Isaac Newton. He is credited with persuading Newton to write up his studies on planetary motion, showing that the planets travel in ellipses around the sun.

Let  $f$  be a  $C^2$  function with  $f'(z) \neq 0$  on the domain of  $f$ . We have seen that if  $N_f(z) = z - f(z)/f'(z)$  is the Newton function of  $f(z)$  for which  $f(\alpha) = 0$  and  $f'(\alpha) \neq 0$ , then  $\alpha$  is a super-attracting fixed point of  $N_f$ . The generalization of this result, due to Halley, was used by Schröder to study the basins of attraction of fixed points of the Newton's function, arising from quadratic polynomials.

The idea is to apply Newton's method to the function  $g(z) = f(z)/f'(z)$ . The Newton function for  $g$  is

$$N_g(z) = z - \frac{f(z)f'(z)}{[f'(z)]^2 - f(z)f''(z)},$$

so that  $f(z) = 0$  if and only if  $N_g(z) = z$ . Differentiating gives

$$N'_g(z) = 1 - \left[ \frac{(f'(z))^4 - [f(z)f''(z)]^2 - [f'(z)f''(z) - f(z)f'''(z)]f(z)f'(z)}{[(f'(z))^2 - f(z)f''(z)]^2} \right],$$

and substituting  $z = \alpha$  gives  $N'_g(\alpha) = 0$ , so that  $\alpha$  is a super-attracting fixed point for  $N_g$ .

We will use Halley's method in the proof of the following theorem:

**Theorem 14.7.13** (Schröder [113], 1872, Cayley [27], 1882.) *Let  $f(z) = az^2 + bz + c$  be a complex quadratic function having two distinct roots  $\alpha_1$  and  $\alpha_2$ . Join  $\alpha_1$  and  $\alpha_2$  by a straight line, and denote the perpendicular bisector of this line by  $L$ . The basin of attraction of the fixed points  $\alpha_1$  and  $\alpha_2$  of the Newton function  $N_f$ , consists of all those points in the same open half plane determined by the line  $L$ .  $N_f$  is chaotic on  $L$ .*

**Proof.** It suffices to prove the theorem for  $f(z) = z^2 - \alpha$ ,  $\alpha \neq 0$ . We do this for  $\alpha = 1$  as the general case is similar. When  $\alpha = 1$ , if  $g(z) = f(z)/f'(z)$ , then  $g(z) = (z^2 - 1)/2z$ , and the generalized Newton function is

$$M(z) = N_g(z) = z - \frac{g(z)}{g'(z)} = z - \frac{(z^2 - 1)}{2z} \frac{4z^2}{((z^2 - 1)2 - (2z)^2)} = \frac{2z}{1 + z^2}.$$

If  $\tan(z) \neq \pm 1$ , we can set  $w = \tan(z)$  in the identity

$$\frac{2\tan(z)}{1 - \tan^2(z)} = \tan(2z),$$

to give

$$\frac{2w}{1 - w^2} = \tan(2\text{Arctan}(w)).$$

Replacing  $w$  by  $iz$  gives

$$M(z) = \frac{2z}{1 + z^2} = -i \tan(2\text{Arctan}(iz)),$$

provided  $z \neq \pm i$ . Iterate this function, using the identity  $\tan(2\text{Arctan}(\tan(iz))) = \tan(2iz)$  (see Exercises 14.6, where this identity can be seen to hold for any  $z$  not belonging to the imaginary axis) to find  $M^2$ :

$$M^2(z) = -i \tan(2\text{Arctan}(i(-i \tan(2\text{Arctan}(iz)))) = -i \tan(2^2 \text{Arctan}(iz)).$$

It is easily seen by induction that

$$M^n(z) = -i \tan(2^n \text{Arctan}(iz)).$$

This formula does not hold for certain values of  $z$  on the imaginary axis (such as  $z = \pm i$ ). However, we are only interested in the iterates on either side of this axis, so this is not important. Our aim is to show that  $\lim_{n \rightarrow \infty} M^n(z) \rightarrow 1$ , if  $\text{Re}(z) > 0$ , and  $\lim_{n \rightarrow \infty} M^n(z) \rightarrow -1$ , if  $\text{Re}(z) < 0$  (see Lemma 14.7.14 below). We then argue as follows to prove the theorem:

Suppose that  $N(z) = N_f(z) = z/2 + 1/(2z)$ . Then clearly  $N(1/z) = N(z)$  and  $M(1/z) = M(z)$ . Define  $h : \mathbb{C} \cup \{\infty\} \rightarrow \mathbb{C} \cup \{\infty\}$  by  $h(z) = 1/z$  (so that  $h(0) = \infty$  and  $h(\infty) = 0$ , a homeomorphism on the extended complex plane - see Section 14.3). It follows that

$$N \circ h(z) = N(1/z) = N(z) \quad \text{and} \quad h \circ M(z) = 1/M(z) = (z^2 + 1)/2z = N(z),$$

so that  $M$  and  $N$  are conjugate via  $h$ . In particular,  $N^n \circ h = h \circ M^n$  for all  $n \in \mathbb{Z}^+$ .

Since  $h(1) = 1$ , and  $h$  maps the imaginary axis to itself, we deduce that  $\lim_{n \rightarrow \infty} N_f^n(z) = \pm 1$ , on the required sets.

□

**Lemma 14.7.14** *If  $M(z) = -i \tan(2\text{Arctan}(iz))$ , then*

$$\lim_{n \rightarrow \infty} M^n(z) = \begin{cases} 1; & \text{Re}(z) > 0 \\ -1; & \text{Re}(z) < 0. \end{cases}$$

**Proof.** Consider

$$i \tan(2^n w) = \frac{e^{2^{n+1}iw} - 1}{e^{2^{n+1}iw} + 1} \quad \text{as } n \rightarrow \infty.$$

Set  $w = w_1 + iw_2$  where  $w_1$  and  $w_2$  are real. Then

$$i \tan(2^n w) = \frac{e^{2^{n+1}iw_1} e^{-2^{n+1}w_2} - 1}{e^{2^{n+1}iw_1} e^{-2^{n+1}w_2} + 1},$$

where  $|e^{2^{n+1}iw_1}| = 1$ .

There are two cases to consider:

**Case 1.**  $w_2 > 0$ , then  $e^{-2^{n+1}w_2} \rightarrow 0$  and  $i \tan(2^n w) \rightarrow -1$  as  $n \rightarrow \infty$ .

**Case 2.**  $w_2 < 0$ , then  $e^{2^{n+1}w_2} \rightarrow 0$  and

$$i \tan(2^n w) = \frac{e^{2^{n+1}iw_1} - e^{2^{n+1}w_2}}{e^{2^{n+1}iw_1} + e^{2^{n+1}w_2}} \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Now suppose that  $w = \operatorname{Arctan}(iz)$ ,  $z \neq \pm 1$ . Then by Theorems 14.7.7 and 14.7.11,

$$w = \frac{1}{2i} \log \left( \frac{1-z}{1+z} \right) = \frac{1}{2i} \left\{ \operatorname{Log} \left| \frac{z-1}{z+1} \right| + i\theta \right\},$$

where  $\theta = \operatorname{Arg} \left( \frac{1-z}{1+z} \right)$  is the principal argument.

Thus,

$$w = \frac{\theta}{2} - \frac{i}{2} \ln \left| \frac{z-1}{z+1} \right| = w_1 + iw_2,$$

where  $w_2 = -\frac{1}{2} \ln \left| \frac{z-1}{z+1} \right|$ . Note that  $w_2 < 0$  when  $\left| \frac{z-1}{z+1} \right| > 1$ , and this happens when  $|z-1| > |z+1|$ , or when  $\operatorname{Re}(z) > 0$ . Similarly  $w_2 > 0$  when  $\left| \frac{z-1}{z+1} \right| < 1$ , i.e., when  $\operatorname{Re}(z) < 0$ . We have shown that  $\operatorname{Re}(z) < 0$  implies  $\operatorname{Im}(w) < 0$ , and  $\operatorname{Re}(z) > 0$  implies  $\operatorname{Im}(w) > 0$ , so the lemma follows from Cases 1 and 2, (also using  $M^n(z) = -i \tan(2^n w)$ ). The proves the first part of the theorem.

To complete the proof of Theorem 14.7.13, we need to show that the Newton function is chaotic on the imaginary axis. This was done in Section 14.6, but we give an alternative proof in the exercises (see Exercises 14.7 # 10).

□

## Exercises 14.7

1. Prove Proposition 14.7.2.
2. Show that  $(e^z)^w \neq e^{zw}$ , in general.
3. Prove Proposition 14.7.4.
4. Let  $S = \{z \in \mathbb{C} : z = x + iy, x, y \in \mathbb{R}, -\pi < y \leq \pi\}$ . Prove that if  $f : S \rightarrow \mathbb{C} \setminus \{0\}$ ,  $f(z) = e^z$ , then  $f$  is one-to-one and onto.
5. Let  $T = \{z \in \mathbb{C} : z = x + iy, x, y \in \mathbb{R}, -\pi/2 < y < \pi/2\}$ . Prove that if  $f : T \rightarrow \mathbb{C} \setminus \{\pm 1\}$ ,  $f(z) = \tan z$ , then  $f$  is one-to-one and onto.
6. Prove that if  $z \in \mathbb{C}$ , then  $\tan(2 \arctan(\tan(z))) = \tan(2z)$ .
- 7\*. Give a direct proof that  $f(z) = e^z$  is a continuous function (i.e., using the definition of  $e^z$ ). Deduce that  $\sin z$  is continuous on  $\mathbb{C}$ . Discuss the continuity of  $\text{Log}(z)$ .
8. Let  $f(z) = \sin(z)$ . We have seen in Chapter 1, that if  $z$  is purely real, then  $f^n(z) \rightarrow 0$  as  $n \rightarrow \infty$ . Prove that if  $z$  is purely imaginary, then  $f^n(z) \rightarrow \infty$  as  $n \rightarrow \infty$ .
9. Let  $f(x) = x^2 + a^2$ , with Newton function  $N_f(x) = (x - a^2/x)/2$ . Show that  $N_f(x)$  is conjugate to the doubling map  $D(x) = \begin{cases} 2x; & 0 \leq x < 1/2 \\ 2x - 1; & 1/2 \leq x < 1, \end{cases}$  using the following steps: Set  $h(x) = a \tan(\pi x/2)$ .  
 Show that  $N_f \circ h(x) = h \circ C(x)$ , where  $C(x) = \begin{cases} 2x + 1; & -1 \leq x < 0 \\ 2x - 1; & 0 \leq x < 1. \end{cases}$  Now show that  $C$  is conjugate to  $D$  via  $k(x) = (x + 1)/2$ .

10. Use the result of the last problem to complete the proof of the Schröder-Cayley Theorem: Show that the Newton function from the Schröder-Cayley Theorem, restricted to the imaginary axis can be represented as  $N_f$  above, and hence it is chaotic on the imaginary axis.

11. If  $g = f/f'$ , where  $f$  is differentiable sufficiently many times, derive the formulas of Section 14.7.12 for  $N_g$  and  $N'_g$ .

Prove that  $N'_g(\alpha) = N''_g(\alpha) = 0$ , and  $N'''_g(\alpha) = -Sf(\alpha)$ , where  $Sf$  is the Schwarzian derivative of  $f$ .



## CHAPTER 15

### Examples of Substitutions.

A substitutions is a special type of function that gives rise to non-periodic, semi-infinite sequences. These sequences lie in spaces such as  $\Sigma = \mathcal{A}^{\mathbb{N}} = \{(a_0, a_1, a_2, \dots) : a_i \in \{0, 1\}\}$ , of sequences of zeros and ones, and turn out to be fixed points of the substitution map. In this chapter, we will derive some of the properties of the semi-infinite sequences that are defined by substitutions, giving an intuitive, non-rigorous approach to the theory of substitutions with an emphasis on examples. Besides being intrinsically interesting, we study substitutions for two main reasons: they give rise to an important class of shift dynamical systems, which, although not chaotic in the sense we have seen so far, have properties of a chaotic nature such as transitivity and sensitive dependence, and secondly, the fixed points of substitutions generate some beautiful examples of fractal curves that can be seen most easily using a computer algebra system. Special attention will be given to sequences that arise from some of the more famous substitutions. These are the Thue-Morse sequence, the Rudin-Shapiro sequence, the Fibonacci sequence and the Toeplitz sequence. We also take a look at paperfolding sequences and derive their properties.

In Chapter 16, we will see how substitutions can be used to generate fractals, and briefly discuss two-dimensional substitutions, and their related fractals. These fractals arise from the self-similarity property possessed by the fixed point of a substitution. In Chapters 18 and 19, we give a rigorous introduction to the mathematical theory of substitutions. This requires a study of compact metric spaces, especially the compactness of the sequence spaces such as  $\Sigma$  defined above. However, Chapter 15 is independent of the other chapters of this text and assumes only a basic knowledge of recursion, infinite sequences, and series.

#### 15.1 One-dimensional Substitutions and the Thue-Morse Substitution.

Before giving the definition, it is instructive to start with a simple example of a (one-dimensional), substitution. One-dimensional substitutions give rise to infinite sequences of integers, whereas two-dimensional substitutions give rise to two-dimensional arrays (infinite matrices), of integers.

**15.1.1 The Thue-Morse sequence.** The Thue-Morse sequence was discovered by E. Prouhet in 1851, rediscovered by Axel Thue in 1912 as an example of a non-periodic sequence with some other special properties (in his study of formal languages), and independently discovered by Marston Morse in 1917, (in his study of the dynamics of geodesics). This sequence is usually referred to as the *Thue-Morse sequence*, but sometimes just the *Morse sequence*. There are a number of equivalent ways of defining this sequence.

If  $n$  is a positive integer, we can write  $n = a_0 + a_1 2 + a_2 2^2 + \dots + a_k 2^k$ , uniquely, for some integer  $k$ , where  $a_i \in \{0, 1\}$  and  $2^k \leq n < 2^{k+1}$ . For example,  $53 = 1 + 0 \cdot 2 + 1 \cdot 2^2 + 0 \cdot 2^3 + 1 \cdot 2^4 + 1 \cdot 2^5$ .

**Definition 15.1.2** Define a sequence  $s(n)$ ,  $n = 0, 1, 2, \dots$ , by

$$s(n) = a_0 + a_1 + \dots + a_k, \pmod{2},$$

the sum of the digits, reduced modulo 2, in the binary expansion of  $n$ . This sequence is called the *Thue-Morse sequence*.

To be more explicit, if we write the integers  $0, 1, 2, \dots$ , in binary form:

$$0, 1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111, \dots$$

The sequence obtained by adding the digits of these numbers (reduced modulo 2) is the *Thue-Morse sequence*:

$$0110100110010110 \dots$$

It is a remarkable fact, that the Thue-Morse sequence may be obtained using a *substitution*, called the *Morse-substitution*. Set

$$\mathcal{A} = \{0, 1\},$$

a set with 2 distinct members.

Define a map  $\theta : \mathcal{A} \rightarrow \mathcal{A} \times \mathcal{A}$  in the following way (with some abuse of notation):

$$\begin{aligned}\theta(0) &= 01 \\ \theta(1) &= 10\end{aligned}$$

Now think of applying  $\theta$  so that it has a homomorphism type property:

$$\theta^2(0) = \theta(01) = \theta(0)\theta(1) = 0110,$$

$$\theta^3(0) = \theta(0110) = \theta(0)\theta(1)\theta(1)\theta(0) = 01101001,$$

$$\theta^4(0) = 0110100110010110,$$

and continue in this way to create a one-sided infinite sequence. We express this as

$$0 \rightarrow 01 \rightarrow 0110 \rightarrow 01101001 \rightarrow 0110100110010110 \rightarrow \dots,$$

and denote the infinite sequence obtained by  $\theta^\infty(0)$ .

We call a sequence of finite length a *word*. If  $\omega$  is such a word, then its length is denoted by  $|\omega|$ , and its *opposite* (or *reflection*), by  $R(\omega)$  (so  $R(0110) = 1001$ ). It can be seen that the Morse substitution sequence can be defined recursively by setting

$$\theta(0) = 01, \quad \theta(1) = 10 \quad \text{and} \quad \theta^{n+1}(0) = \theta^n(0)R(\theta^n(0)).$$

Clearly  $|\theta^n(0)| = 2^n$  is the length of the sequence at the  $n$ th stage of its construction. Our first proposition shows that the Thue-Morse sequence is the same sequence as that obtained from the Morse substitution.

**Proposition 15.1.3** *Let  $u_n$  be the  $n$ th term of  $\theta^\infty(0)$ , obtained from the Morse substitution. Then  $u_n = s(n)$ , ( $n = 0, 1, 2, \dots$ ), where  $s(n)$  is the  $n$ th term of the Thue-Morse sequence.*

**Proof.** We use induction. The result is clearly true for  $n = 0$ . Suppose it is true for all  $m < n$ . We may choose  $k$  so that  $2^k \leq n < 2^{k+1}$ .

Then  $u_n$  is the  $n$ th term in  $\theta^{k+1}(0) = \theta^k(0)\theta^k(1) = X_k R(X_k)$  (where  $X_k = \theta^k(0)$ ). Thus  $u_n$  is the  $(n - 2^k)$ th term in  $R(X_k)$ , and

$$u_n = (u_{n-2^k} + 1) \bmod 2.$$

By the inductive hypothesis,  $u_{n-2^k} = s(n - 2^k)$ .

Since  $2^k \leq n < 2^{k+1}$ , we have  $s(n) = s(n - 2^k) + 1 \pmod 2$  (for in this case we must have  $a_k = 1$ ), so that

$$u_n = (s(n - 2^k) + 1) \bmod 2 = s(n).$$

□

As we let  $n \rightarrow \infty$ , we see that the sequence  $\theta^n(0)$  converges to a sequence (the *Thue-Morse sequence*)  $u = 0110100110010110 \dots$ , which has the property that  $\theta(u) = u$ , i.e.,  $u$  is a fixed point of the substitution map  $\theta$  (the notion of convergence in this sense will be discussed in Chapter 17). Clearly, there is nothing special about the symbols 0 and 1, and we can equally use  $a$  and  $b$  to obtain

$$u = abbaababbaababbab \dots$$

In Chapter 16, we shall use the alphabet  $\mathcal{S} = \{L, F\}$  to indicate the commands for a turtle in the plane (in the sense of turtle geometry as in the programming language LOGO, developed by Seymour Papert in the 1980's). The symbol  $F$  represents a forward motion in the plane by one unit, and  $L$  represents a counter clockwise rotation of the turtle by the fixed angle  $\phi = \pi/3$ .

#### 15.1.4 Properties of the Thue-Morse Sequence $u$ .

We state some of the more important properties of the Thue-Morse sequence  $u$ . The first three properties are proved in Exercises 15.1. The second property will be investigated in a more general setting in Chapter 18. The other properties we mention in passing, as they are not directly related to the main theme of this text.

1.  $u$  is non-periodic (not even eventually periodic). It may be defined recursively by:

$$u_0 = 0, \quad u_{2n} = u_n, \quad \text{and} \quad u_{2n+1} = R(u_n),$$

where  $R(\omega)$  is the reflection of  $\omega$ .

2.  $u$  is *recurrent*: this means that every word that occurs in  $u$ , occurs infinitely often.

3. If we write the sequence as a formal power series:

$$F(x) = 0 + 1 \cdot x + 1 \cdot x^2 + 0 \cdot x^3 + 1 \cdot x^4 + \dots,$$

then  $F(x)$  satisfies the equation

$$(1+x)F^2 + F = \frac{x}{1+x^2} \pmod{2}.$$

This equation has two solutions,  $F$  and  $F'$ , ( $F' = R(F)$ , the reflection of  $F$ ), which satisfy

$$F + F' = 1 + x + x^2 + x^3 + \dots = \frac{1}{1+x} \pmod{2}.$$

In addition,

$$\prod_{i=0}^{\infty} (1 - x^{2^i}) = (1-x)(1-x^2)(1-x^4)\dots = 1 - x - x^2 + x^3 + \dots = \sum_{j=0}^{\infty} (-1)^{u_j} x^j.$$

4. It has been shown that if we think of  $u = .011010011001\dots$  as representing the binary expansion of a real number, then  $u$  is *transcendental* (it is not the root of any polynomial equation with rational coefficients), so in particular it is irrational.

5. Product formulas such as

$$\prod_{n=0}^{\infty} \left( \frac{2n+1}{2n+2} \right)^{2u_n} \left( \frac{2n+3}{2n+2} \right) = \frac{\sqrt{2}}{\pi},$$

have been established (where  $u_n$  is the  $n$ th term in the Thue-Morse sequence). As an example of this, we prove the following result, whose elementary proof is due to J. -P. Allouche (see [2]).

**Proposition 15.1.5** *Let  $\epsilon_n = (-1)^{u_n}$ , where  $(u_n)_{n \geq 0}$  is the Thue-Morse sequence. Then*

$$\prod_{n=0}^{\infty} \left( \frac{2n+1}{2n+2} \right)^{\epsilon_n} = \frac{1}{\sqrt{2}}.$$

**Proof.** Set  $P = \prod_{n=0}^{\infty} \left( \frac{2n+1}{2n+2} \right)^{\epsilon_n}$ ,  $Q = \prod_{n=1}^{\infty} \left( \frac{2n}{2n+1} \right)^{\epsilon_n}$ , then

$$PQ = \frac{1}{2} \prod_{n=1}^{\infty} \left( \frac{n}{n+1} \right)^{\epsilon_n} = \frac{1}{2} \prod_{n=1}^{\infty} \left( \frac{2n}{2n+1} \right)^{\epsilon_{2n}} \prod_{n=0}^{\infty} \left( \frac{2n+1}{2n+2} \right)^{\epsilon_{2n+1}},$$

(These can be seen to converge using Abel's Theorem).

From Property 1,  $\epsilon_{2n} = \epsilon_n$  and  $\epsilon_{2n+1} = -\epsilon_n$ , we get

$$PQ = \frac{1}{2} \prod_{n=1}^{\infty} \left( \frac{2n}{2n+1} \right)^{\epsilon_n} \left( \prod_{n=0}^{\infty} \left( \frac{2n+1}{2n+2} \right)^{\epsilon_n} \right)^{-1} = \frac{Q}{2P}.$$

Since  $Q \neq 0$ , we obtain  $P^2 = 1/2$ , and the result follows as  $P$  is positive.  $\square$

### 15.1.6 The Fibonacci Substitution.

The Morse substitution is an example of a substitution of *constant length*. For each  $x \in \mathcal{A} = \{0, 1\}$ ,  $\theta(x)$  is always the same length. The Fibonacci substitution is a substitution of *non-constant length*, and is defined on  $\mathcal{A} = \{0, 1\}$  as follows:

$$\theta(0) = 01, \quad \theta(1) = 0.$$

We can check that

$$\theta^2(0) = 010, \quad \theta^3(0) = 01001, \quad \theta^4(0) = 01001010$$

$$\theta^5(0) = 0100101001001.$$

It is easy to show using the identity  $\theta^{n+2}(0) = \theta^{n+1}(0)\theta^n(0)$ ,  $n \geq 1$ , (see Exercises 15.1), that  $|\theta^{n-2}(0)| = F_n$ ,  $n = 2, 3, \dots$ , is the  $n$ th term in the Fibonacci sequence (where  $F_0 = 0$ ,  $F_1 = 1$  and  $F_{n+2} = F_{n+1} + F_n$ ,  $n \geq 0$ ).

### 15.1.7 Definition of a Substitution.

In order to gain a better understanding of how substitutions work, we have been looking at examples before giving the formal definition. We define substitutions formally as follows: start with a finite set  $\mathcal{A}$ , and denote by  $\mathcal{A}^*$ , the set of all *words* on  $\mathcal{A}$ . This means all possible finite strings of letters using our *alphabet*  $\mathcal{A}$ . Thus, if  $w \in \mathcal{A}^*$ , then  $w = a_1a_2\dots a_n$  for some  $a_1, a_2, \dots, a_n \in \mathcal{A}$ , where  $n$  the length of  $w$ . Two words,  $w$  and  $w'$ , are joined by *concatenation* to form a new word. If  $w = a_1\dots a_n$  and  $w' = b_1\dots b_m$ , then

$$ww' = a_1\dots a_nb_1\dots b_m \in \mathcal{A}^*.$$

The *empty word* (denoted by  $\epsilon$ ), is a word of zero length with the property that  $w\epsilon = w = \epsilon w$ .

**Definition 15.1.8** A *substitution* is a mapping  $\theta$  from  $\mathcal{A}$  into the set of words  $\mathcal{A}^*$  and then extended as a map from  $\mathcal{A}^*$  into  $\mathcal{A}^*$  by concatenation, using the homomorphism property:

$$\theta(w_1w_2) = \theta(w_1)\theta(w_2) \quad \text{for any } w_1, w_2 \in \mathcal{A}^*.$$

If we set

$$\mathcal{A} = \{0, 1, 2, \dots, s-1\}, \quad \text{for some integer } s > 1,$$

then

$$\mathcal{A}^* = \bigcup_{i=0}^{\infty} \mathcal{A}^i,$$

is the set of all words of finite length (where  $\mathcal{A}^i$  is the  $i$ -fold cartesian product and  $\mathcal{A}^0$  contains only the empty word). If  $w \in \mathcal{A}^n$ , strictly speaking,  $w = (a_1, \dots, a_n)$  is an  $n$ -tuple, but we write it as  $w = a_1\dots a_n$ , a word of length  $n$ . The substitution is a map

$$\theta : \mathcal{A} \rightarrow \mathcal{A}^*$$

which is extended to a map  $\theta : \mathcal{A}^{\mathbb{N}} \rightarrow \mathcal{A}^{\mathbb{N}}$  and to  $\theta : \mathcal{A}^* \rightarrow \mathcal{A}^*$  by concatenation, using the homomorphism property.  $\mathcal{A}^{\mathbb{N}}$  is the set of all infinite sequences indexed by  $\mathbb{N} = \{0, 1, 2, \dots\}$ ,  $\mathcal{A}^{\mathbb{N}} = \{(a_0, a_1, \dots) : a_i \in \mathcal{A}\}$ , but we think of these as infinite words.

If  $\theta(0)$  begins with 0, and  $|\theta(0)| > 1$  then the sequence

$$u = \theta^{\infty}(0) = \lim_{n \rightarrow \infty} \theta^n(0) \in \mathcal{A}^{\mathbb{N}},$$

(called a *substitution sequence* for  $\theta$ ), satisfies  $\theta(u) = u$ .

The requirement that  $\theta(0)$  begins with 0 and has length greater than one, ensures that when we apply  $\theta$  to  $\theta^n(0)$ , its length increases exponentially, and the first  $|\theta^n(0)|$  terms of  $\theta^{n+1}(0)$  coincide with  $\theta^n(0)$  (there is no “flipping” of the sequence). This ensures that  $u = \theta^\infty(0)$  is well defined with  $\theta(u) = u$ , a *fixed point of the substitution*.

In Chapters 17 and 18, we shall see how substitutions give rise to dynamical systems having many of the properties of chaotic maps.

### Exercises 15.1

1. (a) If  $\theta$  is a substitution on  $\{0, 1\}$  for which  $|\theta(0)| > 1$ , and with  $\theta(0)$  starting with 0, show that  $|\theta^n(0)| \rightarrow \infty$  as  $n \rightarrow \infty$ , and  $\theta^{n+1}(0)$  starts with  $\theta^n(0)$ . Deduce that  $u = \theta^\infty(0)$  is a fixed point of the substitution.  
(b) Find the fixed points (if any), of the following substitutions on  $\{0, 1\}$ . (i)  $\theta(0) = 01$ ,  $\theta(1) = 1$ , (ii)  $\theta(0) = 010$ ,  $\theta(1) = 101$ , (iii)  $\theta(0) = 11$ ,  $\theta(1) = 10$ .
2. Note that the Morse substitution has fixed points  $u = 01101\dots$ , and  $R(u) = 10010\dots$ . Show that there are no other fixed points.
3. (a) If  $(u_n)$  is the Thue-Morse sequence, prove that  $u_{2n} = u_n$  and  $u_{2n+1} = R(u_n)$  for  $n \geq 0$ . (Hint: It is easier to show this for the sequence  $s(n)$ : If  $n = a_0 + a_12 + a_22^2 + \dots + a_k2^k$ , for  $2^k \leq n < 2^{k+1}$ , then  $s(n) = a_0 + a_1\dots + a_k \pmod{2}$ . Now think about what  $2n$  and  $s(2n)$  are equal to).  
(b) Prove that if  $(u_n)$  is the Thue-Morse sequence, then  $u_0u_1\dots u_{2^{2n}-1}$ , is a *palindrome* (i.e., a word that reads the same forwards and backwards). (Hint: This is most easily done for the sequence  $s(n)$ ).
4. Let  $w = u_su_{s+1}\dots u_t$  be a finite string of consecutive terms from the Thue-Morse sequence. Prove that there exists a number  $n_w$  such that every string of  $n_w$  consecutive terms  $u_{k+1}u_{k+2}\dots u_{k+n_w}$  from the sequence must contain  $w$ . (Hint: Take  $w' = u_0u_1\dots u_{2^n-1}$ , the shortest string of  $2^n$  consecutive terms of the sequence  $u$  containing  $w$  and deduce that any sequence of length  $2^{n+2}$  consecutive terms has the required property).

5. Denote by  $s_3(n)$  the sum of the digits (modulo 3) in the ternary expansion of  $n \in \mathbb{N}$ . What substitution on  $\mathcal{A} = \{0, 1, 2\}$  gives rise to the sequence  $s_3(n)$ ?
6. Show that for the Fibonacci substitution  $\theta$ ,  $F_n = |\theta^{n-2}(0)|$  for each  $n \geq 2$ , where  $F_n$  is the  $n$ th term in the Fibonacci sequence.
7. Prove that the Thue-Morse sequence  $u$  is recurrent (every word appearing in  $u$  appears infinitely often). (Hint: 0 appears infinitely often in  $u$ . If a word  $w$  appears in  $u$ , it must appear in  $\theta^k(0)$  for some  $k \in \mathbb{N}$ ).
8. (a) If the Thue-Morse sequence is written as a formal power series:

$$F(x) = 0 + 1 \cdot x + 1 \cdot x^2 + 0 \cdot x^3 + 1 \cdot x^4 + \dots,$$

show, by checking the first few terms, that  $F(x)$  satisfies the quadratic equation

$$(1+x)F^2 + F = \frac{x}{1+x^2} \pmod{2}.$$

- (b) Show in a similar way that

$$\prod_{i=0}^{\infty} (1-x^{2^i}) = (1-x)(1-x^2)(1-x^4)\dots = 1 - x - x^2 + x^3 + \dots = \sum_{j=0}^{\infty} (-1)^{u_j} x^j.$$

9. A substitution  $\theta$  is defined on  $S = \{0, 1, 2\}$  by

$$\theta(0) = 010, \quad \theta(1) = 121, \quad \theta(2) = 202.$$

Write down  $\theta^n(0)$  for  $n = 2, 3, 4$ , and show that if  $(x_n)$  is the  $n$ th term of the sequence,

$$x_{3n} = x_n, \quad x_{3n+1} = x_n + 1, \quad x_{3n+2} = x_n, \quad n \geq 0,$$

(addition being modulo three).

## 15.2 The Toeplitz Substitution.

In this section we give another important example of a class of substitutions, to be defined formally in 15.2.1. Set  $\mathcal{A} = \{0, 1\}$ , and define a substitution  $\phi$  on  $\mathcal{A}$  by

$$\phi(0) = 11, \quad \phi(1) = 10,$$

so that

$$1 \rightarrow 10 \rightarrow 1011 \rightarrow 10111010 \rightarrow 1011101010111011\dots$$

The fixed point (which is found by iterating 1 in this case), of the substitution is an example of a *Toeplitz sequence* (sometimes called the  $2^\infty$ -sequence or the period doubling sequence). We shall call the corresponding substitution the *Toeplitz substitution*. Generally, Toeplitz sequences can be constructed in the following way:

A sequence  $t = t_0t_1t_2\dots \in \mathcal{A}^{\mathbb{N}}$  is defined using the following steps: start with a semi-infinite sequence of blank spaces and place a ‘1’ in the 0’th place, and a ‘1’ in every other blank space thereafter (so  $t_{2n} = 1$  for all  $n \in \mathbb{N}$ ). This gives:

$$1 \cdot 1 \cdot \dots$$

Now place a ‘0’ in the first blank space, and in every other blank space thereafter (we are missing every other blank space, so  $t_{4n+1} = 0$  for all  $n \in \mathbb{N}$ ):

$$101 \cdot 101 \cdot 101 \cdot 101 \cdot 101 \cdot 101 \cdot \dots$$

Next place a ‘1’ in the first remaining blank space, and a ‘1’ in every other remaining blank space thereafter (so  $t_{8n+3} = 1$  for all  $n \in \mathbb{N}$ ):

$$1011101 \cdot 1011101 \cdot 1011101 \cdot \dots$$

Continue indefinitely in this way, alternately placing 0’s in every other blank space, then 1’s in every other blank space. The construction can be varied with different alphabets. Our construction gives one of the simplest examples of a *Toeplitz sequence*:

$$101110101011101110111010\dots$$

Comparing this sequence with the fixed point of the Toeplitz substitution, we see that they are identical in the first few terms. In the following proposition, we show that the two ways of defining the Toeplitz sequence are equivalent. Later, we shall see a connection between the Toeplitz sequence and the Morse sequence.

**Definition 15.2.1** A point  $t = t_0t_1t_2 \dots \in \mathcal{A}^{\mathbb{N}}$  is a *Toeplitz sequence*, if for all  $k \in \mathbb{N}$ , there exists  $p_k \in \mathbb{Z}^+$ , such that

$$t_k = t_{k+n \cdot p_k}, \quad \text{for all } n \in \mathbb{N}.$$

The next proposition tells us that the Toeplitz sequence generated in the above way is exactly the sequence generated by the Toeplitz substitution (the  $2^\infty$ -sequence).

**Proposition 15.2.2** *The fixed point of the Toeplitz substitution  $\phi$  defined on  $\mathcal{A} = \{0, 1\}$  by  $\phi(0) = 11$ ,  $\phi(1) = 10$  is exactly the Toeplitz sequence generated above.*

**Proof.** Define a word  $A_n$  of length  $2^n$  recursively, by setting  $A_0 = 1$ ,  $A'_0 = 0$ , and  $A_n = A_{n-1}A'_{n-1}$  where  $A'_{n-1} = A_{n-1}$  except they differ in the last letter. We can show by induction, that  $A_n = \phi^n(1)$  for  $n \in \mathbb{N}$  (see Exercises 15.2).

Thus, we see that  $\phi^n(0)$  and  $\phi^n(1)$  are identical except that they differ in the very last letter. We now show using induction, that  $\phi^n(1)$  coincides with the first  $2^n$  terms of the Toeplitz sequence  $t$ .  $\phi(1) = 10$ . This is clearly true when  $n = 1$ . Suppose that  $\phi^n(1)$  coincides with the first  $2^n$  terms of  $t$ . Then we have

$$\phi^{n+1}(1) = \phi^n(\phi(1)) = \phi^n(10) = \phi^n(1)\phi^n(0) = A_nA'_n.$$

Denote by  $B_n$  the word  $A_n$  with the last letter omitted. Then we must have the  $n$ th stage of the construction of  $t$  as,

$$B_n \cdot B_n \cdot B_n \cdot B_n \cdot \dots,$$

where we still have a space between the different versions of the word  $B_n$ . The first space gets filled with a 0 if  $n$  is odd, and the second space with a 1, and vice-versa if  $n$  is even. Thus  $t = A_n A'_n \dots$ , as required. □

## Exercises 15.2

1. A sequence  $A_n$  is defined recursively as follows:

$$A_0 = 1, \quad A'_0 = 0, \quad A_n = A_{n-1}A'_{n-1},$$

where  $A'_{n-1}$  is the word  $A_{n-1}$ , except that the last letter (say  $a$ ) is replaced by  $1 - a = R(a)$ . Write down  $A_2$  and  $A_3$ , and use induction to show that  $\lim_{n \rightarrow \infty} A_n$  is the sequence obtained from the Toeplitz substitution.

2. Define a map  $b : \{00, 01, 10, 11\} \rightarrow \{0, 1\}$  by  $b(ab) = a + b \pmod{2}$ , and extend it as a map  $b : \mathcal{A}^{\mathbb{N}} \rightarrow \mathcal{A}^{\mathbb{N}}$  in the obvious way ( $\mathcal{A} = \{0, 1\}$  and  $b(a b c \dots) = a + b b + c \dots$ ). Find  $b(0110100110010110 \dots)$ , the image of the Morse sequence. What is special about the sequence obtained?
3. Let  $\mathcal{A} = \{0, 1, 2\}$  and  $\theta$  be the substitution defined on  $\mathcal{A}$  by  $\theta(0) = 010$ ,  $\theta(1) = 121$ ,  $\theta(2) = 202$ . If  $b(a b c) = a + b + c \pmod{3}$ , show that the image of  $\theta^\infty(0)$  under  $b$  is a type of Toeplitz sequence (in a similar way to the previous example).

4\*. If  $u = u_0 u_1 u_2 \dots$  is the Morse sequence, and  $t = t_0 t_1 t_2 \dots$  is the Toeplitz sequence, show that

$$u_{n+1} = \sum_{k=0}^n t_k \pmod{2}.$$

### 15.3 The Rudin-Shapiro Sequence.

Another famous substitution arises in connection with the *Rudin-Shapiro sequence*, introduced by H. S. Shapiro (1951) and W. Rudin (1959), to answer a question of R. Salem (1950) in harmonic analysis. Specifically, they constructed a sequence  $\varepsilon = \varepsilon_0 \varepsilon_1 \dots$ , where  $\varepsilon_n \in \{-1, 1\}$  has the property that for all integers  $N \geq 0$ ,

$$\sup_{\theta \in [0, 1]} \left| \sum_{n=0}^{N-1} \varepsilon_n e^{2\pi i n \theta} \right| \leq (2 + \sqrt{2}) \sqrt{N}.$$

The Rudin-Shapiro sequence was subsequently used to solve an open question in ergodic theory: that there exists an ergodic measure preserving transformations having a finite Lebesgue component in its spectrum (see [104]). The areas of ergodic theory and harmonic analysis are topics beyond the scope of this text. The interested reader should consult the texts of Walters [127], Queffelec [104], and Petersen [102]. However, we will see that the Rudin-Shapiro sequence has some interesting topological and combinatorial properties.

Just as the Morse sequence gives the *parity* of the number of 1's in the binary expansion of  $n \in \mathbb{N}$  (i.e., the number of 1's reduced modulo 2), the Rudin-Shapiro sequence gives the parity of the number of words '11' in the binary expansion of  $n$ . More explicitly, it gives the number of (possibly overlapping), occurrences of the word 11 in the base-2 expansion of  $n$ . This can be proved using induction, and the relations

$\varepsilon_{2n} = \varepsilon_n$  and  $\varepsilon_{2n+1} = (-1)^n \varepsilon_n$ , the parity being  $r_n = 0$  when  $\varepsilon_n = 1$  and  $r_n = 1$  when  $\varepsilon_n = -1$ , so that  $\varepsilon_n = (-1)^{r_n}$ .

**Definition 15.3.1** The *Rudin-Shapiro sequence*  $\varepsilon = \varepsilon_0 \varepsilon_1 \dots$ , where  $\varepsilon_n \in \{-1, 1\}$ , is defined recursively by:  $\varepsilon_0 = 1$ , and if  $n \geq 1$  then

$$\varepsilon_{2n} = \varepsilon_n \quad \text{and} \quad \varepsilon_{2n+1} = (-1)^n \varepsilon_n.$$

If we list  $n$ , the base-2 expansion of  $n$ ,  $r_n$  and  $\varepsilon_n$  we have

$n =$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	...
$n_2 =$	0	1	10	11	100	101	110	111	1000	1001	1010	1011	1100	1101	...
$r_n =$	0	0	0	1	0	0	1	0	0	0	0	1	1	1	...
$\varepsilon_n =$	1	1	1	-1	1	1	-1	1	1	1	1	-1	-1	-1	...

**Proposition 15.3.2** The Rudin-Shapiro sequence  $\varepsilon_n$ , and the sequence that gives the parity of the word 11 in the binary expansion of  $n \in \mathbb{N}$  are identical.

**Proof.** See Exercises 15.3. □

### The Rudin-Shapiro Substitution 15.3.3

The Rudin-Shapiro sequence is closely related to a substitution  $\zeta$  (the Greek letter zeta), called the *Rudin-Shapiro substitution*. This substitution is defined on the alphabet  $\mathcal{A} = \{0, 1, 2, 3\}$  in the following way:

$$\zeta(0) = 02, \quad \zeta(1) = 32, \quad \zeta(2) = 01, \quad \zeta(3) = 31.$$

If  $u = u_0 u_1 \dots = \zeta^\infty(0)$  denotes the fixed point of  $\zeta$ , then  $\varepsilon = \tau(u)$ , where  $\tau$  is the map  $\tau : \mathcal{A} \rightarrow \{-1, 1\}$ ,  $\tau(0) = 1 = \tau(2)$  and  $\tau(1) = -1 = \tau(3)$ , extended in the usual way. We call the map  $\tau$ , a *recoding* of the substitution. The fixed point is generated as:

$$0 \rightarrow 02 \rightarrow 0201 \rightarrow 02010232 \rightarrow 0201023202013101\dots,$$

**Proposition 15.3.4** The recoding of the fixed point of the Rudin-Shapiro substitution under the map  $\tau$  gives the Rudin-Shapiro sequence.

**Proof.** See Exercises 15.3. □

### Exercises 15.3

1. Prove Proposition 15.3.2: that the Rudin-Shapiro sequence  $\epsilon_n$ , and the sequence that gives the parity of the word 11 in the binary expansion of  $n \in \mathbb{N}$ , are identical.
2. Prove Proposition 15.3.4: that the recoding of the fixed point of the Rudin-Shapiro substitution, under the map  $\tau$ , gives the Rudin-Shapiro sequence.

3. Let  $f(x) = \sum_{n=0}^{N-1} c_n e^{2\pi i n x}$ ,  $0 \leq x \leq 1$ , where each of the coefficients  $c_n$  is  $\pm 1$ .

(a) Show that  $\int_0^1 |f(x)|^2 dx = \sum_{n=0}^{N-1} |c_n|^2 = N$ .

(b) Deduce that  $\sup_{x \in [0,1]} \left| \sum_{n=0}^{N-1} c_n e^{2\pi i n x} \right| \geq \sqrt{N}$ .

(c)\*\* (See [45])). If  $\epsilon_n$  is the Rudin-Shapiro sequence, show that

$$\sup_{x \in [0,1]} \left| \sum_{n=0}^{N-1} \epsilon_n e^{2\pi i n x} \right| \leq (2 + \sqrt{2})\sqrt{N}.$$

4\*. Given a sequence  $(\alpha(n))$ ,  $n \geq 0$ , of complex numbers having absolute value equal to 1, the *correlation function*  $\sigma(n)$  of the sequence  $\alpha(n)$  is defined by:

$$\sigma(n) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=0}^{m-1} \alpha(n+k) \overline{\alpha(k)}, \quad n > 0,$$

$$\sigma(-n) = \overline{\sigma(n)}, \quad n < 0, ; \quad \sigma(0) = 1.$$

The correlation function  $\sigma(n)$ , was introduced by N. Wiener. The *Wiener class*  $\mathcal{S}$  are those sequences for which this limit exists. The sequence  $\sigma(n)$  is a positive definite sequence. (A sequence  $(u_n)$  of complex numbers is *positive definite* if  $\sum_{n,m=0}^N u_{n-m} z_n \bar{z}_m \geq 0$  for all sequences  $(z_n)$  of complex numbers, and all  $N \in \mathbb{N}$  (see [99] or [104]).

The sequence  $(\sigma(n))$  gives information about what is called the *spectrum* of the sequence, and related operators (see [59] or [104] for example).

(a) If  $\alpha(n) = \epsilon_n = (-1)^{u_n}$ , where  $u_n$  is the Thue-Morse sequence, show that the correlation function  $\sigma$  satisfies  $\sigma(2n) = \sigma(n)$ ,  $\sigma(2n+1) = (\sigma(n+1) - \sigma(n))/2$ .

(b) If  $\alpha(n) = \omega^{x_n}$  where  $x_n$  is the sequence defined in Exercises 15.1 # 9, and  $\omega = e^{2\pi i/3}$ , show that

$$\sigma(3n) = \sigma(n), \quad \sigma(3n+1) = \frac{1}{3}(\sigma(n+1) - \sigma(n)), \quad \sigma(3n+2) = \frac{1}{3}(\sigma(n) - \sigma(n+1)).$$

(c) Show that the sequence  $\sigma(n)$ , is positive definite when the limit exists. (Hint: We can do this for any sequence of the form:  $\sigma(k) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{n=0}^{m-1} u_n \bar{u}_{n+k}$ . Start by showing that  $\sum_{n=0}^{m-1} u_n \bar{u}_{n+k} = \int_0^{2\pi} \left| \sum_{n=0}^{m-1} u_n e^{int} \right|^2 e^{ikt} dt.$ )

**Remarks.** It can be shown that the correlation function  $\sigma(n)$  of the Rudin-Shapiro sequence satisfies  $\sigma(n) = 0$  for all  $n \neq 0$  (see [45] page 44).

## 15.4 Paperfolding Sequences.

In Section 15.4, we examine the connection between the sequences obtained by looking at the creases in the folding of a sheet of paper, and certain types of substitutions. More detail can be found in the beautiful expositions of Dekking, Mendés France, and van der Poorten [33], [34], and [35], where the relations between paperfolding, automata, dragon curves, and substitutions are examined (see also [90]). In Chapter 16 we will look at so called *dragon curves*. These are the fractal curves that arise from paperfolding.

Put a sheet of paper on the desk in front of you, and fold the right side on top of the left side, and continue doing this repeatedly. When we fold a sheet of paper (which is hard to do more than 7 times, but we can imagine repeating this process indefinitely), and then unfold it, a pattern of folds indented in the paper can be seen. We represent the folds by  $\vee$  and  $\wedge$ , and see that the pattern obtained for the first few folds is:

fold # 1		∨									
fold # 2		∨	∨	∧							
fold # 3	∨	∨	∧	∨	∨	∧	∧				
fold # 4	∨	∨	∧	∨	∨	∧	∧	∨	∧	∧	∧

To obtain the pattern at fold  $n+1$ , first write down the pattern at fold  $n$  (say  $w$ ), add the middle fold  $\vee$ , followed by the “reflection”  $R(\tilde{w})$ . This is not the reflection we have used previously. Denote by  $\tilde{w}$ , the word  $w$  with order reversed. Then interchange  $\vee$  and  $\wedge$ . For example, if  $w = \vee \vee \wedge \vee \vee \wedge \wedge$ , then  $\tilde{w} = \wedge \wedge \vee \vee \wedge \vee \vee$  and  $R(\tilde{w}) = \vee \vee \wedge \wedge \vee \wedge \wedge$ . We then obtain by concatenation, the new word  $w \vee R(\tilde{w}) = \vee \vee \wedge \vee \vee \wedge \wedge \vee \vee \vee \wedge \wedge \vee \wedge \wedge$ . Representing  $\vee$  by 1 and  $\wedge$  by 0, we obtain the sequence

$$1 1 0 1 1 0 0 1 1 1 0 0 1 0 0 \underline{1} 1 1 0 1 1 0 0 1 1 0 0 1 0 0 \dots$$

(where the middle term of the 5th fold has been underlined). The length of the sequence at the  $n$ th, fold is  $2^n - 1$ . This sequence is well defined, and can be seen to converge to a unique infinite sequence.

Denote by  $f = f_1 f_2 f_3 \dots$  the paper folding sequence of 0’s and 1’s. Notice that the subsequence  $f_1 f_3 f_5 \dots$  is just the period-2 sequence 1 0 1 0 1 0 …, and  $f_2 f_4 f_6 \dots$ , coincides with the original sequence  $f$ . Suppose we use the sequence  $(f_n)$  to define a formal power series:

$$f(X) = \sum_{n=1}^{\infty} f_n X^n.$$

Then working formally, and using these recurrence relations, we observe that

$$\begin{aligned} f(X) - f(X^2) &= (f_1 X + f_2 X^2 + f_3 X^3 + \dots) - (f_1 X^2 + f_2 X^4 + f_3 X^6 + \dots) \\ &= f_1 X + f_3 X^3 + f_5 X^5 + \dots = X + X^5 + X^9 + \dots \\ &= X(I + X^4 + X^8 + \dots) = \frac{X}{I - X^4}. \end{aligned}$$

We have shown:

**Proposition 15.4.1** *If  $f = f_1 f_2 f_3 \dots$  is the paper folding sequence and  $f(X) = \sum_{n=1}^{\infty} f_n X^n$ , then*

$$f(X) - f(X^2) = \frac{X}{I - X^4}.$$

Our original construction of the paperfolding sequence may be varied. Instead of repeatedly folding right over left, we may fold right under left in a random way. We

call the resulting sequence a *paperfolding sequence*, and the sequence described above is the *positive paperfolding sequence*. These sequences can be seen to converge to some particular infinite sequence. This leads to the following observation characterizing paperfolding sequences, indicating that paperfolding sequences are a type of Toeplitz sequence.

**Property 15.4.2** A *paperfolding sequence*  $(g_n)_{n \geq 1}$  on the alphabet  $\mathcal{A} = \{0, 1\}$  has the following properties:

- (a)  $g_{2n-1} = (g_1 + n) \pmod{2}$ .
- (b)  $(g_{2n})$  is again a paperfolding sequence.

In particular, our original paperfolding sequence  $(f_n)$  has these properties, and any sequence with the properties  $f_1 = 1$ ,  $f_{2n-1} = f_1 + n \pmod{2}$  and  $f_{2n} = f_n$ ,  $n \geq 1$ , must be the positive paperfolding sequence. It follows that  $(f_n)$  is a Toeplitz sequence based on the periodic sequence  $1 \cdot 0 \cdot 1 \cdot 0 \cdot 1 \dots$ , as follows:

$$\begin{array}{cccccccccccccccccccc}
 1 & \cdot & 0 & \cdot & \cdots \\
 1 & & \cdot & & 0 & & \cdot & & 1 & & \cdot & & 0 & & \cdot & & \cdots \\
 & 1 & & & & \cdot & & & & 0 & & & & \cdot & & & 1 & \cdots \\
 & & & & & & 1 & & & & & & & & & \cdot & & \cdots \\
 & & & & & & & & 1 & & & & & & & & & \cdots \\
 & & & & & & & & & & & & & & & 1 & & \cdots
 \end{array}$$
  

$$1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 1 \ \dots$$

**Proposition 15.4.3** *The positive paperfolding sequence is a Toeplitz sequence.*

**Proof.** This is just the fact that  $f_{2n-1}$ ,  $n \geq 1$ , is the periodic sequence  $1010101\dots$ , and  $f_{2n}$  is the sequence  $f_n$  (see Exercises 15.4).

□

Just as for the Morse sequence, it has been shown that the real number whose binary digits are the  $f_n$ :

$$\alpha = \sum_{n=1}^{\infty} \frac{f_n}{2^n} = .110110011100100\dots,$$

is transcendental, and this is generally true for paperfolding sequences. In particular:

**Proposition 15.4.4** *The positive paperfolding sequence is non-periodic and not ultimately periodic.*

**Proof.** Use the fact that the positive paperfolding sequence is a Toeplitz sequence (see Exercises 15.4). □

#### 15.4.5 The Substitution Associated with a Paperfolding Sequence.

Notice that if we define a substitution using the alphabet  $\{00, 01, 10, 11\}$  in the following way:

$$00 \rightarrow 1000, \quad 01 \rightarrow 1001, \quad 10 \rightarrow 1100, \quad 11 \rightarrow 1101,$$

then it has a fixed point given by:

$$11 \rightarrow 1101 \rightarrow 11011001 \rightarrow 1101100111001001$$

$$\rightarrow 110110011100100111011001100011001001\dots,$$

which is exactly the paperfolding sequence. Now replace 00 by 0, 01 by 1, 10 by 2 and 11 by 3 (a *recoding*), then we can write the substitution as

$$0 \rightarrow 20, \quad 1 \rightarrow 21, \quad 2 \rightarrow 30, \quad 3 \rightarrow 31,$$

which has a fixed point given by the limit of

$$3 \rightarrow 31 \rightarrow 3121 \rightarrow 31213021 \rightarrow 3121302131203021 \rightarrow \dots$$

**Proposition 15.4.6** *The positive paperfolding sequence is the fixed point of the substitution*

$$\eta(0) = 20, \quad \eta(1) = 21, \quad \eta(2) = 30, \quad \eta(3) = 31,$$

when it is recoded via the map  $\tau : \{0, 1, 2, 3\} \rightarrow \{00, 01, 10, 11\}$ ,  $\tau(0) = 00$ ,  $\tau(1) = 01$ ,  $\tau(2) = 10$ , and  $\tau(3) = 11$ .

**Proof.** It suffices to show that the fixed point  $t$ , of the substitution  $\theta$ :

$$\theta(00) = 1000, \quad \theta(01) = 1001, \quad \theta(10) = 1100, \quad \theta(11) = 1101,$$

is the positive paperfolding sequence. Write  $t = x_1x_2\dots x_n\dots$ . Then clearly  $x_1 = 1$  and  $(x_{2n-1})$  is the sequence 101010..., because of the way the substitution is defined. It therefore suffices to show that  $x_{2n} = x_n$  for  $n \geq 1$ . This follows from the

fact that when we delete the first and third terms from  $\theta(i j)$  we get  $i j$ . Thus we have

$$(x_n) = \theta^\infty(11) = \theta(11)\theta(01)\theta(10)\theta(01)\theta(11)\dots = 1101100111\dots = (x_{2n}).$$

□

### Exercises 15.4

1. Let  $(f_n)_{n \geq 1}$  denote the positive paperfolding sequence.
  - (a) Show that the sequence  $(f_{2n-1})_{n \geq 1}$  is the periodic sequence  $101010\dots$ , and that the sequence  $(f_{2n})_{n \geq 1}$  coincides with the original sequence  $(f_n)$ .
  - (b) Use (a) to show that the paperfolding sequence  $(f_n)_{n \geq 1}$ , is a Toeplitz sequence.
2. Show that the positive paperfolding sequence is non-periodic and not ultimately periodic.

3. **Stacking Transformations** (see [48] or [49]). We define a *stacking transformation*  $T : [0, 1] \rightarrow [0, 1]$  as follows: map  $[0, 1/2]$  linearly onto the interval  $[1/2, 1]$ , so that  $T(x) = x + 1/2$ , for  $0 \leq x < 1/2$ .

Imagine “stacking” the interval  $[1/2, 1]$  on top of  $[0, 1/2]$  in such a way that a point is mapped vertically to the point above. Map  $[1/2, 3/4]$  linearly onto  $[1/4, 1/2]$ , so that  $T(x) = x - 1/4$  for  $1/2 \leq x < 3/4$ . We can picture this as cutting our stack into two equal pieces, and placing the right piece on top of the left piece to form a single stack consisting of 4 levels, each of length  $1/4$ . Each point is mapped to the point vertically above, but the map remains undefined on the top level. We continue this procedure so that at the  $n$ th stage we have a stack of  $2^n$  intervals, each of length  $1/2^n$ . We cut each column into two equal parts, and stack the right column on top of the left column to extend the definition of  $T$ . As  $n \rightarrow \infty$ ,  $T$  will be defined on all of  $[0, 1]$ . This map was constructed by J. von Neumann and S. Kakutani (1940, unpublished), and is known as the *von Neumann - Kakutani adding machine*.

- (a) Graph  $T$  as a function  $T : [0, 1] \rightarrow [0, 1]$ . Note that  $T$  is continuous on intervals of the form  $[1 - 1/2^n, 1 - 1/2^{n+1}]$ ,  $n = 1, 2, \dots$  and that the graph has a certain type of self-similarity.

(b) If we label the intervals  $[0, 1/2), [1/2, 1 - 1/2^2), \dots, [1 - 1/2^n, 1 - 1/2^{n+1}) \dots$  alternately “1” and “0” and transfer this labeling to the stack, show that the resulting sequence is the Toeplitz sequence 1 0 1 1 1 0 ....

4. The Morse sequence can be obtained from a stacking transformation in a similar way. Define  $S : [0, 1] \rightarrow [0, 1]$  by mapping  $[0, 1/4)$  linearly onto  $[3/4, 1)$  and mapping  $[1/2, 3/4)$  linearly onto  $[1/4, 1/2)$ . Now we stack using 2 columns, so that the base consists of the two intervals:  $[0, 1/4)$  and  $[1/2, 3/4)$ , and points are mapped vertically upwards. Continue the construction by cutting each column into two equal columns. Place the right stack of the 2nd column above the left stack of the first column, and the right stack of the 1st column above the left stack of the 2nd column, and map vertically above. If we continue in exactly this way, eventually  $S$  will be defined at every point except  $x = 1$  and  $x = 1/2$ . This transformation is called the *Morse automorphism*.

(a) Graph  $S$  as a map on  $[0, 1)$  and note the self-similarity.

(b) Let  $\beta = \{[0, 1/2), [1/2, 1)\} = \{A_0, A_1\}$  be a two set partition of  $[0, 1)$ . Define a sequence  $(z_n)$  by following the orbit of 0 under  $S$ , i.e.,

$$z_n = 0 \text{ if } S^n(0) \in A_0, \text{ and } z_n = 1 \text{ if } S^n(0) \in A_1.$$

Show that  $(z_n) = u$ , the Thue-Morse sequence.

5\*. Construct a stacking transformation that gives rise to the sequence defined by the substitution on  $\mathcal{A} = \{0, 1, 2\}$ :  $\theta(0) = 010$ ,  $\theta(1) = 121$ ,  $\theta(2) = 202$ . (Hint: Generalize the construction of the Morse automorphism, but use three columns - see [59] and [60]).

6\*. Construct a stacking transformation using a single column that gives rise to the positive paperfolding sequence.

**Remarks.** The von Neumann - Kakutani adding machine and the Morse automorphism are important examples in ergodic theory, of *invertible ergodic measure preserving transformations*. Roughly speaking, this means that they are essentially one-to-one and onto, they preserve *Lebesgue measure* (sets get mapped to sets of the

same length), and they have no non-trivial invariant sets (see [127] or [102] for more precise definitions). These maps have no fixed points, so cannot be chaotic. However, they are related to the substitution dynamical systems to be studied in subsequent chapters, and have a particular type of chaotic behavior.

## CHAPTER 16

### Fractals Arising from Substitutions.

In this chapter we look at the fractal nature of substitutions and study a surprising connection between the Thue-Morse sequence and the Koch curve. If we use the Morse substitution to trace out a path in the plane with the 0's and 1's being interpreted as a particular direction, the result is the Koch curve. In Section 16.2, we look at *Dragon curves*, which were first considered by J. E. Heighway. These arise from paper-folding sequences, resulting in interesting fractals and tilings of the plane. They are generated by substitutions, and we shall see that they also arise from iterated function systems.

One-dimensional substitutions can be regarded as giving rise to tilings of  $\mathbb{R}$ . The fractals arising from two-dimensional substitutions are of interest as they can be interpreted as non-periodic tilings of the plane. In Chapters 18 and 19 we look at the dynamical systems that one-dimensional substitutions generate. In an analogous way, two-dimensional substitutions generate “tiling dynamical systems”. However, we shall not pursue the theory of two-dimensional substitutions in this text.

In Section 16.4, we define the Rauzy fractal using the tribonacci substitution, which is a generalization of the Fibonacci substitution. The study of the tribonacci substitution was initiated by Gérard Rauzy, with the aim of gaining a better understanding of substitutions of non-constant length.

Our approach in this chapter is mostly non-rigorous, and example oriented. Chapter 16 follows directly from Chapter 15, and is independent of the remaining chapters of this text.

#### 16.1 A Connection Between the Morse Substitution and the Koch Curve.

We have already seen a type of self similarity arising from the definition of a substitution, so it is not surprising that substitutions can give rise to interesting fractals.

**Example 16.1.1** 1. Consider the Thue-Morse sequence written using the alphabet  $\mathcal{A} = \{a, b\}$ :

$$\theta(a) = a b, \quad \theta(b) = b a.$$

This gives rise to the sequence

$$abbabaabbababbabba\dots$$

Replacing  $ab$  by  $A$  and  $ba$  by  $B$ , we then obtain:

$$ABBAABAAB\dots,$$

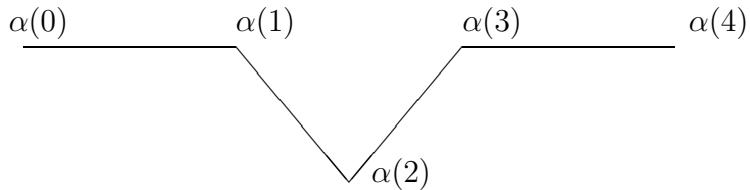
essentially the same sequence using a different symbol, i.e., the Morse substitution has a self-similarity property, despite its non-periodicity. This self-similarity can be seen for any substitution  $\theta$  having a fixed point  $u$ :  $\theta(u) = u$ .

2. Various authors have shown (independently), that there is a connection between the Morse substitution and the Koch curve (see [36], [2] and [86]). We will present results from the latter two papers to show how the Morse substitution can be used to construct the Koch curve. Recall that the Koch curve is constructed by starting with an equilateral triangle of side length 1 unit. Now add equilateral triangles of side length  $1/3$  to the middle of each side of the triangle (see Section 10.1). Continue indefinitely in this way, adding equilateral triangles on each line segment.

Let  $u = (u_n) = 0110100110010110\dots$ , be the Thue-Morse sequence, so  $u_0 = 0$ ,  $u_1 = 1$ ,  $u_{2n} = u_n$  and  $u_{2n+1} = R(u_n)$ , where  $R(0) = 1$ ,  $R(1) = 0$ . Set

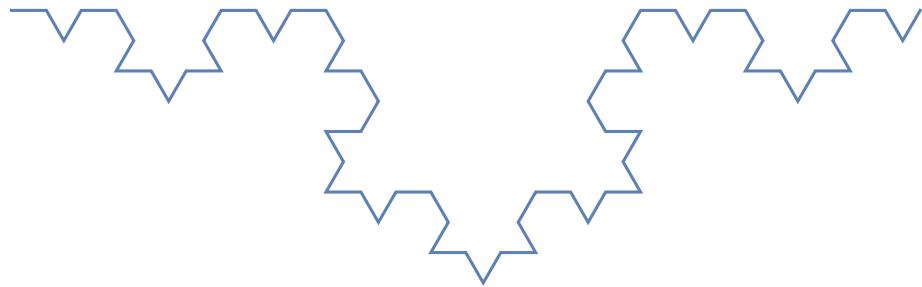
$$\alpha(n) = \sum_{k=0}^{n-1} (-1)^{u_k} e^{2\pi i k/3} = 1 - e^{2\pi i/3} - e^{4\pi i/3} + e^{6\pi i/3} - e^{8\pi i/3} + \dots + (-1)^{u_{n-1}} e^{2\pi(n-1)i/3},$$

and plot these points in the complex plane for  $n = 1, 2, \dots$ . We start at the origin  $(0, 0)$  and set  $\alpha(0) = 0$ . Then  $\alpha(1) = 1$ . We represent this geometrically by drawing a line from  $(0, 0)$  to  $(1, 0)$ . Now  $\alpha(2) = 1 - e^{2\pi i/3}$ , so we move 1 unit down to the right, at an angle of  $\pi/3$  to the  $x$ -axis, joining  $\alpha(1)$  to  $\alpha(2)$  with a straight line. We continue in this way with  $\alpha(3) = 1 - e^{2\pi i/3} - e^{4\pi i/3}$ ,  $\alpha(4) = 1 - e^{2\pi i/3} - e^{4\pi i/3} + 1$ . This gives the first step in the construction of one-side of the Koch snow-flake.



The first steps in the construction of the Koch Snowflake.

Continuing thus we plot  $\alpha(k)$ ,  $0 \leq k < 2^{2n}$ , giving a polygonal approximation to the bottom edge of the Koch Snowflake:



An approximation to the Koch Snowflake, generated by the Morse substitution.

If we scale by a factor of  $1/3$  at each stage of the construction, convergence to the Koch curve results. This gives an easy way of graphing the first steps in the approximation to the Koch Snowflake using a computer algebra system - we simply plot the points of the function  $\alpha(n)$ , then join them to form a polygonal path.

A slightly different construction is given in [86], but the resulting limiting curve is again the Koch curve. Let  $\Sigma = \{F, L\}$  be the alphabet where  $F$  denotes a move of one unit forward, and  $L$  is an anticlockwise rotation through an angle  $\phi = \pi/3$ . Denote by  $\theta$  the Morse substitution.

**Definition 16.1.2** The *Thue-Morse turtle programs of degree n*, denoted by  $TM_n$  and  $\overline{TM}_n$  are defined to be the following words in  $\Sigma^*$ :

$$TM_n = \theta^n(F) \quad \text{and} \quad \overline{TM}_n = \theta^n(L),$$

so that

$$TM_n = FLLFLFLFFLLFFLFLFLLF \dots FLLF,$$

the first  $2^n$  terms of the Thue-Morse sequence using the alphabet  $\Sigma$ .

If we interpret this sequence as a polygonal path in the plane, then when the sequence is plotted, it does not give the usual approximation to the Koch curve, but the limiting curve can be shown to be the Koch curve.

## Exercises 16.1

1. Let  $u = u_0u_1\dots$  be the Thue-Morse sequence. To illustrate the fractal nature of this sequence, show that deleting the odd terms:  $u_1, u_3, \dots$ , results in the original sequence.
2. Illustrate the self-similarity of the fixed point of the following substitutions (as in Example 16.1.1), by replacing groups of symbols by a single symbol to obtain the same sequence:
  - (i)  $\mathcal{A} = \{a, b\}$ ,  $\theta(a) = ab$ ,  $\theta(b) = a$ .
  - (ii)  $\mathcal{A} = \{a, b, c\}$ ,  $\theta(a) = ab a$ ,  $\theta(b) = b c b$ ,  $\theta(c) = c a c$ .
3. Write down a substitution on  $\mathcal{A} = \{0, 1\}$  that gives rise to the Cantor set (associate 0 and 1 with suitable length black and white lines in  $[0, 1]$ ).

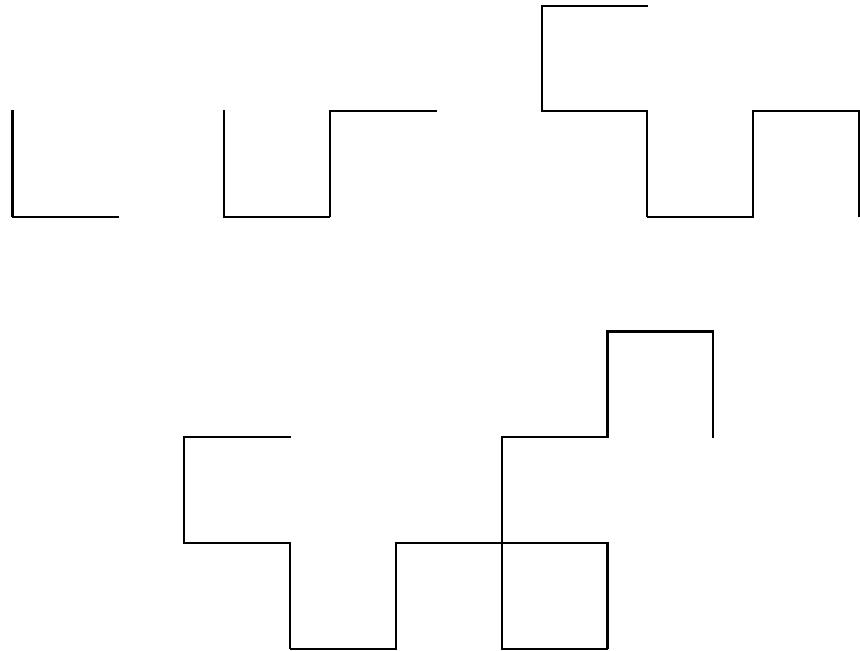
4\*. Exercises of the following type can be used to show that the sequence  $\alpha(n)$  arising from the Thue-Morse sequence, generates the Koch Snowflake curve:

- (a) Show that  $\alpha(2^{2n}) = 3^n$  for  $n \in \mathbb{N}$ . (Hint: Use the fact that the sequence  $u_0u_1\dots u_{2^{2n}-1}$  reads the same forwards and backwards (called a *palindrome*) - see Exercises 15.1).
- (b) Show that the polygonal curve obtained by joining  $\{\alpha(0), \alpha(1), \dots, \alpha(2^{2m})\}$  with straight lines, is symmetric about the vertical line through  $\alpha(2^m)$ .
- (c) If the polygonal curve joining  $\{\alpha(0), \alpha(1), \dots, \alpha(2^m)\}$  is translated by  $\alpha(2^m)$ , and rotated through  $-\pi/3$  radians, show that it coincides with the polygonal curve joining  $\{\alpha(2^m), \alpha(2^m + 1), \dots, \alpha(2^{2m})\}$ .

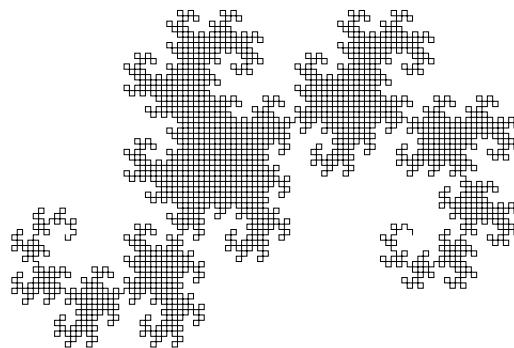
## 16.2 Dragon Curves.

The physicist John E. Heighway used the paperfolding sequence to construct *dragon curves*. Fold our sheet of paper to form the paperfolding sequence of  $\vee$ 's and  $\wedge$ 's as described in Section 15.4, but then open out the folded paper so that each of the corners forms a  $90^\circ$  angle. Look at a cross section of the paper, then we see the first few approximations to the dragon curve. Continuing to carry out the folding will give curves that start to resemble a dragon - called the *Heighway dragon*. The

first four dragon curves arising from the positive paper folding sequence, using one through four folds are given below.



The first four steps in the construction of a Dragon Curve.



A Dragon Curve.

Since the curve is generated by paperfolding, it never crosses itself. It will ultimately tile the whole plane. An iterated function system with  $f_1$  and  $f_2$  defined

below will also result in the dragon curve (see Section 10.5).

$$f_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad f_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1/2 & -1/2 \\ 1/2 & -1/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

### 16.3 Fractals Arising from Two-Dimensional Substitutions

In this section, we briefly look at various examples of 2-dimensional substitutions, and indicate a connection with tiling dynamical systems and certain types of fractals. Substitutions can be used to model tiling dynamical systems (see [108], [46]), where instead of using numbers or letters, we use colored unit square tiles. We start by showing how to construct a tiling resulting from a 2-dimensional substitution of constant length.

**Examples 16.3.1** Consider a map  $\zeta$  defined on  $\mathcal{A} = \{0, 1\}$  as follows:

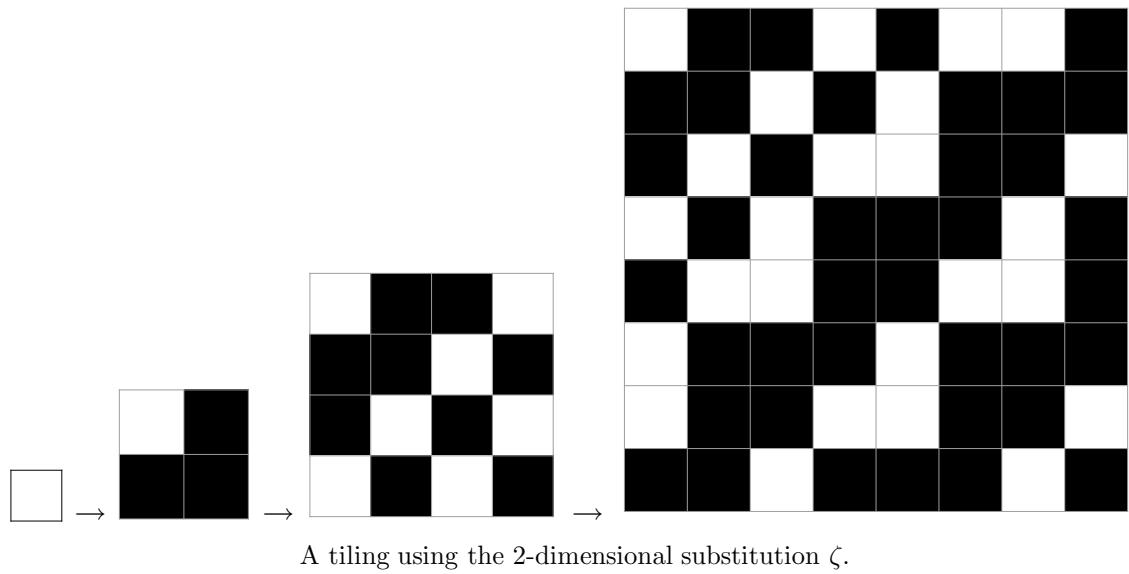
$$\zeta(0) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad \zeta(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

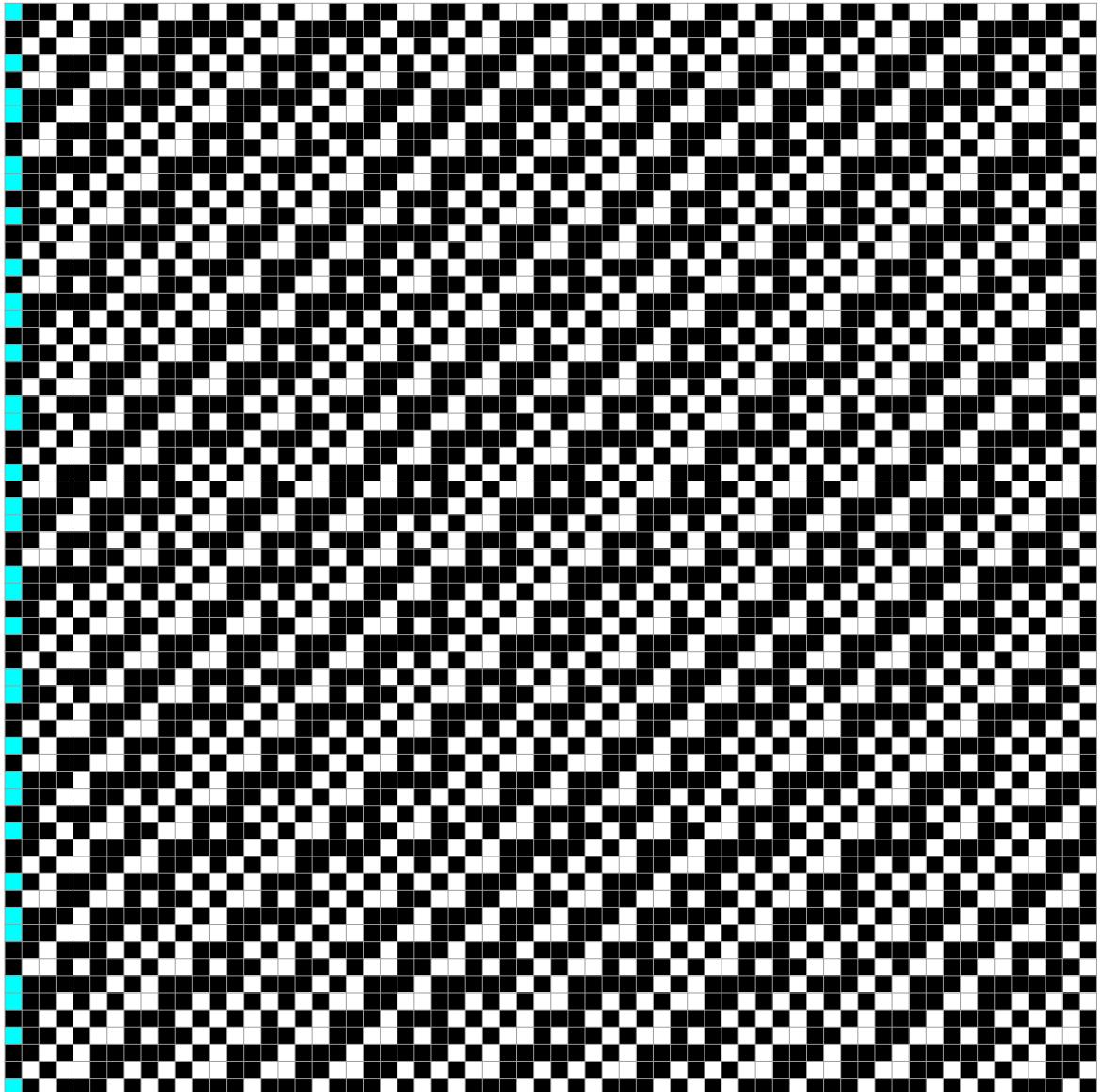
Using the homomorphism property  $\zeta^2(0) = \begin{pmatrix} \zeta(0) & \zeta(1) \\ \zeta(1) & \zeta(1) \end{pmatrix}$ , gives

$$\zeta^2(0) = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad \zeta^3(0) = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

A fixed point in the 2-dimensional sequence space results when we have a 0 in the top left-hand corner (or bottom left if we wish to tile the first quadrant starting at the origin) of the matrix  $\zeta(0)$ . As  $n \rightarrow \infty$ , we obtain a 2-dimensional array of 0's and 1's, fixed by the substitution  $\zeta$ . It is interesting to illustrate this geometrically by replacing '0' by a white unit square and '1' by a black unit square, then iterating to obtain a "*tiling*" of the plane using black and white squares. Repeated iteration emphasizes the fractal nature of the tiling. We expand each colored tile by some integer  $n > 1$ , and then subdivide it into  $n^2$  colored unit squares so that in the limit we have tiled the plane (or rather a quarter plane) using unit squares. A tiling of the whole plane may be obtained by extending the substitution in a natural way. Since the sizes of the matrices in the definition of  $\zeta$  are the same,  $\zeta$  is a 2-dimensional substitution of constant length. It is interesting to notice that such tiles often give

rise to a stereogram whose image depends on how the tile is viewed (see [47]). The resulting tiling using the substitution  $\zeta$  is then:





Eventually a tiling of the first quadrant is obtained.

### 16.3.2 Substitution Tilings.

The table and chair tilings are a type of substitution first introduced and studied by B. Solomyak [118] and then by E. A. Robinson [109], as examples of self-similar tilings of the plane. The table and chair tiles are examples of *rep-tiles* (short for

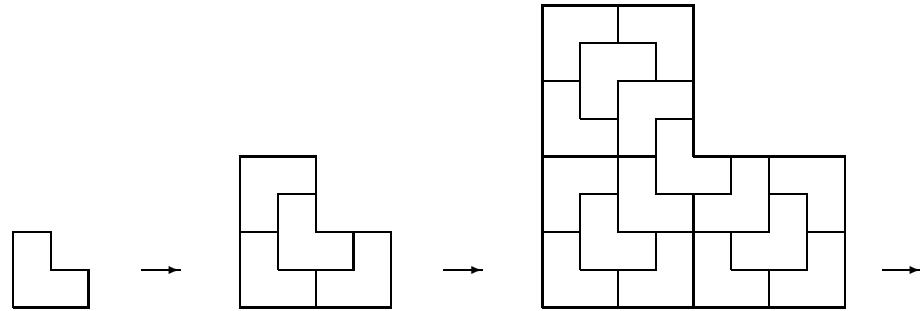
*replicating tile*, introduced by S. Golomb in 1966 [58]), which are polygons that can be tiled by a finite number of smaller, identically shaped copies of themselves. The table and chair rep-tiles are also *polyominoes*, in the sense that they can be written as the union of squares. Golomb showed that if a polyomino tiles itself, then it can tile a quadrant. We start with a brief description of tilings of the plane.

Although tiles can be defined more generally (for example with fractal boundaries such as dragon curves, or in  $\mathbb{R}^d$  for some  $d > 2$ ), by a *tile* we shall mean a connected polygon in  $\mathbb{R}^2$ . Two tiles are *equivalent* if they differ by a translation of the plane. A finite set of inequivalent tiles  $\mathcal{T} = \{T_1, T_2, \dots, T_m\}$ , is called the *prototile set*. The sets  $T_i$  are closed and bounded subsets of  $\mathbb{R}^2$  (compact sets), which have the property that the intersection  $T_i^\circ \cap T_j^\circ$  is the empty set for  $i \neq j$ . (The set  $T^\circ$  is the *interior* of  $T$ , being the largest open set contained in  $T$ ). We have a linear map  $Q : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , which is expanding in the sense that its eigenvalues have absolute value greater than one and is called the *inflation factor*, and we have a rule which tells us how to dissect each scaled tile  $QT_i$  into copies of the original prototiles  $T_1, T_2, \dots, T_m$ . It is also required that the prototiles fit “edge to edge” as we see for the squares in the chair and table substitution, and other examples.

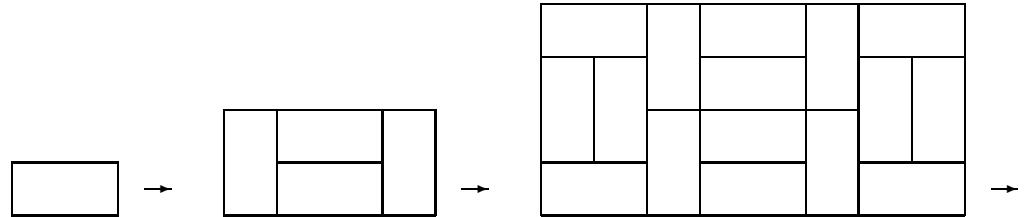
**Examples 16.3.3** 1. The simplest examples are periodic tilings of the plane by rectangles, parallelograms or triangles. For example, using a square as a single prototile, doubling its size and subdividing into four equal squares, and repeating this process indefinitely, gives rise to a tiling of  $\mathbb{R}^2$ . This is a periodic substitution since it has translational symmetry. Without the colorings given in Example 16.3.1, we would have a simple tiling using squares.

2. Probably the most famous examples of tilings were discovered by Roger Penrose in 1973-74, and now known as *Penrose tilings* ([54], [100]). It has been shown that the corresponding prototiles can be put together to form a tiling of the plane, but this cannot be done periodically (we call these tilings *aperiodic*). Penrose tilings have a five-fold symmetry and have been much studied. It is claimed that their discovery helped motivate the discovery of quasicrystals. In 1984, physicists Shechtman, Blech, Gratias and Cahn discovered a metal alloy whose X-ray spectrum has a five-fold rotational symmetry, but no translational symmetry. Despite this work being widely ridiculed (the two-time Nobel prize winner Linus Pauling is quoted as saying “There is no such thing as quasicrystals, only quasi-scientists”), Dan Schechtman received the Nobel prize for Chemistry in 2011.

**3. The Table and Chair Tilings.** There are four chair prototiles (four rotations of the chair), and two table prototiles (two rotations of the table), each giving rise to an aperiodic tiling of the plane [108]. For example, at the first stage the chair is “doubled” in size and partitioned into four congruent (but rotated), copies of itself, and this inflating and subdividing is continued indefinitely to obtain a tiling of the first quadrant of  $\mathbb{R}^2$ .



The chair tiling substitution.



The table tiling substitution.

4. Two-dimensional substitutions are defined in an analogous way to one-dimensional substitutions. Start with a finite alphabet  $\mathcal{A}$  consisting of at least two letters. A *block* (a generalization of the notion of word in the one-dimensional case), is a 2-dimensional array of member of  $\mathcal{A}$  (we can think of it as a matrix with finitely many rows and columns, but without the brackets). The set of all infinite 2-dimensional such arrays is

$$\begin{array}{cccccc} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{3,0} & a_{3,1} & a_{3,2} & a_{3,3} & \dots \\ a_{2,0} & a_{2,1} & a_{2,2} & a_{2,3} & \dots \\ a_{1,0} & a_{1,1} & a_{1,2} & a_{1,3} & \dots \\ a_{0,0} & a_{0,1} & a_{0,2} & a_{0,3} & \dots \end{array}$$

denoted by  $\mathcal{A}^{\mathbb{N}^2}$ , the set of all infinite matrices of the form:

Denote by  $\mathcal{A}^+$  the set of all finite blocks, then a substitution is a map  $\zeta : \mathcal{A} \rightarrow \mathcal{A}^+$ , which replaces a letter by a block. We assume that the iterates are expanding in both the horizontal and vertical directions and the definition on each letter is conformable with the iteration. For example define  $\zeta$  on  $\mathcal{A} = \{a, b, c, d\}$  by

$$\zeta(a) = \begin{bmatrix} c & d \\ a & b \end{bmatrix}, \quad \zeta(b) = \begin{bmatrix} c \\ a \end{bmatrix}, \quad \zeta(c) = \begin{bmatrix} a & b \end{bmatrix}, \quad \zeta(d) = a.$$

Note that  $\zeta(a)$  has an  $a$  in the bottom left hand corner. This ensures that  $\lim_{n \rightarrow \infty} \zeta^n(a)$  generates a fixed point of  $\mathcal{A}^{\mathbb{N}^2}$ .  $\zeta$  is an example of a substitution of non-constant length. The chair substitution can be “modeled” by the following substitution of constant length (the size of each block is the same for each  $\alpha \in \mathcal{A}$ , see [108]):

$$\xi(a) = \begin{bmatrix} d & a \\ a & b \end{bmatrix}, \quad \xi(b) = \begin{bmatrix} b & c \\ a & b \end{bmatrix}, \quad \xi(c) = \begin{bmatrix} d & c \\ c & b \end{bmatrix}, \quad \xi(d) = \begin{bmatrix} d & c \\ a & d \end{bmatrix}.$$

### Exercises 16.3

1. A 2-dimensional substitution of non-constant length  $\zeta$  is defined by

$$\zeta(1) = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}, \quad \zeta(2) = \begin{bmatrix} 1 & 3 \end{bmatrix}, \quad \zeta(3) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \zeta(4) = [1].$$

Write down  $\zeta^n(1)$ , for  $n = 2, 3, 4$ , and if  $\zeta^n(1)$  is an  $M \times M$  matrix, find  $M$ .

2. Write down a 2-dimensional substitution that generates the Sierpinski carpet. Generalize this to find a 3-dimensional substitution that generates the Menger sponge.

3. Let  $\theta$  and  $\phi$  be two 1-dimensional substitutions defined on alphabets  $\mathcal{A}$  and  $\mathcal{B}$ . The *direct product*  $\zeta$  of  $\theta$  and  $\phi$  is defined as follows:

If  $\theta(a) = a_1a_2 \dots a_n$  and  $\phi(b) = b_1b_2 \dots b_m$ , then

$$\zeta(a, b) = \begin{bmatrix} (a_1, b_m) & (a_2, b_m) & \cdots & (a_n, b_m) \\ \vdots & \vdots & \cdots & \vdots \\ (a_1, b_2) & (a_2, b_2) & \cdots & (a_n, b_2) \\ (a_1, b_1) & (a_2, b_1) & \cdots & (a_n, b_1) \end{bmatrix}, \text{ for each } (a, b) \in \mathcal{A} \times \mathcal{B}.$$

- (a) If  $\theta(0) = 01$  and  $\theta(1) = 10$  is the Morse substitution, find the direct product of  $\theta$  with itself.
- (b) For the Fibonacci substitution written as  $\theta(a) = ba$ ,  $\theta(b) = a$ , find its direct product with itself. Using the recoding  $(a, a) = 1$ ,  $(a, b) = 2$ ,  $(b, a) = 3$  and  $(b, b) = 4$ , show that the resulting 2-dimensional substitution is the same as the substitution  $\zeta$  in exercise 1 above.
4. Let  $\theta$  be a one-dimensional substitution of constant length  $q$  on an alphabet  $\mathcal{A}$ . Define functions  $\phi_j : \mathcal{A} \rightarrow \mathcal{A}$ ,  $(0 \leq j \leq q - 1)$ , by  $\phi_j(\alpha) = [\theta(\alpha)]_j$ . For example, if  $\theta$  is the Morse substitution  $\theta(0) = 01$ ,  $\theta(1) = 10$ , then  $\phi_0(0) = [\theta(0)]_0 = 0$ ,  $\phi_0(1) = [\theta(1)]_0 = 1$ ,  $\phi_1(0) = [\theta(0)]_1 = 1$  and  $\phi_1(1) = [\theta(1)]_1 = 0$ .

The substitution  $\theta$  is said to be *bijective* if the maps  $\phi_j$  are bijective (one-to-one and onto).  $\theta$  is *commutative* if the maps  $\phi_j$  commute. Clearly the Morse substitution is both bijective and commutative.

Determine whether or not the following substitutions are bijective or commutative.

		$\theta(0) = 013$	
(i)	$\theta(0) = 010$	$\theta(0) = 021$	$\theta(0) = 012$
	$\theta(1) = 121$	$\theta(1) = 100$	$\theta(1) = 102$
	$\theta(2) = 202$	$\theta(2) = 212$	$\theta(2) = 231$
			$\theta(2) = 320$
			$\theta(2) = 211$

**Remark 16.3.4** For a bijective substitution of constant length  $q$ , it may be assumed that  $\phi_0$  = identity map, and we can associate with the substitution a group  $\{\phi_j : 0 \leq j \leq q - 1\}$ , where the group operation is composition of functions. This group is abelian when the substitution is commutative. For example, in (ii) above the maps are permutations on the set  $\{0, 1, 2\}$ , and the resulting group is the permutation group  $S_3$  (see [104]).

## 16.4 The Rauzy Fractal.

The Rauzy fractal is a fractal subset of the plane  $\mathbb{R}^2$  associated with the *tribonacci substitution*. The tribonacci substitution  $t$ , is defined on the alphabet  $\{1, 2, 3\}$  by:

$$t(1) = 12, \quad t(2) = 13, \quad t(3) = 1.$$

The tribonacci substitution was introduced by Gérard Rauzy in 1981, [106] as a generalization of the Fibonacci substitution, in order to gain a better understanding of substitutions of non-constant length. This fractal gives rise to a tiling of the plane, and an example of a tiling dynamical system (see [6], [45], [106] and [15]). If  $t_n = t^n(1)$ , then we have

$$t_0 = 1, \quad t_1 = 12, \quad t_2 = 1213, \quad t_3 = 1213121, \quad t_4 = 1213121121312,$$

In this case  $u_t = \lim_{n \rightarrow \infty} t^n(1)$ , is a fixed point of the substitution.

**16.4.1 The Incidence Matrix of a Substitution.** In order to define the *Rauzy fractal*, we digress with some general ideas concerning substitutions.

**Definition 16.4.2** (a) Let  $\zeta$  be a substitution on the alphabet  $\mathcal{A} = \{1, \dots, d\}$ . The *incidence matrix* of the substitution is the matrix  $M_\zeta$  defined as follows. Let  $\alpha \in \mathcal{A}$  and let  $B$  be a word using the alphabet  $\mathcal{A}$ . Denote by  $|B|_\alpha$  the number of occurrences of  $\alpha$  in  $B$ . Write  $\ell_{\alpha,\beta} = |\zeta(\beta)|_\alpha$ , then the *incidence matrix* of the substitution  $\zeta$  is

$$M_\zeta = [\ell_{\alpha,\beta}],$$

a  $d \times d$  matrix with entries from  $\mathbb{N}$ .

(b) A  $d \times d$  matrix  $M = [m_{ij}]$  is *non-negative* if  $m_{ij} \geq 0$ , for all  $1 \leq i, j \leq d$ . If  $M$  is non-negative, then  $M$  is said to be *primitive* if there exists  $k \in \mathbb{Z}^+$ , such that every entry of  $M^k$  is positive (i.e., if  $M^k = [m_{ij}^{(k)}]$ , then  $m_{ij}^{(k)} > 0$  for all  $1 \leq i, j \leq d$ , written  $M^k > 0$ ). If the incidence matrix  $M_\zeta$  of the substitution  $\zeta$  is primitive, then  $\zeta$  is said to be a *primitive substitution*.

It will be seen in Section 18.4 that primitive substitutions have interesting dynamical properties. In particular, they always have a fixed point, or point periodic under the substitution map. Another important property of primitiveness is that we can apply the *Perron-Frobenius Theorem* to the incidence matrix (see [104] for a proof of this theorem).

**16.4.3 The Perron-Frobenius Theorem.** *Let  $M$  be a primitive matrix. Then*

(a)  *$M$  has a largest positive eigenvalue  $\gamma$ :  $\gamma > |\lambda|$  for any other eigenvalue  $\lambda$ .*

- (b) *There is a positive eigenvector corresponding to  $\gamma$  (all entries are positive).*  
(c)  *$\gamma$  is a simple eigenvalue (algebraic multiplicity equal to one).*

**Examples 16.4.4** 1. If  $\theta(0) = 01$ ,  $\theta(1) = 10$  is the Morse substitution, then  $M_\theta = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ . Clearly  $M_\theta$  is a primitive matrix.

2. For the Fibonacci substitution,  $\theta(0) = 01$ ,  $\theta(1) = 0$ ,  $M_\theta = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ .  $M_\theta^2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ , so  $\theta$  is a primitive substitution. The largest positive eigenvalue is  $\gamma = (1 + \sqrt{5})/2$ , with corresponding eigenvector  $(1, 1/\gamma)$  (see Exercises 16.4 # 3).

This same exercise shows that for  $n$  large, if  $F_n$  is the  $n$ th Fibonacci number,  $F_n \approx \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^n = \gamma^n / \sqrt{5}$  (approximately equal), so that  $\lim_{n \rightarrow \infty} (\alpha \gamma^n \bmod 1) = 0$ , where  $\alpha = 1/\sqrt{5}$ . This is typical of *Pisot numbers*, which we define below.

3. The *Chacon substitution* is defined by  $\theta(0) = 0010$ ,  $\theta(1) = 1$ . In this case  $M_\theta = \begin{bmatrix} 3 & 0 \\ 1 & 1 \end{bmatrix}$ , and we can check that it is not primitive.

4. In the case of the tribonacci substitution,  $M_t = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ , and we can check that  $M_t^3 > 0$ , so  $t$  is primitive. The eigenvalues satisfy the equation  $z^3 - z^2 - z - 1 = 0$ . A computer algebra system shows the eigenvalues are given by

$$\rho = \frac{1}{3} \left( 1 + (19 - 3\sqrt{33})^{1/3} + (19 + 3\sqrt{33})^{1/3} \right) = 1.83928\dots$$

$\rho$  is a positive eigenvalue which is greater than 1. The other two eigenvalues occur in conjugate pairs, with absolute values less than 1 (see Exercises 16.4 # 4):

$$\lambda = -0.419643 + 0.606291i, \quad \bar{\lambda} = -0.419643 - 0.606291i, \quad |\lambda| = 0.737353.$$

Recall that a *real algebraic number*  $\gamma$  is a real root of a monic irreducible integer polynomial. A monic irreducible integer polynomial is a polynomial whose coefficients are integers with the highest power having coefficient 1, and which cannot be factored into similar polynomials of lesser degree. Pisot numbers, named after Charles Pisot, are real algebraic numbers which are greater than 1, all of whose conjugate elements

have absolute value less than 1. In other words,  $\gamma$  is a real root of a polynomial of the above type which is greater than 1, such that all the other roots (which are real or appear in conjugate pairs) have absolute value less than one. It turns out that Pisot numbers always result from primitive matrices that have determinant equal to  $\pm 1$ . They also have the following characterization due to Charles Pisot, which says (roughly speaking)  $\alpha\lambda^n \approx M \in \mathbb{N}$  for  $n$  large:

**Proposition 16.4.5** *A real algebraic number  $\lambda > 1$  is a Pisot number if and only if there exists  $\alpha$  satisfying*

$$\lim_{n \rightarrow \infty} (\alpha\lambda^n \bmod 1) = 0.$$

It follows that if  $\gamma = (1 + \sqrt{5})/2$ , then  $\gamma$  is a Pisot number.

#### 16.4.6 The Fibonacci Quasicrystal.

Consider the Fibonacci substitution, written as  $\theta(1) = 12$ ,  $\theta(2) = 1$ . We plot points in  $\mathbb{R}^2$  according to the fixed point of  $\theta$ , using the standard basis of  $\mathbb{R}^2$ ,  $e_1 = (1, 0)$  and  $e_2 = (0, 1)$ .

Suppose that  $w = w_1w_2 \dots w_n$  is a word using the alphabet  $\mathcal{A} = \{1, 2\}$ . Define a map  $P : \mathcal{A}^n \rightarrow \mathbb{R}^2$  by

$$P(w_1w_2 \dots w_n) = e_{w_1} + e_{w_2} + \dots + e_{w_n},$$

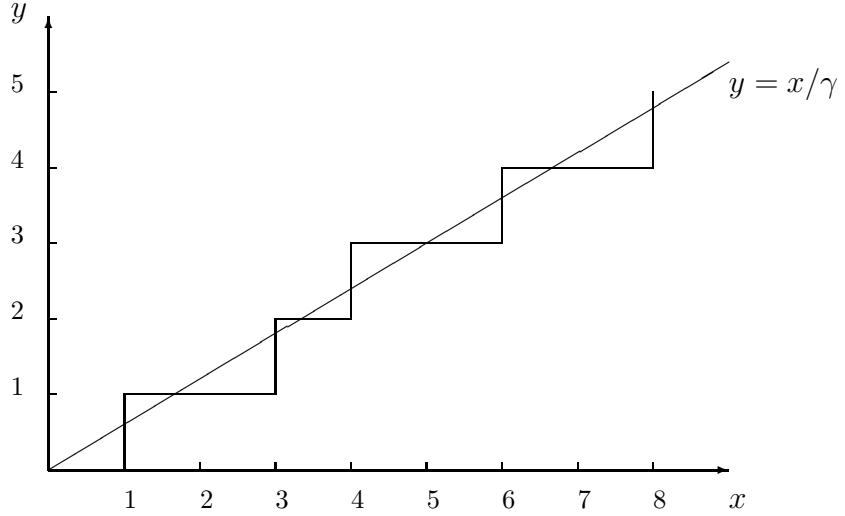
using the same notation for any  $n \in \mathbb{Z}^+$ . We plot these points in the plane as follows:

Consider the sequence obtained from  $\theta$ :  $u = 1211212112112\dots$  and interpret each 1 as indicating movement of one unit horizontally and 2 as movement one unit vertically. Start at the origin  $(0, 0)$ . When we meet a 1, we move one unit in the positive  $x$ -direction. When we meet a 2, we move one unit in the positive  $y$ -direction. Two 1's mean that we move two units in the  $x$ -direction. Thus the points are:

$$(0, 0) \rightarrow (1, 0) \rightarrow (1, 1) \rightarrow (2, 1) \rightarrow (3, 1) \rightarrow (3, 2) \rightarrow (4, 2) \rightarrow (4, 3) \rightarrow (5, 3) \rightarrow \dots$$

Join these points to form a polygonal path (staircase). They seem to oscillate around some fixed line, remaining a finite distance from that line. The incidence matrix of  $\theta$ ,  $M_\theta$  is a  $2 \times 2$  matrix having an eigenvalue  $\gamma$  with  $\gamma > 1$ . The other eigenvalue has absolute value less than 1. In Section 13.1 we studied the dynamics of such matrices, showing that as a map on  $\mathbb{R}^2$ ,  $M_\theta$  expands in the direction of the line  $y = x/\gamma$  (the eigenspace  $E_\gamma$ ), and will contract in the direction  $y = -\gamma x$  (the eigenspace  $E_{-\gamma}$ ). These directions are mutually orthogonal. Our aim is to show that the line

of oscillation for the substitution is  $E_\gamma$ , and we study the projection of the points generated by the substitution, onto the subspace  $E_{-1/\gamma}$ .



The Fibonacci Staircase.

Generally, suppose that  $\theta$  is a substitution on  $\mathcal{A} = \{1, 2, \dots, d\}$ . Let  $\{e_1, e_2, \dots, e_d\}$  be the standard basis for  $\mathbb{R}^d$ , writing  $e_i$  as a row vector (so  $e_1 = (1, 0, \dots, 0)$ ,  $e_2 = (0, 1, 0, \dots, 0), \dots, e_n = (0, 0, \dots, 1)$ ).  $P : \mathcal{A}^n \rightarrow \mathbb{R}^d$  is defined by  $P(w_1 w_2 \dots w_n) = e_{w_1} + e_{w_2} + \dots + e_{w_n}$ .  $P$  is thought of as a *linearization* of the substitution.

**Definition 16.4.7** The substitution  $\theta$  is a *unit Pisot irreducible substitution* if its incidence matrix  $M_\theta$  is primitive with determinant  $\pm 1$  and the dominant eigenvalue is a Pisot number.

The requirement that the incidence matrix  $M_\theta$  has a simple dominant eigenvalue (called a *Perron-Frobenius eigenvalue*), ensures that there is one real expanding eigenvalue  $\gamma$ , and all other eigenvalues are contracting. It can be shown that in this case  $\gamma$  is a Pisot number.

In the case of the tribonacci substitution  $\theta$ , we plot points in  $\mathbb{R}^3$  using the standard basis  $e_1 = (1, 0, 0)$ ,  $e_2 = (0, 1, 0)$  and  $e_3 = (0, 0, 1)$ .  $P(t_1 t_2 \dots t_n) = e_{t_1} + e_{t_2} + \dots + e_{t_n}$ , giving as “staircase” oscillating around a particular line. The matrix  $M_t$  has a dominant eigenvalue  $\rho > 1$ , the other eigenvalues being complex conjugates with absolute value less than one. The plotted points (the *stair of the sequence*), oscillate around the line given by the eigenspace  $E_\rho$ . The other two eigenvectors generate

a two-dimensional subspace  $\mathbb{H}_c$ , called the *contracting subspace*, (where unlike the Fibonacci case, it is not orthogonal to  $E_\rho$ ).

We project the points of the substitution  $\theta$  parallel to the unstable direction, onto  $\mathbb{H}_c$ . The closure of these projected points is called the Rauzy fractal.

In the following lemma, where we treat members of  $\mathbb{R}^d$  as column vectors, we see that  $P$  intertwines the incidence matrix and the substitution:

**Lemma 16.4.8** *Let  $M_\theta$  be the incidence matrix of the substitution  $\theta$  on  $\mathcal{A}$ . If  $w \in \mathcal{A}^n$ , then*

$$P(\theta(w)) = M_\theta P(w).$$

**Proof.** We use induction on the length of the word  $w$ . Suppose that  $|w| = 1$ , say  $w = i$ . Then  $P(w) = e_i$ , and  $M_\theta P(w) = M_\theta e_i$ . The latter product picks out the  $i$ th column of the incidence matrix. This column has first entry counting the number of 1's in  $\theta(i)$ , the second entry counts the number of 2's, etc.

On the other hand,  $P(\theta(w)) = P(\theta(i)) = e_{w_1} + e_{w_2} + \dots + e_{w_k}$  if  $\theta(i) = w_1 w_2 \dots w_k$ . Again, the first entry counts the number of 1's in  $\theta(i)$  etc, so these two expressions are equal.

Suppose that the identity holds for all words of length  $k$ . Let  $w = w_1 w_2 \dots w_k w_{k+1}$ . Then

$$\begin{aligned} M_\theta P(w) &= M_\theta(e_{w_1} + \dots + e_{w_k} + e_{w_{k+1}}) = M_\theta(e_{w_1} + \dots + e_{w_k}) + M_\theta(e_{w_{k+1}}) \\ &= P(\theta(w_1 w_2 \dots w_k)) + P(\theta(w_{k+1})), \end{aligned}$$

(using the induction hypothesis and the fact that the identity holds for words of length 1),

$$= P(\theta(w_1 w_2 \dots w_k) \theta(w_{k+1})) = P(\theta(w_1 \dots w_k w_{k+1})) = P(\theta(w)),$$

and so the result holds for all  $n \in \mathbb{Z}^+$ .

□

**Proposition 16.4.9** *If  $\theta$  is a unit Pisot irreducible substitution, then the stair of any fixed point (or periodic point) of  $\theta$ , oscillates around the expanding line, and remains within a fixed distance from it.*

**Proof.** We see from Lemma 16.4.6 that  $P(\theta^n(1)) = M_\theta^n P(1) = M_\theta^n(e_1)$ . Since the matrix  $M_\theta$  has a single expanding direction (the eigenspace  $E_\rho$  where  $\rho$  is the Perron-Frobenius eigenvalue), iterates of any vector become arbitrarily close to this line for  $n$  large enough. In particular, points associated with the Fibonacci word  $\theta^n(1)$  tend to  $E_\rho$ .

Suppose that  $w$  is an initial word in the fixed point  $u = \theta^\infty(1)$ . We can write  $w = \theta^k(i_k)\theta^{k-1}(i_{k-1})\dots\theta(i_1)i_0$ , for some finite sequence  $i_0, i_1, \dots, i_k$ , where each  $i_j$  is either 1 or  $\epsilon$ , the empty word (see the exercises). Then

$$\begin{aligned} P(w) &= P(\theta^k(i_k)\theta^{k-1}(i_{k-1})\dots\theta(i_1)i_0) \\ &= P(\theta^k(i_k)) + P(\theta^{k-1}(i_{k-1})) + \dots + P(\theta(i_1)) + P(i_0) \\ &= M_\theta^k(P(i_k)) + M_\theta^{k-1}(P(i_{k-1})) + \dots + M_\theta(P(i_1)) + P(i_0). \end{aligned}$$

The coordinates of the corresponding points are given by finite sums, bounded by geometric series, convergent in the contracting direction (for example, in the case of the tribonacci substitution, if we diagonalize the matrix  $M_t$ , sums involving  $M_t^k(e_1)$ , give rise to geometric series in each coordinate). The result now follows.  $\square$

#### 16.4.10 Construction of the Rauzy Fractal.

Let  $t$  be the tribonacci substitution, with incidence matrix  $M_t$  and eigenvalues  $\rho$ ,  $\lambda$  and  $\bar{\lambda}$  described above. From Exercises 16.4 # 5, we see that  $E_\rho = \{(1, 1/\rho, 1/\rho^2)a : a \in \mathbb{R}\}$ , is the expanding line, and  $\mathbb{H}_c$  (a copy of  $\mathbb{R}^2$ ), is the contracting plane. Denote by  $\pi : \mathbb{R}^3 \rightarrow \mathbb{H}_c$  the projection sending points from  $\mathbb{R}^3$  to  $\mathbb{H}_c$  in a direction parallel to  $E_\rho$ .

**Definition 16.4.11** The *Rauzy fractal*  $\mathcal{R}$ , is the closure in the plane  $\mathbb{H}_c$  of the set

$$\{\pi\left(\sum_{i=1}^n e_{w_i}\right) : n \in \mathbb{Z}^+\},$$

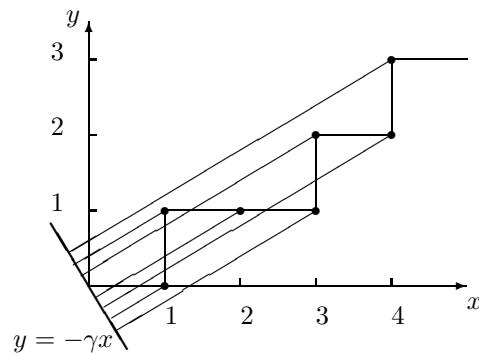
where  $w_1, w_2, \dots, w_n$  are the first  $n$  terms in the fixed point of the tribonacci substitution, and  $\{e_1, e_2, e_3\}$  is the standard basis of  $\mathbb{R}^3$ .

In other words, we project the points in  $\mathbb{R}^3$  generated by the tribonacci substitution, parallel to the unstable direction, onto the attracting plane. The same method can be used to generate fractals of any Pisot type substitution. If the substitution is defined using  $d$  symbols, the fractal (sometimes called a *central tile*), is contained in a *hyperplane*  $\mathbb{H}_c$ , (a  $(d-1)$ -dimensional subspace of  $\mathbb{R}^d$ ). The resulting fractal turns out to be a compact (closed and bounded) subset of  $\mathbb{R}^{d-1}$ , which is connected, with

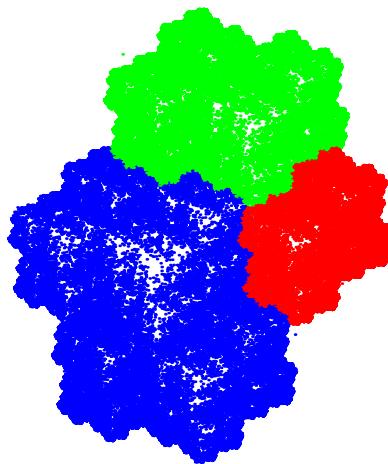
no holes (simply connected). In the case of the Fibonacci substitution, the resulting set is a closed bounded interval on the line  $y = -\gamma x$ .

**Theorem 16.4.12** *The Rauzy fractal (central tile)  $\mathcal{R}$ , for the tribonacci substitution  $\theta$ , is a compact subset of the plane. More generally, if  $\sigma$  is a unit Pisot irreducible substitution, the central tile  $T_\sigma$  associated with  $\sigma$ , is a compact set.*

**Proof.** By definition, the Rauzy fractal is closed, so it suffices to show that it is bounded. This follows directly from Proposition 16.4.9. □



The projection onto the contracting subspace.



The Rauzy Fractal  $\mathcal{R}$ .

**16.4.13 Additional Properties for the Rauzy Fractal.** We mention without proof some important properties of the tribonacci substitution and the Rauzy fractal (see [45], [107] and [6]).

1. When we talk about the Rauzy fractal  $\mathcal{R}$ , it is the boundary that has been shown to have a fractal nature, with a fractal dimension greater than 1.
2. In the picture of the Rauzy fractal,  $\mathcal{R}$  is made up of three distinct regions  $\mathcal{R}_j$ , for  $j = 1, 2, 3$ , where  $\mathcal{R}_j$  is the closure of the set of those projected points associated with the vector  $e_j$ :

$$\mathcal{R}_j = \overline{\left\{ \pi\left(\sum_{i=0}^n e_{u_i}\right) : u_i = j, n \in \mathbb{Z}^+ \right\}},$$

and  $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$ . Rauzy has shown that the overlap of the three sets is a set of measure zero.

3. The incidence matrix of the tribonacci substitution has characteristic equation  $z^3 - z^2 - z - 1 = 0$ , with a real root  $\rho > 1$  and two complex conjugate roots, say  $\alpha$  and  $\bar{\alpha}$ , so that  $\alpha^3 = \alpha^2 + \alpha + 1$ .  $\alpha$  is called a *Tribonacci number*.

If we consider the Rauzy fractal as a subset of  $\mathbb{C}$ , Rauzy has shown that

$$\mathcal{R} = \left\{ \sum_{i \geq 0} \epsilon_i \alpha^i : \epsilon_i \in \{0, 1\}, \epsilon_i \epsilon_{i+1} \epsilon_{i+2} = 0 \right\}.$$

Setting  $\epsilon_0 = 0$  gives  $\mathcal{R}_1$ ,  $\epsilon_0 \epsilon_1 = 10$  gives  $\mathcal{R}_2$ , and  $\epsilon_0 \epsilon_1 = 11$  gives  $\mathcal{R}_3$ .

Then it can be shown that

$$\mathcal{R}_1 = \alpha \mathcal{R}, \quad \mathcal{R}_2 = \alpha^2 \mathcal{R} + 1, \quad \mathcal{R}_3 = \alpha^3 \mathcal{R} + \alpha + 1,$$

so  $\mathcal{R}$  is partitioned by contractions of itself, and is a self-similar fractal. These three pieces admit both a periodic tiling and a non-periodic self-similar tiling of the contracting plane. They can be assembled in two ways to partition the Rauzy fractal, giving rise to a dynamical system that is an exchange of the three pieces. This dynamical system is semi-conjugate to the tribonacci substitution dynamical system (see 4).

4. We shall see in Chapter 18, that a substitution  $\theta$  generates a dynamical system in the following way. If  $u$  is a fixed point of  $\theta$ , and  $\sigma$  is the shift map, then  $\sigma$  restricted to the closure of the orbit of  $u$  under  $\sigma$ ,  $\overline{\mathcal{O}(u)}$ , is an invariant subset of the shift space.  $(\overline{\mathcal{O}(u)}, \sigma)$  is called a *shift dynamical system*. A slightly weaker condition than conjugacy, called *semi-topological conjugacy* is defined in Section 19.5. It can

be shown that the dynamical system  $(\overline{O(u)}, \sigma)$ , where  $u$  is the fixed point of the tribonacci substitution  $t$ , is semi-topologically conjugate to the dynamical system:

$$R_\rho : [0, 1) \times [0, 1) \rightarrow [0, 1) \times [0, 1), \quad R_\rho(x, y) = (x + \frac{1}{\rho}, y + \frac{1}{\rho^2}) \mod 1.$$

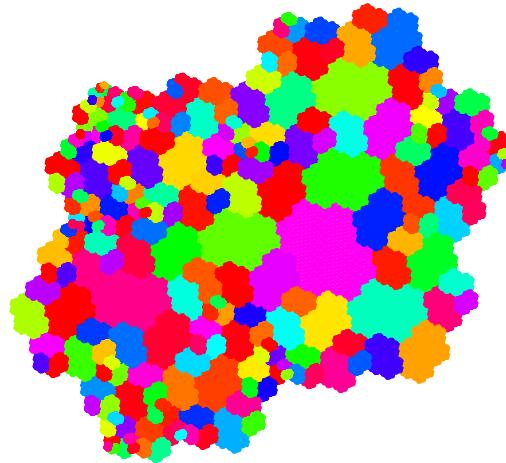
In the case of the Fibonacci substitution, the conjugacy is between the shift map and  $T_\gamma(x) = x + \gamma \mod 1$ , on  $[0, 1)$ , the rotation by the golden number.

5. We consider some properties of the contracting subspace  $\mathbb{H}_c$  for the case of the tribonacci substitution, and which also hold for general Pisot type substitutions. The incidence matrix  $M_t$  has eigenvalues  $\rho$ ,  $\alpha$  and  $\bar{\alpha}$  say. Denote the corresponding eigenvectors by  $u_\rho$ ,  $u_\alpha$  and  $u_{\bar{\alpha}}$ , where we may assume  $u_{\bar{\alpha}} = \bar{u}_\alpha$ .

Denote by  $M_t^T$  the transpose of the matrix  $M_t$ , a matrix having the same eigenvalues as  $M_t$ . However, its eigenvectors are different. Denote these by  $v_\rho$ ,  $v_\alpha$  and  $v_{\bar{\alpha}}$ . Since  $M_t^T$  is also a positive matrix, both of the eigenvectors  $u_\rho$  and  $v_\rho$  have positive entries, and can be chosen so that  $\langle v_\rho, u_\rho \rangle = 1$  (where  $\langle v, u \rangle$  denotes the inner product in a complex vector space). We claim that

$$\langle v_\rho, u_\alpha \rangle = 0, \quad \langle v_\rho, u_{\bar{\alpha}} \rangle = 0, \quad \langle v_\alpha, u_\rho \rangle = 0, \quad \langle v_{\bar{\alpha}}, u_\rho \rangle = 0.$$

$\mathbb{H}_c$  is the two-dimensional subspace of  $\mathbb{R}^3$ , generated by the eigenvectors  $\{u_\alpha, u_{\bar{\alpha}}\}$  of  $M_t$ , so is a plane containing the origin (strictly speaking,  $\mathbb{H}_c$  is generated by the vectors  $\text{Re}(u_\alpha)$  and  $\text{Im}(u_\alpha)$ ). We therefore see that the contracting space  $\mathbb{H}_c$  is orthogonal to the vector  $v_\rho$ , but  $\mathbb{H}_c$  is not orthogonal to  $u_\rho$ .



A tiling of the plane using Rauzy fractals of different sizes.

### Exercises 16.4

1. (a) Find the incidence matrices of the following substitutions using a suitable alphabet:
  - (i)  $\theta(0) = 01$ ,  $\theta(1) = 1$ ,
  - (ii)  $\theta(0) = 010$ ,  $\theta(1) = 121$ ,  $\theta(2) = 202$ ,
  - (iii)  $\theta(0) = 010$ ,  $\theta(1) = 111$ ,
  - (iv)  $\theta(1) = 12$ ,  $\theta(2) = 23$ ,  $\theta(3) = 23$ .
 (b) Determine the eigenvalues of the incidence matrices from (a).
   
 (c) Which of the matrices from (a) is primitive?
  
2. Let  $t$  be the tribonacci substitution defined on the alphabet  $\{1, 2, 3\}$ . If  $t_n = t^n(1)$ , show that for  $n > 2$ ,  $t_n = t_{n-1}t_{n-2}t_{n-3}$ . Deduce that  $|t^n(1)| = T_n$ , the  $n$ th tribonacci number (see Exercise 1.1 # 12, for the definition of  $T_n$ ).
  
3. Let  $(F_n)$  be the Fibonacci sequence. We saw in Exercise 1.1 # 12(b) that if  $v_n = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}$ , and  $F = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ , then  $v_{n+1} = F \cdot v_n$ ,  $n \geq 0$ .
  - (a) Show that the eigenvalues of the matrix  $F$  are of the form  $\gamma$  and  $-1/\gamma$ . Find the corresponding eigenvectors, and show that they are orthogonal. Deduce a diagonalization for  $F$ :
$$F = P \cdot \begin{bmatrix} \gamma & 0 \\ 0 & -1/\gamma \end{bmatrix} \cdot P^{-1}.$$
  
  - (b) Using  $v_n = F^n v_0$  and the diagonalization, deduce *Binet's formula*:
$$F_n = \frac{1}{\sqrt{5}} \left( \gamma^n - \left( \frac{-1}{\gamma} \right)^n \right) = \frac{1}{\sqrt{5}} \left( \left( \frac{1+\sqrt{5}}{2} \right)^n - \left( \frac{1-\sqrt{5}}{2} \right)^n \right), \quad n \geq 1.$$
  
  - (c) Deduce that  $\lim_{n \rightarrow \infty} \frac{F_{n+1}}{F_n} = \gamma$ .  
  4. (a) Do a similar analysis as in Exercise 3 above, with the tribonacci sequence  $(T_n)$ . Use the matrix in Exercises 1.1 # 12(c), and show that it is primitive. Show that there are three eigenvalues:  $\rho > 1$  (real), and  $\lambda, \bar{\lambda}$ , complex conjugates, and these

satisfy the equation  $z^3 - z^2 - z - 1 = 0$  (compare with the equation satisfied by the eigenvalues of  $F$ ). Deduce that

$$T_n = a_1\rho^n + a_2\lambda^n + a_3\bar{\lambda}^n,$$

for constants  $a_1, a_2$  and  $a_3$ .

- (b) Show that  $a_2 = \bar{a}_3$ .
- (c) Deduce that  $\rho$  is a Pisot number.

5. (a) Show that the expanding direction for the tribonacci substitution (i.e., the direction given by the eigenvalue  $\rho$  of  $M_t$ ) is parallel to the vector  $(1, 1/\rho, 1/\rho^2)$ .

(b) Show that  $(T_n/T_{n+1}, T_{n-1}/T_{n+1})$  is a good approximation to  $(1/\rho, 1/\rho^2)$ , for  $n$  large.

6. Generalize the Fibonacci and tribonacci substitutions in a natural way, to find a substitution on  $\mathcal{A} = \{1, 2, 3, 4\}$  with analogous properties. Find the characteristic equation of the resulting incidence matrix. Conjecture what the natural generalization on  $\mathcal{A} = \{1, 2, \dots, d\}$  should look like.

7. Determine whether the following are unit Pisot irreducible substitutions:

- (i)  $\theta(1) = 121$ ,  $\theta(2) = 1$ , (ii)  $\theta(1) = 1121$ ,  $\theta(2) = 11$ , (iii)  $\theta(1) = 12$ ,  $\theta(2) = 23$ ,  $\theta(3) = 312$ .

8. (a) Show that if  $w$  is an initial word of the fixed point of the Fibonacci substitution ( $u = \theta^\infty(1)$ ), then we can write  $w = \theta^k(i_k)\theta^{k-1}(i_{k-1})\dots\theta(i_1)i_0$ , for some finite sequence  $i_0, i_1, \dots, i_k$ , where each  $i_j$  is either 1 or  $\epsilon$  (the empty word).

(b) If  $u = u_1u_2\dots u_n\dots$ , show that  $u_n = \lfloor (n+1)/\gamma \rfloor - \lfloor n/\gamma \rfloor$ , where  $\lfloor x \rfloor$  = the largest integer  $n \leq x$ .

(c) Show that the ratios of the frequencies of 1 and 2, in the sequence  $u = \theta^\infty(1)$  tend to the golden number  $\gamma = (1 + \sqrt{5})/2$ .

9. Prove the statement in 16.4.13 # 3, that if

$$\mathcal{R} = \left\{ \sum_{i \geq 0} \epsilon_i \alpha^i : \epsilon_i \in \{0, 1\}, \epsilon_i \epsilon_{i+1} \epsilon_{i+2} = 0 \right\},$$

where  $\epsilon_0 = 0$  gives  $\mathcal{R}_1$ ,  $\epsilon_0 \epsilon_1 = 10$  gives  $\mathcal{R}_2$ , and  $\epsilon_0 \epsilon_1 = 11$  gives  $\mathcal{R}_3$ , then  $\mathcal{R}_1 = \alpha \mathcal{R}$ ,  $\mathcal{R}_2 = \alpha^2 \mathcal{R} + 1$ ,  $\mathcal{R}_3 = \alpha^3 \mathcal{R} + \alpha + 1$ .

10. Find the vectors  $v_\rho$ ,  $v_\alpha$  and  $v_{\bar{\alpha}}$ , and prove all the statements in 16.4.13 # 5.

## CHAPTER 17

# Compactness in Metric Spaces and an Introduction to Topological Dynamics.

In this chapter and also in Chapter 18, we continue our study of substitutions, but in a more rigorous manner. In particular, we develop their topological properties and look at the dynamical systems they generate. In order to do this, we need to digress with a brief study of compactness in metric spaces and continuous functions on compact metric spaces. This leads to an introduction to *topological dynamics* on compact metric spaces, which is the general study of the dynamics of continuous functions defined on compact metric spaces. Given a continuous map  $f : X \rightarrow X$  on a compact metric space  $X$  (such as an interval  $X = [0, 1]$  in  $\mathbb{R}$ ), we can think of  $(X, f)$  as defining a (*topological*) dynamical system which is an abstraction of many of the examples we have met so far. The aim is to see what we can say in this general setting about the dynamics of  $f$ . A familiarity with Chapters 1, 2, 4, 5, 6 and 7 is assumed in this chapter, together with the contraction mapping theorem in Chapter 10, and substitutions in Chapter 15.

### 17.1 Compactness in Metric Spaces.

In earlier chapters, we have avoided the use of compactness. However, it is required when discussing properties of sequence spaces that arise in the mathematical theory of substitutions. In the case of the real line, the Heine-Borel Theorem is important, and was used in Section 5.1:

**17.1.1 The Heine-Borel Theorem.** *Every cover of a closed interval  $[a, b]$  by a collection of open sets, has a finite subcover.*

Recall that a cover of an interval  $I = [a, b]$  is simply a collection of sets whose union contains  $I$ . In the case of a metric space  $(X, d)$ , we generalize this as follows:

**Definition 17.1.2** Let  $(X, d)$  be a metric space. A *cover* of  $X$  is a collection of sets whose union is  $X$ . An *open cover* of  $X$  is a collection of open sets whose union is  $X$ .

**Definition 17.1.3** The metric space  $(X, d)$  is said to be *compact* if every open cover of  $X$  has a finite subcover, i.e., whenever  $X = \cup_{\lambda \in J} \mathcal{O}_\lambda$ , for some index set  $J$ , and open sets  $\mathcal{O}_\lambda$ , there is a finite subset  $K \subset J$  with  $X = \cup_{\lambda \in K} \mathcal{O}_\lambda$ .

The Heine-Borel theorem says that any closed interval  $[a, b]$  in  $\mathbb{R}$  is compact. It will be shown that a subset of  $\mathbb{R}^n$  is compact if and only if it is both closed and bounded.

Recall that in a metric space  $(X, d)$ ,  $\alpha \in X$  is a *limit point* of  $A \subseteq X$ , if every open ball  $B_\delta(\alpha)$  ( $\delta > 0$ ), centered on  $\alpha$ , contains points of  $A$  other than  $\alpha$ .

**Definition 17.1.4** The metric space  $(X, d)$  is *sequentially compact* if every infinite sequence  $(x_n)$  in  $X$  has a limit point in  $X$ . We shall see that this is equivalent to saying that every sequence  $(x_n)$  has a convergent subsequence.

For example, the space  $X = \mathbb{Q} \cap [0, 1]$  with usual metric is not sequentially compact since there are sequences of rationals in  $X$  that converge to an irrational. If  $X = [0, \infty)$ , then there are infinite sequences, such as  $x_n = n$ , which do not have limit points in  $X$ , and so  $X$  is not sequentially compact.

For a finite subset  $A$  of the metric space  $(X, d)$ ,  $d(x, A)$  is the distance between  $x$  and the nearest point of  $A$ . For more general sets  $A$ , we set

$$d(x, A) = \inf_{y \in A} d(x, y).$$

**Definition 17.1.5** The metric space  $(X, d)$  is *totally bounded*, if for every  $\epsilon > 0$ ,  $X$  can be covered by a finite number of open balls of radius  $\epsilon$ . If  $A$  is a finite subset of  $X$  with the property that  $d(x, A) < \epsilon$  for all  $x \in X$ , then we call  $A$  an  $\epsilon$ -net of  $X$ . The existence of an  $\epsilon$ -net for each  $\epsilon > 0$  is clearly equivalent to  $X$  being totally bounded.

The space  $X = \mathbb{Q} \cap [0, 1]$  is easily seen to be totally bounded, but it is not complete (there are Cauchy sequences in  $X$  that do not converge to a point of  $X$ - see Section 10.4).

These ideas lead to the following important theorem characterizing the compactness of metric spaces.

**Compactness of Metric Spaces Theorem 17.1.6** *The following are equivalent for a metric space  $(X, d)$ :*

- (a)  $X$  is compact.

- (b) If  $F_1 \supseteq F_2 \supseteq F_3 \supseteq \dots$  is a nested sequence of non-empty closed sets in  $X$ , then  $\cap_{n=1}^{\infty} F_n$  is non-empty.
- (c)  $X$  is sequentially compact.
- (d)  $X$  is totally bounded and complete.

**Proof.** (a)  $\Rightarrow$  (b) Let  $(F_n)_{n=1}^{\infty}$  be a decreasing sequence of non-empty closed sets in  $X$  and suppose that  $\cap_{n=1}^{\infty} F_n = \emptyset$ . Then  $(X \setminus F_n)$  is an increasing sequence of open sets with

$$\bigcup_{n=1}^{\infty} (X \setminus F_n) = X \setminus \bigcap_{n=1}^{\infty} F_n = X,$$

so that the collection of sets  $\{X \setminus F_n : n \in \mathbb{N}\}$  is an open cover of  $X$ . Since  $X$  is compact, there is a finite subcover: there is a finite index set  $I \subset \mathbb{N}$  with

$$X \setminus \bigcap_{n \in I} F_n = \bigcup_{n \in I} (X \setminus F_n) = X.$$

A contradiction since  $\cap_{n \in I} F_n = F_m \neq \emptyset$ , where  $m = \max(I)$ .

(b)  $\Rightarrow$  (c) Let  $(x_n)$  be an infinite sequence in  $X$ , and let  $F_n$  be the closure of the set  $\{x_n, x_{n+1}, x_{n+2}, \dots\}$ .  $(F_n)$  is a decreasing sequence of non-empty closed sets, so by (b),  $\cap_{n=1}^{\infty} F_n$  is non-empty. Let  $\alpha \in \cap_{n=1}^{\infty} F_n$ . Then we show that there is a subsequence  $(x_{n_k})$  which converges to  $\alpha$ .

Since  $\alpha \in F_1$ , there exists  $x_{n_1}$  with  $d(\alpha, x_{n_1}) < 1$ . Let  $n > n_1$ . Then  $\alpha \in F_n$ , so there exists  $n_2 > n_1$  with  $d(\alpha, x_{n_2}) < 1/2$ . Continue in this way to find  $x_{n_k}$  with  $n_k > n_{k-1}$  and  $d(\alpha, x_{n_k}) < 1/k$ . Clearly  $\alpha = \lim_{k \rightarrow \infty} x_{n_k}$ , so that (c) holds.

(c)  $\Rightarrow$  (d) First we show that  $X$  is complete. Let  $(x_n)$  be a Cauchy sequence in  $X$ . Since  $X$  is sequentially compact,  $(x_n)$  has a limit point  $\alpha \in X$ . Using the previous argument, there is a subsequence  $(x_{n_k})$  that converges to  $\alpha$ . Let  $\epsilon > 0$ . Then there exists  $N \in \mathbb{N}$  such that if  $m, n \geq N$ , we have  $d(x_n, x_m) < \epsilon/2$ , and such that if  $k \geq N$  then  $d(\alpha, x_{n_k}) < \epsilon/2$ . Clearly  $n_k \geq k \geq N$ , so that

$$d(\alpha, x_n) \leq d(\alpha, x_{n_k}) + d(x_{n_k}, x_n) < \epsilon/2 + \epsilon/2 = \epsilon.$$

Hence the sequence  $(x_n)$  is convergent. We have shown that every Cauchy sequence is convergent, so  $X$  is complete.

If  $X$  is not totally bounded, then there exists  $\epsilon > 0$  for which  $X$  has no finite covering by balls of radius  $\epsilon$ . The idea is to construct an infinite sequence of points

$(x_n)$  in  $X$  having distance apart at least  $\epsilon$ :  $d(x_i, x_j) \geq \epsilon$  for all  $i \neq j$ , contradicting the sequential compactness of  $X$ .

We proceed inductively taking  $x_1 \in X$  arbitrarily. Suppose that  $x_1, x_2, \dots, x_{n-1}$  have been chosen, then the open balls  $B_\epsilon(x_i)$ ,  $(1 \leq i \leq n-1)$ , have a union which cannot be the whole space. Consequently, if  $x_n \in X \setminus \cup_{i=1}^{n-1} B_\epsilon(x_i)$ , then  $x_n$  satisfies  $d(x_n, x_i) \geq \epsilon$  for all  $i = 1, \dots, n-1$ . It follows by the Principle of Induction that the required infinite sequence exists. This sequence cannot have a limit point, for if  $\alpha \in X$  were a limit point, say within  $\epsilon/2$  of some  $x_{n_0}$ , then the distance of  $\alpha$  from every other member of the sequence will be at least  $\epsilon/2$  since

$$\epsilon \leq d(x_{n_0}, x_m) \leq d(x_{n_0}, \alpha) + d(\alpha, x_m) < \epsilon/2 + d(\alpha, x_m).$$

This is clearly impossible for a limit point.

(d)  $\Rightarrow$  (a) If  $X$  is not compact, there is an open cover  $\mathcal{O} = \{O_\lambda : \lambda \in I\}$  of  $X$  that does not have a finite subcover. To obtain a contradiction, we first construct a convergent sequence  $(x_n)$  in  $X$ .

Using the fact that  $X$  is totally bounded, choose an  $\epsilon$ -net  $A_1$  with  $\epsilon = 1/2$ . Let  $x_1 \in A_1$  with the property that no finite sub-collection of  $\{O_\lambda : \lambda \in I\}$  covers  $B_{1/2}(x_1)$ .  $x_1$  exists because if every ball  $B_{1/2}(a)$  of radius  $1/2$  with  $a \in A_1$  had a finite subcover from  $\mathcal{O}$ , then the whole space would have a finite subcover from  $\mathcal{O}$ , because  $A_1$  is a finite set.

Now choose an  $\epsilon$ -net  $A_2$  with  $\epsilon = 1/2^2 = 1/4$ , and let  $x_2 \in A_2$  be chosen so that  $B_{1/2}(x_1) \cap B_{1/4}(x_2) \neq \emptyset$ , and with the property that no finite sub-collection of  $\{O_\lambda : \lambda \in I\}$  covers  $B_{1/4}(x_2)$ , ( $x_2$  exists because we know that  $B_{1/2}(x_1)$  has no finite subcover from  $\mathcal{O}$  so we can apply the previous argument to  $B_{1/2}(x_1)$ ).

We continue in this way so that at the  $n$ th stage we choose an  $\epsilon$ -net  $A_n$ , with  $\epsilon = 1/2^n$ , satisfying  $B_{1/2^{n-1}}(x_{n-1}) \cap B_{1/2^n}(x_n) \neq \emptyset$ , and with the property that no finite sub-cover of  $\mathcal{O}$  is a cover of  $B_{1/2^n}(x_n)$ .

The construction ensures that

$$d(x_{n-1}, x_n) \leq \frac{1}{2^{n-1}} + \frac{1}{2^n} \leq \frac{1}{2^{n-2}},$$

so that for  $m < n$

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) \\ &\leq \frac{1}{2^{m-1}} + \frac{1}{2^m} + \cdots + \frac{1}{2^{n-2}} \leq \frac{1}{2^{m-2}}, \end{aligned}$$

and  $(x_n)$  is a Cauchy sequence in  $X$ . The completeness of  $X$  implies that this sequence converges to some point  $\alpha \in X$ .

Suppose that  $\alpha \in O_{\lambda_0}$  for some  $\lambda_0 \in I$ . Since  $O_{\lambda_0}$  is an open set, we can choose  $\epsilon > 0$  with  $B_\epsilon(\alpha) \subseteq O_{\lambda_0}$ . Since  $\lim_{n \rightarrow \infty} x_n = \alpha$ , there exists  $n \in \mathbb{N}$  with  $d(x_n, \alpha) < \epsilon/2$  and  $1/2^n < \epsilon/2$ . If  $x \in B_{1/2^n}(x_n)$ , then

$$d(x, \alpha) \leq d(x, x_n) + d(x_n, \alpha) < 1/2^n + \epsilon/2 < \epsilon,$$

so that

$$B_{1/2^n}(x_n) \subseteq B_\epsilon(\alpha) \subseteq O_{\lambda_0}.$$

This contradicts the fact that no finite sub-collection of  $\mathcal{O}$  covers  $B_{1/2^n}(x_n)$ .  $\square$

**Examples 17.1.7** Every non-empty closed and bounded subset of  $\mathbb{R}$  or  $\mathbb{R}^n$  for  $n \geq 1$ , is a compact set (see the exercises). In particular, any non-empty closed interval  $[a, b]$  and the Cantor set  $C$ , are compact in  $\mathbb{R}$ . The unit circle  $\mathbb{S}^1$  and the closed unit disc in  $\mathbb{R}^2$  are compact. Julia sets arising from polynomials, and the Mandelbrot set are compact in  $\mathbb{C}$ .

2. In Chapter 18, we will show that the set of all one-sided sequences of 0's and 1's (with a suitable metric), is a compact metric space. This is a special case of Tychonoff's Theorem which states that the cartesian product of compact spaces is compact.
3. We say that a subset  $A$  of the metric space  $X$  is compact, if  $A$ , considered as a metric space in its own right (in this case,  $A$  is a *subspace* of  $X$ ), is compact. It is clear that any closed subset of a compact metric space is compact, and any compact subset of a metric space is closed.

**Remark 17.1.8** We have seen that any compact metric space  $(X, d)$  is necessarily complete. In addition, such a space is *separable*, i.e., has a countable dense subset. To show this, note that the total boundedness tells us that for each  $n \in \mathbb{N}$  there is a finite set  $A_n \subseteq X$ , such that if  $x \in X$ , then  $d(x, A_n) < 1/n$ . Set  $A = \bigcup_{n=1}^{\infty} A_n$ .  $A$  is a countable set with the property that for each  $x \in X$ ,  $d(x, A) \leq d(x, A_n) < 1/n$ , so that  $x \in \bar{A}$ , i.e.,  $\bar{A} = X$ , and  $X$  is separable (see the exercises).  $\square$

## Exercises 17.1

1. Let  $X = \{0\} \cup \{1, 1/2, 1/3, \dots\}$ . Show directly, that any open cover has a finite subcover, so that  $X$  is a compact metric space (with the induced metric as a subset of  $\mathbb{R}$ ). Why does this not contradict the Baire Category Theorem? (See Appendix B for its statement).
  
2. Show that the open interval  $(0, 1)$  is not compact by exhibiting an open cover with no finite subcover.
  
3. Prove that a compact subset of a metric space is closed, and that a closed subset of a compact metric space is compact.
  
4. Prove that a subset of  $\mathbb{R}^n$  is compact if and only if it is both closed and bounded.
  
5. Show that if  $A$  is a subset of the metric space  $X$  with  $d(x, A) < \epsilon$  for all  $\epsilon > 0$ , then  $x \in \overline{A}$ .
  
6. Let  $(X, d_X)$  and  $(Y, d_Y)$  be metric spaces. Distances  $d$  and  $d'$  are defined on the product space  $X \times Y = \{(x, y) : x \in X, y \in Y\}$  by  $d((x_1, y_1), (x_2, y_2)) = \sqrt{[d_X(x_1, x_2)]^2 + [d_Y(y_1, y_2)]^2}$ ,  $d'((x_1, y_1), (x_2, y_2)) = \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}$ .
  - (a) Show that  $(X \times Y, d)$  and  $(X \times Y, d')$  are metric spaces.
  - (b) If  $(X, d_X)$  and  $(Y, d_Y)$  are compact metric spaces, show that  $(X \times Y, d)$  and  $(X \times Y, d')$  are compact metric spaces. (Hint: Show that if  $z_n = (x_n, y_n)$  is a sequence in  $X \times Y$ , it has a convergent subsequence).

## 17.2 Continuous Functions on Compact Metric Spaces.

In Section 4.3 we studied continuous functions on metric spaces, and we saw that for such functions the inverse images of open sets are open, and the inverse images of closed sets are closed. Here we see that the images of compact sets are necessarily compact.

**Theorem 17.2.1** *Let  $f : X \rightarrow Y$  be a continuous map of metric spaces  $(X, d_X)$  and  $(Y, d_Y)$ . If  $X$  is compact, so is  $f(X) = \{f(x) : x \in X\}$ , the image of  $X$  in  $Y$ .*

**Proof.** Let  $\{O_\lambda : \lambda \in I\}$  be an open cover of  $f(X)$ . Then each of the sets  $f^{-1}(O_\lambda)$  is open in  $X$  and  $X = \bigcup_{\lambda \in I} f^{-1}(O_\lambda)$ . As  $X$  is compact, there is a finite subcover: a finite set  $J \subset I$  with  $X = \bigcup_{\lambda \in J} f^{-1}(O_\lambda)$ . We can now check that  $\{O_\lambda : \lambda \in J\}$  is a cover of  $f(X)$ . □

An immediate consequence is the following result, which will be needed shortly. It says that a continuous map on a compact metric space into  $\mathbb{R}$ , attains both its maximum and minimum values.

**Corollary 17.2.2** *Let  $(X, d)$  be a compact metric space and  $f : X \rightarrow \mathbb{R}$  a continuous function. Then there are points  $\alpha, \beta \in X$  such that  $f(\alpha) = \inf_{x \in X} f(x)$  and  $f(\beta) = \sup_{x \in X} f(x)$ .*

**Proof.** By Theorem 17.2.1,  $f(X)$  is a compact subset of  $\mathbb{R}$ , so it is both closed and bounded. In particular, if  $h = \sup_{x \in X} f(x)$  and  $k = \inf_{x \in X} f(x)$ , then  $h, k \in f(X)$ . To see this, suppose that  $h \notin f(X)$ . Then we can find an open ball  $B_\delta(h)$ , surrounding  $h$  which does not intersect  $f(X)$ . Thus  $h - \delta/2$  is an upper bound for  $f(X)$  smaller than  $h$ , giving a contradiction. We deduce that  $h = f(\beta)$ , and similarly  $k = f(\alpha)$  for some  $\alpha, \beta \in X$ . □

**Examples 17.2.3** One of the major theorems from real analysis follows from Corollary 17.2.2: *Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function. Then  $f([a, b])$  is a closed and bounded interval.* This follows from the fact that  $[a, b]$  is compact in  $\mathbb{R}$ , so  $f([a, b])$  is compact. In addition, the Intermediate Value Theorem and properties of intervals tell us that  $f([a, b])$  is an interval. (An *interval* is a subset  $I$  of  $\mathbb{R}$ , with the property that if  $a, b \in I$ ,  $a \leq b$ , and  $a \leq y \leq b$ , then  $y \in I$ . See [117] for more details).

## Exercises 17.2

1. Let  $f : X \rightarrow Y$  be a map of metric spaces  $(X, d)$  and  $(Y, \rho)$ . Prove:

- (a) If  $f$  is a constant map (there exists  $a \in Y$  with  $f(x) = a$  for all  $x \in X$ ), then  $f$  is continuous.
- (b) If  $X = Y$  and  $f$  is the identity map ( $f(x) = x$  for all  $x \in X$ ), then  $f$  is continuous.
- (c) If  $A$  is a subspace of  $X$  and  $f : X \rightarrow Y$  is continuous, then the restriction of  $f$  to  $A$  (written  $f|_A$ ), is continuous.
2. Let  $f : X \rightarrow Y$  be a continuous, one-to-one, and onto map between compact metric spaces  $(X, d)$  and  $(Y, \rho)$ . Prove that the inverse of  $f$  is continuous, so  $f$  is a homeomorphism.
3. Let  $f : X \rightarrow Y$  be a continuous map between the metric spaces  $(X, d)$  and  $(Y, \rho)$ , where  $X$  is compact. Prove that  $f$  is *uniformly continuous*, i.e., given  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $d(x, y) < \delta$  implies  $\rho(f(x), f(y)) < \epsilon$ , for all  $x, y \in X$ . (Hint: Given  $x \in X$ , use the continuity of  $f$  at  $x$  to find an open cover of  $X$ , which has a finite subcover).

### 17.3 The Contraction Mapping Theorem for Compact Metric Spaces.

In Section 10.5, we proved a fixed point theorem for contraction mappings on complete metric spaces. The requirement was that  $d(f(x), f(y)) \leq \alpha \cdot d(x, y)$  for all  $x, y \in X$ , where  $0 < \alpha < 1$ . Compact metric spaces are complete, and we can weaken the conditions of the contraction mapping theorem slightly in this case. Our aim is to apply this result in the next chapter to show that substitutions always have a fixed point.

**Theorem 17.3.1** *Let  $f : X \rightarrow X$  be a function defined on the compact metric space  $(X, d)$ , having the property*

$$d(f(x), f(y)) < d(x, y), \quad \text{for all } x, y \in X, \quad x \neq y.$$

*Then  $f$  has a unique fixed point in  $X$ .*

We first prove a lemma:

**Lemma 17.3.2** *If  $f : X \rightarrow X$  is continuous, then the function  $k : X \rightarrow [0, \infty)$ ,  $k(x) = d(x, f(x))$  is continuous.*

**Proof.** From the triangle inequality we obtain  $|d(a, c) - d(c, b)| \leq d(a, b)$  for all  $a, b, c \in X$ .

Let  $\epsilon > 0$  and fix  $x \in X$ . We must show that there is a  $\delta > 0$  such that if  $y \in X$  with  $d(x, y) < \delta$ , then

$$|d(x, f(x)) - d(y, f(y))| < \epsilon.$$

We may assume that there exists  $\delta_1 > 0$ , such that if  $y \in X$  with  $d(x, y) < \delta_1$ , then  $d(f(x), f(y)) < \epsilon/2$ , (using the continuity of  $f$  at  $x$ ).

The above inequality and the usual triangle inequality gives:

$$\begin{aligned} |d(x, f(x)) - d(y, f(y))| &= |d(x, f(x)) - d(f(x), y) + d(f(x), y) - d(y, f(y))| \\ &\leq |d(x, f(x)) - d(f(x), y)| + |d(f(x), y) - d(y, f(y))| \\ &\leq d(x, y) + d(f(x), f(y)) < \epsilon/2 + \epsilon/2 = \epsilon, \end{aligned}$$

if  $d(x, y) < \delta = \min\{\delta_1, \epsilon/2\}$ .

□

**Proof of the Theorem 17.3.1** Clearly  $f$  is continuous on  $X$ . By the lemma,  $k(x) = d(x, f(x))$  is a continuous function defined on a compact space. Consequently, its range is a closed and bounded subset of  $\mathbb{R}^+$ . It follows that  $\alpha = \inf_{x \in X} k(x)$  is attained, and there is an  $x_0 \in X$  with  $\alpha = k(x_0)$ .

Suppose that  $\alpha > 0$ . Since  $\alpha = d(x_0, f(x_0)) \neq 0$ , we have  $f(x_0) \neq x_0$ . In addition

$$d(f^2(x_0), f(x_0)) < d(f(x_0), x_0) = \alpha,$$

a contradiction since  $\alpha = \inf_{x \in X} d(x, f(x))$ . Therefore we must have  $\alpha = d(x_0, f(x_0)) = 0$ , so  $f(x_0) = x_0$ , a fixed point of  $f$ . As usual, if  $y_0$  is another fixed point then the inequality  $d(f(x_0), f(y_0)) < d(x_0, y_0)$  leads to a contradiction unless  $x_0 = y_0$ .

□

## 17.4 Basic Topological Dynamics.

In this section we develop the topological dynamics needed to study substitutions. For more detail, see the excellent books by P. Walters [127] and K. Petersen [102]. The main object of study here is the (topological) dynamical system  $(X, T)$ . Throughout this section we assume that  $(X, d)$  is a compact metric space, and  $T : X \rightarrow X$  is a continuous map (often a homeomorphism). As usual,

$O(x) = \{T^n(x) : n \in \mathbb{N}\}$  is the orbit of  $x$  under  $T$ . It is usual to assume that  $X$  is not a finite set.

**Examples 17.4.1** 1. Any continuous map  $f : I \rightarrow I$  on an interval  $I \subseteq \mathbb{R}$  gives rise to a dynamical system  $(I, f)$ . For example, maps from the logistic family or tent family, can be thought of as defining a topological dynamical systems.

2. The rotation  $R_a : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $R_a(z) = az$ , where  $\mathbb{S}^1$  is the unit circle in  $\mathbb{C}$ , and  $a \in \mathbb{S}^1$  is fixed, gives rise to a dynamical system. In this case,  $R_a$  is actually a homeomorphism on  $\mathbb{S}^1$ .  $R_a$  is closely related to the map  $T_\alpha : [0, 1) \rightarrow [0, 1)$  defined by  $T_\alpha(x) = x + \alpha \pmod{1}$ , when  $a = e^{2\pi i \alpha}$ . The squaring map  $S : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $S(z) = z^2$  is continuous, and so gives rise to another topological dynamical system.

3. In Section 6.5, the shift map  $\sigma$  on the sequence space  $\mathcal{A}^\mathbb{N}$ , was seen to be continuous, and it defines a topological dynamical system (in Chapter 18, we will show that when  $\mathcal{A} = \{0, 1\}$ ,  $\mathcal{A}^\mathbb{N}$  is a compact metric space).

**17.4.2 Minimality.** The dynamical system  $(X, T)$  is said to be *minimal* if for every  $x \in X$ ,  $\overline{O(x)} = X$ , i.e., the orbit closure of every point is dense in  $X$ . In this case we say that  $T : X \rightarrow X$  is a *minimal transformation*. This seems to be a very strong condition (recall that  $T$  is *transitive* if there exists  $x \in X$  with  $\overline{O(x)} = X$ ), but we will see that there is a wide range of examples that are minimal. If  $X$  is a finite set, then  $T$  can only be minimal if  $X$  consists of the orbit of a single point.

**Definition 17.4.3** Let  $(X, T)$  be a dynamical system. A non-empty, closed invariant subset  $A \subset X$  is said to be a *minimal set*, if the restriction of  $T$  to  $A$  is a minimal map. Thus a minimal set is one in which every point has a dense orbit.

It can be shown that any continuous map  $T : X \rightarrow X$  of a compact metric space, has at least one minimal set. For example if  $f : [0, 1] \rightarrow [0, 1]$  is defined by  $f(x) = x^2$ , then the only minimal sets are the end points  $x = 0$  and  $x = 1$  (see Exercises 17.4 # 10). For the shift map  $\sigma : \Sigma \rightarrow \Sigma$  where  $\Sigma = \{0, 1\}^\mathbb{N}$  (see Exercise 17.4 # 12), there are infinitely many minimal sets different from those arising from periodic points. The orbit closure  $\overline{O(\omega)}$  of a point  $\omega \in \Sigma$  often turns out to be a minimal set for  $(\Sigma, \sigma)$ .

If  $T$  is minimal, the only closed invariant subsets are the whole space or the empty set. In particular, if  $X$  is infinite, there cannot be any fixed points or periodic points. This rules out most of the examples we have seen thus far, as minimal maps (on

infinite spaces), cannot have period points, so cannot be chaotic. Later, we shall see that substitution dynamical systems give rise to minimal maps.

**Theorem 17.4.4** *Let  $X$  be a compact metric space. The following are equivalent for the dynamical system  $(X, T)$ .*

- (a)  $(X, T)$  is minimal.
- (b) If  $E \subseteq X$  is a closed set with  $T(E) \subset E$ , then  $E = X$  or  $E = \emptyset$ .
- (c) If  $U$  is a non-empty open subset of  $X$ , then  $\cup_{n=0}^{\infty} T^{-n}U = X$ .

**Proof.** (a)  $\Rightarrow$  (b) Suppose  $T$  is minimal and  $x \in E$ , where  $E$  is non-empty and closed. If  $T(E) \subset E$ , we must have  $O(x) \subset E$ , so  $X = \overline{O(x)} = E$ , since  $E$  is closed.

(b)  $\Rightarrow$  (c) If  $U$  is a non-empty open set, then  $E = X \setminus \cup_{n=0}^{\infty} T^{-n}U \neq X$  is closed (we are assuming that  $T$  is continuous). We can now check that standard set theoretic arguments involving functions give  $T(E) \subset E$ , and so we must have  $E = \emptyset$ .

(c)  $\Rightarrow$  (a) If  $x \in X$  and  $U$  is a non-empty open subset of  $X$ , then from  $\cup_{n=0}^{\infty} T^{-n}U = X$ ,  $T^n x \in U$  for some  $n > 0$ . It follows that  $O(x)$  is dense in  $X$ , since the orbit of  $x$  intersects any open set.

□

The next result tells us that for transitive maps, the only continuous invariant functions are constant functions.

**Proposition 17.4.5** *Let  $X$  be a compact metric space. If  $T : X \rightarrow X$  is transitive and  $f : X \rightarrow \mathbb{C}$  is a continuous function with  $f(Tx) = f(x)$  for all  $x \in X$ , then  $f$  is a constant function.*

**Proof.** We have  $\overline{O(x_0)} = X$  for some  $x_0 \in X$ . Now  $f(T^n x_0) = f(x_0)$  for all  $n \in \mathbb{N}$ . This implies that  $f$  is a constant  $c$  on a dense subset of  $X$ . Let  $\alpha \in X$ . Then  $\alpha$  is a limit point of  $O(x_0)$ , so there is a sequence  $(x_{n_k})_{k \geq 1}$  in  $O(x_0)$ , with  $\lim_{k \rightarrow \infty} x_{n_k} = \alpha$ . From the continuity of  $f$ ,  $f(\alpha) = \lim_{k \rightarrow \infty} f(x_{n_k}) = c$ , so  $f$  is constant on  $X$ .

□

**Examples 17.4.6** 1. Consider  $R_a : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $R_a(z) = az$ , where  $a \in \mathbb{S}^1$ , and  $a$  is not a root of unity (i.e.,  $a^n \neq 1$  for all  $n \in \mathbb{Z}^+$ ).  $R_a$  is said to be an *irrational rotation*. Irrational rotations are always minimal.

**Proof.** The set  $\{a^n : n \in \mathbb{N}\}$  is an infinite sequence in  $\mathbb{S}^1$ , so by the compactness of  $\mathbb{S}^1$ , it has a limit point, say  $\alpha \in \mathbb{S}^1$ . Consequently, given  $\epsilon > 0$  there are  $p > q \in \mathbb{N}$  with  $d(a^p, a^q) < \epsilon$  ( $d$  is the usual metric on  $\mathbb{S}^1$ , giving the shortest distance between two points around the circle). Since a rotation is an isometry on  $\mathbb{S}^1$ , it follows that  $d(a^{p-q}, 1) = d(a^p, a^q) < \epsilon$ . Set  $b = a^{p-q}$ , then  $d(b^n, b^{n-1}) < \epsilon$  for  $n > 1$ . The points  $1, b, b^2, b^3, \dots$  are equally spaced around the circle, and so form an  $\epsilon$ -net for  $\mathbb{S}^1$ . Since  $\epsilon$  is arbitrary, it follows that the set  $\{a^n : n \in \mathbb{Z}^+\}$  is dense in  $\mathbb{S}^1$ .

In particular, the sequence  $R_a^n(1) = a^n$  is dense in  $\mathbb{S}^1$ . So for any  $z, w \in \mathbb{S}^1$  we can find a sequence of integers  $(k_n)$  with  $a^{k_n} \rightarrow wz^{-1}$ . Consequently  $R_a^{k_n}(z) = a^{k_n}z \rightarrow wz^{-1}z = w$  as  $n \rightarrow \infty$ , so  $O(z)$  is dense in  $\mathbb{S}^1$ , and  $R_a$  is minimal. □

2. Clearly  $T : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $T(z) = z^2$  is not minimal since  $T(1) = 1$  (in fact  $T$  has uncountably many minimal sets). However,  $T$  is transitive. We indicated why this is true in Section 6.2 using:  $T$  is transitive if and only if, given any open intervals  $U, V \subset \mathbb{S}^1$ , we have  $U \cap T^nV \neq \emptyset$  for some  $n > 0$ . The transitivity of  $T$  is a consequence of Theorem 17.4.7 of this section.

3. Recall that in Chapter 7, the notions of conjugacy and factor map were defined. The dynamical system  $(Y, S)$  is said to be a *factor* of the dynamical system  $(X, T)$ , if there is a continuous map  $\phi$  from  $X$  onto  $Y$  with  $\phi \circ T = S \circ \phi$ .  $(X, T)$  is called an *extension* of  $(Y, S)$  and  $\phi$  is called a *factor map*. The two dynamical systems are *conjugate* (with *conjugacy*  $\phi$ ), if  $\phi : X \rightarrow Y$  is a homeomorphism. In Section 7.2, we saw that if the two dynamical systems are conjugate, then  $T$  is transitive if and only if  $S$  is transitive. In a similar way,  $T$  is minimal if and only if  $S$  is minimal.

For example, if  $R_a : \mathbb{S}^1 \rightarrow \mathbb{S}^1$  is  $R_a(z) = az$ , the rotation of Example 1, and if  $\phi(z) = z^2$ , then  $\phi \circ R_a = R_a^2 \circ \phi$ , so  $R_a^2$  is a factor of  $R_a$ . This is not a conjugacy since  $\phi$  is continuous and onto, but not one-to-one. On the other hand,  $R_a^{-1}(z) = a^{-1}z$ , and if  $\psi(z) = z^{-1}$ , then  $R_a \circ \psi = \psi \circ R_a^{-1}$ , so that  $R_a$  and  $R_a^{-1}$  are conjugate (since  $\psi$  is continuous, one-to-one and onto). It can be shown that every conjugacy between an irrational rotation  $R_a$  and its inverse  $R_a^{-1}$ , is of the form  $\psi(z) = cz^{-1}$  for some  $c \in \mathbb{S}^1$ . Thus every such conjugacy is an *involution* ( $\psi^2 = I$ , the *identity map*).

4. Consider the map  $T_\alpha : [0, 1] \rightarrow [0, 1]$  defined by  $T_\alpha(x) = x + \alpha \pmod{1}$ , where  $\alpha$  is irrational. If we define  $\phi : [0, 1] \rightarrow \mathbb{S}^1$  by  $\phi(x) = e^{2\pi i x}$ , we see that if  $R_a(z) = az$ , where  $a = e^{2\pi i \alpha}$ , then  $\phi(T_\alpha(x)) = R_a(\phi(x))$ , for  $x \in [0, 1]$ . However,  $\phi$  is not quite a conjugacy.  $T_\alpha$  has a point of discontinuity, whereas  $R_a$  is continuous. We regard both

of these maps as irrational rotations. The spaces  $[0, 1)$  and  $\mathbb{S}^1$  are quite different, and  $\phi$  is not a homeomorphism ( $\phi^{-1}$  is not continuous at  $z = 1$ ). If we identify 0 and 1 in  $[0, 1)$ ,  $\phi$  can be considered as a homeomorphism. However, we will not pursue this line of thought, and we rather give a direct proof that  $T_\alpha$  is minimal when  $\alpha$  is irrational. Our proof uses the Pigeonhole Principle:

Denote by  $\{x\}$  the fractional part of  $x \in \mathbb{R}$ , i.e.,  $x$  reduced modulo one. We show that for any  $0 < a < b < 1$ , there exists  $n \in \mathbb{N}$  with  $\{n\alpha\} \in (a, b)$ . Choose  $N \in \mathbb{N}$  so large that  $1/N < b - a$ . Divide  $[0, 1)$  into  $N$  segments of length  $1/N$ . Let  $m > N$  and consider the set  $\{\{\alpha\}, \{2\alpha\}, \dots, \{m\alpha\}\}$ , all distinct since  $\alpha$  is irrational. By the Pigeonhole Principle, there are  $j$  and  $k$  with  $\{j\alpha\}$  and  $\{k\alpha\}$  belonging to the same segment of length  $1/N$ . Set  $r = j - k$ . Then we have  $0 < \{r\alpha\} < 1/N$ . Set  $\beta = \{r\alpha\}$ . The set  $\{\beta, 2\beta, \dots, n\beta, \dots\}$  subdivides  $[0, 1)$  into intervals of length less than  $1/N$ , and so one of them must fall into the interval  $(a, b)$ . This is the required  $\{n\alpha\}$ .

□

5. It has been shown by Auslander and Yorke [7], that if  $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$  is a continuous minimal map, then  $f$  is conjugate to an irrational rotation.

To conclude this section, we give a result similar to Theorem 17.4.4, but requiring only that  $T$  be transitive instead of minimal. The proof requires the use of the Baire Category Theorem (see Appendix B for its statement). Recall that a *nowhere dense set* is one whose closure does not contain any non-empty open sets. The equivalence of (a) and (c) below was mentioned in Section 5.2. This result can be given in greater generality (see [79]), but the following is sufficient for our needs. We require  $T$  to be onto (a transitive map need not be onto), as the equivalences in Theorem 17.4.7 fail without this additional condition (see the exercises).

**Theorem 17.4.7** *Let  $(X, d)$  be a compact metric space with  $T : X \rightarrow X$  a continuous, onto map. The following are equivalent:*

- (a)  *$T$  is transitive.*
- (b) *If  $E$  is a closed set with  $T(E) \subset E$ ,  $E \neq X$ , then  $E$  is nowhere dense.*
- (c) *If  $U$  and  $V$  are non-empty open sets in  $X$ , there exists  $n \geq 1$  with  $T^n U \cap V \neq \emptyset$ .*
- (d) *The set  $\{x \in X : \overline{O(x)} = X\}$  is dense in  $X$ .*

**Proof.** (a)  $\Rightarrow$  (b) Suppose that  $O(x_0)$  is dense in  $X$  and  $U$  is a non-empty open set with  $U \subset E$ . There exists  $m \in \mathbb{N}$  with  $T^m(x_0) \in U$ . It follows that  $\{T^n(x_0) : n \geq m\} \subset E$ , so

$$\{x_0, Tx_0, \dots, T^{m-1}x_0\} \cup E = X,$$

(this is because the left hand side is a closed set that contains all of the orbit of  $x_0$ ).

Since  $T$  is onto, applying  $T$  to both sides gives  $\{Tx_0, \dots, T^m x_0\} \cup E = TX = X$ . Doing this repeatedly, we see that  $E = X$ .

(b)  $\Rightarrow$  (c) Let  $U$  and  $V$  be non-empty and open, then  $\cup_{n=0}^{\infty} T^{-n}V$  is open and

$$E = X \setminus \bigcup_{n=0}^{\infty} T^{-n}V = \bigcap_{n=0}^{\infty} (X \setminus T^{-n}V)$$

is closed and  $T(E) \subset E$ . But  $E \neq X$  as  $V$  is non-empty, and so contains no non-empty open sets. It follows that  $\cup_{n=0}^{\infty} T^{-n}V$  is dense in  $X$ , and thus intersects  $U$ . In particular,  $T^m U$  must intersect  $V$  for some  $m \in \mathbb{N}$ .

(c)  $\Rightarrow$  (d) Since  $X$  is separable, there is a countable dense set  $A = \{x_n : n \in \mathbb{N}\} \subset X$ . Enumerate the positive rationals  $\mathbb{Q}^+ = \{r_n : n \in \mathbb{N}\}$ . Then the set

$$\mathcal{U} = \{B_{r_n}(x_m) : m, n \in \mathbb{N}\} = \{U_n : n \in \mathbb{N}\} \text{ say,}$$

is a countable collection of open balls, and any open set contains members of  $\mathcal{U}$ . We can check that

$$\{x \in X : \overline{O(x)} = X\} = \bigcap_{n=0}^{\infty} \left( \bigcup_{m=0}^{\infty} T^{-m}U_n \right).$$

For each  $n \in \mathbb{N}$ , it follows as in the last proof, that  $\cup_{m=0}^{\infty} T^{-m}U_n$  is an open dense subset of  $X$ , and so  $\{x \in X : \overline{O(x)} = X\}$  is the countable intersection of open dense sets. The Baire Category Theorem (see Appendix B), now implies that this intersection is dense in  $X$ . (d) then follows.

(d)  $\Rightarrow$  (a) This is now clear. □

We modify the proof of Theorem 17.4.7 to give the proof of the Birkhoff Transitivity Theorem, mentioned in Chapter 6.

**Corollary 17.4.8** (The Birkhoff Transitivity Theorem). *Let  $(X, d)$  be a compact metric space with no isolated points, and let  $T : X \rightarrow X$  be a continuous map. Then  $T$  is transitive if and only if for every pair of non-empty open subsets  $U$  and  $V$  of  $X$ , there exists  $n > 0$  with  $T^n(U) \cap V \neq \emptyset$ .*

**Proof.** Suppose that  $\overline{O(x_0)} = X$  and  $U$  and  $V$  are non-empty open sets in  $X$ . There exists  $n \geq 1$  with  $T^n(x_0) \in U$ . The set  $V \setminus \{x_0, T(x_0), \dots, T^n(x_0)\}$  is open, and non-empty because  $X$  has no isolated points. It follows that there exists  $m > 0$  with  $T^m(x_0) \in V \setminus \{x_0, T(x_0), \dots, T^n(x_0)\}$ . In particular,  $m > n$  and  $T^m(x_0) \in T^{m-n}(U) \cap V$ , so  $T^{m-n}(U) \cap V \neq \emptyset$ .

The other direction follows directly from the proof of Theorem 17.4.7 using (c)  $\Rightarrow$  (d)  $\Rightarrow$  (a). □

### Exercises 17.4

1. (a) Let  $T : X \rightarrow X$  be an isometry ( $d(Tx, Ty) = d(x, y)$  for all  $x, y \in X$ ), which is transitive. Prove that  $T$  is minimal.  
 (b) Show that two minimal sets  $A$  and  $B$ , for the topological dynamical system  $(X, T)$ , are either disjoint or equal.  
 (c) If  $x_0 \in X$  is a point of period  $n$  for the dynamical system  $(X, T)$ , show that  $O(x_0)$  is a minimal set for  $T$ . Deduce that if  $X \neq O(x_0)$ ,  $T$  cannot be minimal.
  
2. (a) Show that if  $f : I \rightarrow I$  is a continuous, transitive map on an interval  $I \subseteq \mathbb{R}$ , then  $f$  is onto.  
 (b) Show that if  $(X, T)$  is a minimal dynamical system, then  $T$  is onto.  
 (c) Show that the following example is continuous, transitive and not onto. Let  $X = \{0, 1, 1/2, 1/3, \dots, 1/n, \dots\}$  and define  $T : X \rightarrow X$  by  $T(0) = 0$  and  $T(1/n) = 1/(n+1)$ .  $X$  is a metric space as a subspace of  $\mathbb{R}$  with its usual metric. Deduce that not all of the equivalences in Theorem 17.4.7 need hold without the onto assumption.
  
3. Let  $T : X \rightarrow X$  be a continuous map on the compact metric space  $X$ .  $\lambda \in \mathbb{C}$  is an *eigenvalue* of  $T$  if there exists  $f : X \rightarrow \mathbb{C}$ ,  $f \neq 0$ , continuous with  $f(Tx) = \lambda f(x)$  for all  $x \in X$ .  $f$  is an *eigenfunction* of  $T$  corresponding to the eigenvalue  $\lambda$ . Suppose that  $T$  is transitive and onto.

- (a) Show that if  $\lambda$  is an eigenvalue of  $T$  with corresponding eigenfunction  $f$ , then  $|\lambda| = 1$  and  $|f(x)| = \text{constant}$ , for all  $x \in X$ . (Hint: Note that  $\sup_{x \in X} |f(Tx)| = \sup_{x \in X} |f(x)|$ ).
- (b) If  $f$  and  $g$  are eigenfunctions corresponding to the same eigenvalue  $\lambda$ , then show that  $f(x) = c \cdot g(x)$  for all  $x \in X$ , for some  $c \in \mathbb{C}$ .
- (c) Prove that conjugate maps have the same eigenvalues.
- (d) Find the eigenvalues of the rotation  $R_a : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $R_a(z) = az$ . Deduce that the rotations  $R_a$  and  $R_b$  cannot be conjugate if  $\{a^n : n \in \mathbb{Z}\} \neq \{b^n : n \in \mathbb{Z}\}$ . (Hint: Show that the maps  $f_n(z) = z^n$  are eigenfunctions).
- (e)\* Show that the eigenvalues of  $T$  form a subgroup of  $\mathbb{S}^1$ . (This subgroup can be shown to be countable).

4. Let  $T : X \rightarrow X$  be a continuous map of the compact metric space  $(X, d)$ .  $x \in X$  is a *wandering point* of  $T$  if there is a non-empty open set  $U$  containing  $x$  such that  $U \cap T^n(U) = \emptyset$  for all  $n \in \mathbb{Z}^+$ . A point  $x \in X$  is called *non-wandering* if it is not wandering. The set of non-wandering points of  $T$  is denoted by  $\Omega(T)$ .

- (a) If  $L_\mu(x) = \mu x(1-x)$ ,  $0 < \mu < 4$ , show that  $x = 1$  is a wandering point, but  $x = 0$  is not. What happens when  $\mu = 4$ ?
- (b) Find the wandering points of  $f : [0, 1] \rightarrow [0, 1]$ ,  $f(x) = x^2$ .

5. Let  $T : X \rightarrow X$  be a continuous map of the compact metric space  $(X, d)$ . Show that

- (a) The set of all wandering points is open. Deduce that  $\Omega(T)$  is a closed set which contains all of the periodic points of  $T$ .
- (b)  $\Omega(T)$  is a set invariant under  $T$ .
- (c)  $\Omega(T^n) \subset \Omega(T)$  for all  $n \in \mathbb{Z}^+$ .
- (d) If  $T$  is a homeomorphism, then  $T(\Omega(T)) = \Omega(T)$  and  $\Omega(T^{-1}) = \Omega(T)$ .

6. Let  $T : X \rightarrow X$  be a continuous map of the compact metric space  $(X, d)$ . If  $x \in X$ , then the  $\omega$ -limit set  $\omega(x)$  is defined to be the set:

$$\omega(x) = \bigcap_{n=0}^{\infty} \overline{\{T^k(x) : k > n\}}.$$

Prove the following:

- (a)  $y \in \omega(x)$  if and only if there is an increasing sequence  $(n_k)$  such that  $T^{n_k}(x) \rightarrow y$  as  $k \rightarrow \infty$ .
- (b)  $\omega(x)$  is a non-empty, closed and  $T$ -invariant set.
- (c) If  $x$  is a periodic point of  $T$ , then  $\omega(x) = O(x)$ . Also, if  $x$  is eventually periodic with  $y \in O(x)$ , then  $\omega(x) = O(y)$ .
- (d) If  $\omega(x)$  consists of a single point, then that point is a fixed point.

7. Let  $T : X \rightarrow X$  be a continuous map of the compact metric space  $(X, d)$ .  $x \in X$ , is said to be *recurrent* if  $x \in \omega(x)$ , i.e.,  $x$  belongs to its  $\omega$ -limit set. Recurrent points are a generalization of periodic points. Periodic points return to themselves, whereas recurrent points return closely to themselves infinitely often, becoming closer as the iteration procedure progresses. Show

- (a)  $x \in X$  is recurrent, if and only if for every  $\epsilon > 0$ , the set

$$\{n \in \mathbb{Z}^+ : d(T^n x, x) < \epsilon\} \text{ is infinite.}$$

- (b) A periodic point is recurrent but an eventually periodic point is not recurrent.
- (c) Any recurrent point  $x$ , belongs to  $\Omega(T)$ .
- (d) If  $f : [0, 1] \rightarrow [0, 1]$  is a homeomorphism, then the only recurrent points are the periodic points.
- (e)  $\omega(x) \subset \overline{O(x)}$ , and  $\omega(x) = \overline{O(x)}$  if and only if  $x$  is recurrent.

8. Let  $(X, T)$  be a dynamical system with  $X$  compact. If  $\{U_\alpha : \alpha \in \Lambda\}$  is an open cover of  $X$ , show that  $U_\alpha \cap T^{-n}(U_\alpha) \neq \emptyset$ , for some  $\alpha \in \Lambda$ , and for infinitely many

$n \in \mathbb{Z}^+$ . (Hint: Use compactness, and the fact that if  $\mathbb{Z}^+$  is partitioned into finitely many sets, one of the sets must contain infinitely many integers).

9. (a) Let  $T_2 : [0, 1] \rightarrow [0, 1]$  be the tent map. Show that there are points  $x \in [0, 1]$  with  $\omega(x) = [0, 1]$ . (Hint: Use a variation of the transitive point for  $T_2$ ).

(b) Deduce that there are points  $x \in [0, 1]$  with  $\omega(x) = [0, 1]$  for the Logistic map  $L_4(x) = 4x(1 - x)$ . (Hint: Use the conjugacy between  $L_4$  and  $T_2$ ).

10\*. The purpose of this exercise is to show that if  $T : X \rightarrow X$  is continuous on a compact metric space  $X$ , then  $T$  has a minimal set  $E$  (so  $E$  is a  $T$ -invariant set for which the restriction of  $T$  to  $E$  is minimal). The proof follows [57], avoiding the usual approach using Zorn's Lemma (see [127]).

Let  $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$  be the collection of open sets defined in the proof of Theorem 17.4.7 ( $\mathcal{U}$  is said to be a *countable generator* of the topology of  $X$ ).

Set  $X_0 = X$ . For  $i = 1, 2, 3, \dots$ , if  $\bigcup_{n=-\infty}^{\infty} T^{-n} U_i \supset X_{i-1}$ , set  $X_i = X_{i-1}$ , otherwise set  $X_i = X_{i-1} \setminus \bigcup_{n=-\infty}^{\infty} T^{-n} U_i$ .

Note that each  $X_i$  is closed and non-empty. Show that  $X_\infty = \bigcap_{i=0}^{\infty} X_i$  is a non-empty  $T$ -invariant set which is minimal for  $T$ .

11. Use the following steps to prove the following form of **Birkhoff's Recurrence Theorem**: *If  $T : X \rightarrow X$  is a continuous map on the compact metric space  $X$ , then  $T$  has a recurrent point* (i.e., there is a point  $x \in X$  and a sequence  $n_k \rightarrow \infty$  with  $T^{n_k} x \rightarrow x$  as  $k \rightarrow \infty$ ).

- (a) First suppose that the dynamical system  $(X, T)$  is minimal. Then if  $x \in X$ , there is a sequence  $n_k \rightarrow \infty$  and  $y \in X$  with  $T^{n_k} x \rightarrow y$  as  $k \rightarrow \infty$ .
- (b) Show that there is a sequence  $m_j$  with  $T^{m_j} y \rightarrow x$  as  $j \rightarrow \infty$ , and deduce that  $x$  is a recurrent point for  $T$ .
- (c) Use the result of the previous exercise, to show that we can drop the assumption of minimality.

12. Let  $\sigma : \Sigma \rightarrow \Sigma$  be the Bernoulli shift with usual metric ( $\Sigma = \{0, 1\}^{\mathbb{N}}$  - see Sections 4.2 and 6.5). Which of the following are minimal sets?

- (i)  $\{(x_n) \in \Sigma : x_n = 1, n \geq 0\}$  (ii)  $\{(x_n) \in \Sigma : x_n = 1, n \geq 1, x_0 = 0\}$ , (iii)\*  $\overline{O(u)}$ ,

where  $u$  is the sequence generated by the Morse substitution.

### 17.5 Topological Mixing and Exactness.

Our aim now is to give some simple criteria that will enable us to show more easily, that a dynamical system is chaotic. We define the notion of (topological) *mixing* for dynamical systems, and we show that it is a stronger condition than being transitive.

**Definition 17.5.1** Let  $T : X \rightarrow X$  be a continuous map on a compact metric space  $X$ .  $T$  is *mixing*, if for any non-empty open sets  $U$  and  $V$  in  $X$ , there exists  $N \in \mathbb{Z}^+$  with

$$U \cap T^n(V) \neq \emptyset, \text{ for all } n \geq N.$$

$T$  is mixing if iterates of any non-empty open set eventually become spread out “uniformly” over the whole space to intersect any other open set.

The following proposition follows immediately from Theorem 17.4.7 (c).

**Theorem 17.5.2** Let  $T : X \rightarrow X$  be a continuous map defined on a compact metric space  $X$ .

(a) If  $T$  is mixing, then  $T$  is transitive.

(b) If for every non-empty open set  $U \subset X$  there exists  $n \in \mathbb{Z}^+$  with  $T^n(U) = X$ , then  $T$  is mixing.

The converse of Theorem 17.5.2 (a) is false. For example, an irrational rotation is not mixing (see Exercise 17.5 # 2(b)). Maps that satisfy condition (b) in Theorem 17.5.2, are given a special name:

**Definition 17.5.3** A continuous map  $T : X \rightarrow X$  on a compact metric space  $X$  is *topologically exact* if for any non-empty open set  $U \subset X$ , there exists  $n \in \mathbb{Z}^+$  with  $T^n(U) = X$ .

For interval maps, Theorem 17.5.2 (b) can often be used to show that  $T$  is chaotic.

**Proposition 17.5.4** *Let  $f : [a, b] \rightarrow [a, b]$  be topologically exact. Then  $f$  is chaotic on  $[a, b]$ .*

**Proof.** We have seen that  $f$  is mixing and is therefore transitive. The periodic points are dense, for if  $I$  is open in  $[a, b]$ , there exists  $n > 0$  with

$$f^n(I) = [a, b] \supset I.$$

It follows that there exists  $x \in I$  with  $f^n(x) = x$ . Since this holds for any open interval, the periodic points must be dense.

Sensitive dependence on initial conditions now follows, since  $f$  is continuous on a metric space. Sensitive dependence can also be proved directly: if  $I$  is open with  $f^n(I) = [a, b]$  and  $x \in I$ , there exists  $y \in I$  with  $|f^n(x) - f^n(y)| \geq (b - a)/2$ .

□

An exact map  $f$  cannot have any attracting fixed points, since if  $p$  were such a point, its immediate basin of attraction  $B_f(p)$ , is an open set, invariant under  $f$ . This implies  $B_f(p) = X$ , contradicting the periodic points being dense in  $X$ .

**Examples 17.5.5** 1. The angle doubling map  $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $f(z) = z^2$  is mixing, since if  $I$  and  $J$  are intervals in  $\mathbb{S}^1$ , then  $f^n(J)$  is doubled in length after each iteration, and so  $I \cap f^n(J) \neq \emptyset$  for all  $n$  large enough. In fact, it is clear that  $f$  is exact.  $f$  is not minimal since  $z = 1$  is a fixed point.

2. The full tent map  $T = T_2$  is mixing. We demonstrate this by using Theorem 17.5.2 (b) together with properties of the tent map established in Example 6.4.2 (see Exercises 17.5).

3. Suppose that  $f : [0, 1] \rightarrow [0, 1]$  is a *unimodal map*. Here we mean that  $f$  is continuous on  $[0, 1]$ , increasing on  $[0, c]$ , decreasing on  $[c, 1]$ , with exactly one critical point at  $x = c$ . We have

**Proposition 17.5.6** *If the unimodal map  $f : [0, 1] \rightarrow [0, 1]$  is transitive, then  $f$  is either exact, or there is a fixed point  $d \in [0, 1]$  with  $f[0, d] = [d, 1]$ ,  $f[d, 1] = [0, d]$ .*

**Proof.** We sketch the proof: Since  $f$  is transitive, it has to be onto (see Exercises 17.4), and  $f(c) = 1$ . Now either  $f[c, 1] \subset [c, 1]$  or  $f[c, 1] \supset [c, 1]$ . In either case there must be a fixed point  $d \in [c, 1]$ .

We claim that since  $f$  is transitive,  $f(1) = 0$ . Suppose that  $f(1) = \alpha > 0$ , then  $f(\alpha) > \alpha$ , so that  $[\alpha, 1]$  is an invariant set, contradicting the transitivity of  $f$ . In particular  $f[c, 1] = [0, 1]$ . Now we can find  $c_1 > d$  with  $f(c_1) = c$ , so that if  $I_0 = [c, 1]$ ,

$I_1 = [c, c_1]$ ,  $I_2 = [c_2, c_1]$  etc., where  $c_2 < d$  and  $f(c_2) = c_1$ , we obtain a nested sequence of closed intervals

$$I_0 \supset I_1 \supset I_2 \supset \cdots,$$

each containing  $d$  with  $f(I_k) = I_{k-1}$ . By compactness, the intersection is non-empty, but cannot be an interval of positive length, as this would contradict the transitivity of  $f$  (the intersection being invariant), so  $\cap_k I_k = \{d\}$ . It follows that if we iterate any interval containing  $d$ , we must eventually get all of  $[0, 1]$ .

□

In fact, it has been shown (see [19]), that for continuous maps  $f : I \rightarrow I$ , where  $I$  is a compact interval, the following are equivalent: (i)  $f^2$  is transitive, (ii)  $f^n$  is transitive for every  $n > 0$ , (iii)  $f$  is transitive and has a point of odd period greater than 1, (iv)  $f$  is mixing.

4. It is a consequence of Theorem 17.4.7 that if  $f : X \rightarrow X$  is a chaotic dynamical system on a compact metric space, it can have no “non-trivial” invariant subsets. Specifically, if  $f(E) \subseteq E$  with  $E \neq X$  closed, then  $E$  is nowhere dense. Clearly any periodic point gives rise to a closed invariant subset which is nowhere dense. An example of a chaotic map which has a non-trivial invariant subset (namely the Cantor set, which is of course nowhere dense), is defined by:

$f : [0, 1] \rightarrow [0, 1]$  by

$$f(x) = \begin{cases} 3x; & x \in L = [0, 1/3] \\ 2 - 3x; & x \in N = [1/3, 2/3] \\ 3x - 2; & x \in R = [2/3, 1]. \end{cases}$$

We leave it as an exercise to show that  $f$  is mixing, and also chaotic. Let  $\Lambda$  be the set

$$\Lambda = \{x \in [0, 1] : f^n(x) \in L \cup R \text{ for all } n \in \mathbb{N}\}.$$

Clearly  $\Lambda = \bigcap_{n=0}^{\infty} f^{-n}(L \cup R)$ , a closed and bounded set. We claim that  $\Lambda = C$ , the Cantor set (see Exercises 17.5), and that  $C$  is a set invariant under  $f$ .

We end this chapter with a result mentioned in Chapter 7, namely that sensitive dependence is a conjugacy invariant when the spaces are compact. We saw in Example 7.2.2 that this is not generally true for non-compact spaces.

**Theorem 17.5.7** *Let  $(X, d_X)$  and  $(Y, d_Y)$  be compact metric spaces and  $f : X \rightarrow X$ ,  $g : Y \rightarrow Y$ , maps conjugate via a homeomorphism  $h : X \rightarrow Y$ . If  $g$  has sensitive dependence on initial conditions, then so does  $f$ .*

**Proof.** Since  $X$  and  $Y$  are compact spaces and  $h : X \rightarrow Y$  is continuous,  $h$  is *uniformly continuous* (see Exercises): given  $\epsilon > 0$ , there exists  $\delta > 0$  such that

$$d_X(x_1, x_2) < \delta \implies d_Y(h(x_1), h(x_2)) < \epsilon.$$

It follows that if  $d_Y(h(x_1), h(x_2)) \geq \epsilon$ , then  $d_X(x_1, x_2) \geq \delta$ , or

$$d_Y(y_1, y_2) \geq \epsilon \implies d_X(h^{-1}(y_1), h^{-1}(y_2)) \geq \delta.$$

Now let  $x \in X$ ,  $y = h(x)$  and  $\delta_1 > 0$ . Then there exists  $\epsilon_1 > 0$  such that if  $y_1 \in Y$  with

$$d_Y(y_1, h(x)) = d_Y(y_1, y) < \epsilon_1, \quad \text{then} \quad d_X(h^{-1}(y_1), x) < \delta_1.$$

Since  $g$  has sensitive dependence on initial conditions, there exists  $k \in \mathbb{Z}^+$ ,  $\epsilon > 0$  and  $y_2 \in Y$  with  $d_Y(y, y_2) < \epsilon_1$  and  $d_Y(g^k(y), g^k(y_2)) \geq \epsilon$ .

It follows that  $d_X(h^{-1}(g^k(y)), h^{-1}(g^k(y_2))) \geq \delta$ , ( $\delta$  chosen above). Since  $h^{-1} \circ g^k = f^k \circ h^{-1}$ , we deduce that

$$d_X(f^k(h^{-1}(y)), f^k(h^{-1}(y_2))) \geq \delta, \quad \text{or} \quad d_X(f^k(x), f^k(x_2)) \geq \delta,$$

(where  $h(x_2) = y_2$ ), concluding that  $f$  has sensitive dependence on initial conditions.  $\square$

### Exercises 17.5

1. (a) Prove that mixing and exactness are conjugacy invariants (i.e., show that if  $f : X \rightarrow X$  and  $g : Y \rightarrow Y$  are continuous maps on compact metric spaces and  $f, g$  are conjugate via a homeomorphism  $h : X \rightarrow Y$ , then  $f$  mixing implies  $g$  mixing).  
 (b) Prove that a factor of a mixing map is mixing.
2. (a) Show that if  $T : [0, 1] \rightarrow [0, 1]$  is the full tent map, then  $T$  is mixing.  
 (b) Let  $T : X \rightarrow X$  be a transitive isometry on an infinite metric space (so is minimal). Show that  $T$  is not mixing.
3. Show that the shift map  $\sigma : \Sigma \rightarrow \Sigma$ , where  $\Sigma = \{(a_1, a_2, \dots) : a_i \in \{0, 1\}\}$  and  $\sigma(a_1, a_2, a_3, \dots) = (a_2, a_3, \dots)$ , with its usual topology, is mixing.

4. (a) Prove that the function  $f : [0, 1] \rightarrow [0, 1]$  of Example 17.5.5, defined by

$$f(x) = \begin{cases} 3x; & x \in L = [0, 1/3] \\ 2 - 3x; & x \in N = [1/3, 2/3] \\ 3x - 2; & x \in R = [2/3, 1], \end{cases}$$

is mixing.

- (b) By considering the ternary expansion of  $x \in [0, 1] = L \cup N \cup R$ , show that the set  $\Lambda = \{x \in [0, 1] : f^n(x) \in L \cup R, \text{ for all } n \in \mathbb{N}\}$  is exactly the Cantor set, and that this  $\Lambda$  is invariant under  $f$ .

5. Let  $T$  be a continuous map on a compact metric space  $X$ . Give a direct proof that if  $T$  is topologically exact, then  $T$  has sensitive dependence on initial conditions.

6. Let  $T : X \rightarrow X$  be a continuous map on a compact metric space  $X$ . The map  $T \times T : X \times X \rightarrow X \times X$  is defined by  $T \times T(x, y) = (Tx, Ty)$ . This gives rise to the *direct product* dynamical system  $(X \times X, T \times T)$ . If  $d$  is a metric on  $X$ , a metric  $d'$  can be defined on  $X \times X$  so that if  $X$  is compact, then  $X \times X$  is also compact (see Exercises 17.1 # 6). We say that  $T$  is (topologically), *weakly-mixing* if  $T \times T$  is transitive. Assume that  $X$  has no isolated points, then the Birkhoff Transitivity Theorem may be used.

- (a) Prove that if  $T$  is mixing, then  $T$  is weakly-mixing. (Hint: Use the fact that if  $W$  is non-empty and open in  $X \times X$ , then there exists  $U_1$  and  $U_2$  non-empty, and open in  $X$  with  $U_1 \times U_2 \subset W$ ).
- (b) Prove that if  $T$  is weakly-mixing, then  $T$  is transitive.
- (c) Show that  $T$  is weakly-mixing if and only if for any non-empty open sets  $U_1, U_2, V_1, V_2$  in  $X$ , there exists  $n \geq 1$  such that  $T^n U_1 \cap V_1 \neq \emptyset$  and  $T^n U_2 \cap V_2 \neq \emptyset$ .

It has been shown that the Chacon substitution gives rise to a dynamical systems that is weakly mixing, but not mixing (see [102]).

7. (a) Let  $T : X \rightarrow X$  be a continuous, weakly-mixing and onto transformation on the compact metric space  $X$ . Show that if  $f(Tx) = \lambda f(x)$  for some  $\lambda \in \mathbb{C}$ , and  $f$  a non-zero, continuous function  $f : X \rightarrow \mathbb{C}$ , then  $\lambda = 1$  and  $f = \text{constant}$ . (Hint: Set

$g(x, y) = f(x)\bar{f}(y)$ , noting that  $|\lambda| = 1$  from Exercise 17.4 # 3. Show that  $g$  is  $T \times T$  invariant, and then apply Proposition 17.4.5). This result says that a weakly mixing transformation  $T$ , has no non-trivial *eigenfunctions*.

(b) Let  $R_a : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $R_a(z) = az$  be an irrational rotation. We have seen in Section 17.4 that  $R_a$  is transitive. Use (a) to show that  $R_a$  is not weakly-mixing. (Hint: Show that the maps  $f_n(z) = z^n$  are eigenfunctions of  $R_a$ ).

8. There is a converse to the result from 7(a) which is due to H. Keynes and J. Robertson [74], and independently, K. Petersen [102] (see also [57]). They showed that for  $T : X \rightarrow X$  minimal on a compact metric space  $X$ ,  $T$  is weakly mixing if and only if  $T$  has no non-constant continuous eigenfunctions.

(a) Show that if  $S, T : X \rightarrow X$  are minimal with  $T$  weakly mixing and  $ST = TS$ , then  $S$  is weakly mixing. (Hint: Use the remark above and Exercise 17.4 # 3(b)).

(b) Prove that if  $T^2$  is weakly mixing, then  $T$  is weakly mixing.

(c) Prove that if  $ST = T^{-1}S$  where  $S^2$  and  $T$  are minimal and  $T$  is a homeomorphism, then  $S$  and  $T$  are weakly mixing (see [62] for the measure theoretic version of this result).

9. Suppose that  $ST = T^{-1}S$  where  $T$  is a homeomorphism.

(a) If  $T$  is minimal, show that  $S^{2n} = I$  (the identity map), for some  $n > 0$ , or  $S$  has no periodic points.

(b) If  $S$  is minimal, show that  $T^n = I$  for some  $n > 0$ , or  $T$  has no periodic points.

(c) Show that if  $T : X \rightarrow X$  has an inverse map  $T^{-1}$ , then  $T \times T^{-1}$  is conjugate to its inverse via the map  $R(x, y) = (y, Tx)$ . Note that  $R^2 = T \times T$ . Deduce that the conjugating map need not have periodic points.

10. (a) Suppose that  $S : X \rightarrow X$  is minimal, but  $S^2$  is not minimal. Show that  $-1$  is an eigenvalue of  $S$ . Can you generalize this result to the case where  $S$  is minimal, but  $S^n$  ( $n > 2$ ), is not minimal?

- (b)\* Suppose that  $ST = T^{-1}S$  where  $T$  is a homeomorphism. If  $S$  and  $T$  are minimal, but not weakly mixing, show that  $-1$  is the unique eigenvalue of  $T$ .
- (c) Use (b) to show that if  $S$  and  $T^2$  are minimal, then  $S$  and  $T$  are weakly mixing.

11. Fill in the details in the proof of Proposition 17.5.6.

12. **Topological Joinings**, [56]. Let  $T : X \rightarrow X$  and  $S : Y \rightarrow Y$  be homeomorphisms of compact metric spaces. The product dynamical system is the pair  $(X \times Y, T \times S)$ , where  $T \times S : X \times Y \rightarrow X \times Y$  and  $T \times S(x, y) = (Tx, Sy)$ . A *joining* of  $T$  and  $S$  is a subsystem  $(W, T \times S)$  whose projections on both coordinates are *full*. This means that  $W$  is a non-empty, closed  $T \times S$ -invariant subspace of  $X \times Y$  with the property that if  $\pi_X : X \times Y \rightarrow X$  and  $\pi_Y : X \times Y \rightarrow Y$  are the projection maps ( $\pi_X(x, y) = x$  and  $\pi_Y(x, y) = y$ ), then  $\pi_X(W) = X$  and  $\pi_Y(W) = Y$ . The idea of joining is due to Furstenberg [53]. In order to study the relationship between two dynamical systems, we need to know the nature of all the joinings between the two systems ([57]).

For example, the system  $(X \times Y, T \times S)$  is always a joining of  $T$  and  $S$  (the *trivial joining*). Another example is given by *graph joinings*: let  $\phi : X \rightarrow Y$  be a continuous onto map with  $\phi \circ T = S \circ \phi$  (so  $S$  is a factor of  $T$ ). Set  $W = \{(x, \phi(x)) : x \in X\}$ .

- (a) Show that  $W$  is a closed  $T \times S$ -invariant subset of  $X \times Y$  whose projections on  $X$  and  $Y$  are full. Deduce that  $W$  is a joining of  $T$  and  $S$ .
- (b) The dynamical systems  $(X, T)$  and  $(Y, S)$  are *disjoint* if  $(X \times Y, T \times S)$  is the unique joining between the two systems. If both  $(X, T)$  and  $(Y, S)$  are minimal, show that they are disjoint if and only if  $(X \times Y, T \times S)$  is minimal.
- (c) If  $(X, T)$  and  $(Y, S)$  are disjoint, show that one of them is minimal. (Hint: If  $M$  and  $N$  are minimal subsets for  $T$  and  $S$ , show that  $(M \times Y) \cup (X \times N)$  is a joining of  $T$  and  $S$ ).
- (d) Suppose that  $X$  contains more than one point. Show that there is always a non-trivial joining between  $(X, T)$  and itself. Use (b) to deduce that  $T$  may be minimal, but  $T \times T$  is never minimal.

(e) Let  $(X, T)$  and  $(Y, S)$  be dynamical systems. Suppose that they have a common factor: there is a dynamical system  $(Z, R)$  and continuous onto maps  $\phi_1 : X \rightarrow Z$ ,  $\phi_2 : Y \rightarrow Z$  with  $\phi_1 \circ T = R \circ \phi_1$  and  $\phi_2 \circ S = R \circ \phi_2$ . Show that  $W = \{(x, y) \in X \times Y : \phi_1(x) = \phi_2(y)\}$  is a joining of  $T$  and  $S$ . Deduce that if  $T$  and  $S$  are disjoint, then they have no common factor.

13. Give a non-trivial joining of the rotations  $R_a, R_b : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $R_a(z) = az$  and  $R_b(z) = bz$ , with  $b = a^2$ . If they are both minimal, deduce that  $R_a \times R_b$  is not minimal.

## CHAPTER 18

### Substitution Dynamical Systems.

In Chapter 17 we introduced the notion of compactness in metric spaces. The aim was to study a particular class of topological dynamical systems in more detail, to obtain general properties that would be easily applied to examples such as the logistic and tent maps. In this chapter, we apply these results to substitution dynamical systems. Substitutions were introduced in Chapter 15 in a non-rigorous way. We now aim to give a rigorous foundation to the theory of substitutions, and substitutions dynamics. First we show that with a suitable metric, the space of all sequences  $\Sigma$ , on a finite alphabet is a compact metric space. This is a special case of *Tychonoff's Theorem*.

If  $u \in \Sigma$  is a fixed point of a substitution  $\theta$ , then it follows that the orbit closure  $\overline{O(u)}$  of  $u$  under the shift map  $\sigma$  is a compact invariant subspace of  $\Sigma$ . Hence  $(\overline{O(u)}, \sigma)$  is a compact topological dynamical system, or what we call a *substitution dynamical system*. This is a type of *symbolic dynamical system*, being the action of the shift map on a set of infinite sequences of symbols. Substitution systems often exhibit a type of chaos, being transitive, with sensitive dependence, but without having periodic points. Other types of symbolic systems (for example the full shift map), have all three properties of Devaney chaos.

#### 18.1 Sequence Spaces.

The general form of Tychonoff's Theorem states: *Cartesian products of compact spaces are compact*. A proof will take us too far afield, so we give a direct proof for the case of interest to us:

Set  $\mathcal{A} = \{0, 1, \dots, s - 1\}$ ,  $s \geq 2$ , a finite set which we call the *alphabet*. As usual  $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ , is the set of natural numbers.  $\mathcal{A}^{\mathbb{N}}$  is the set of all one-sided infinite sequences beginning with the index 0. We write  $\omega \in \mathcal{A}^{\mathbb{N}}$ ,

$$\omega = \omega_0 \omega_1 \omega_2 \omega_3 \omega_4 \omega_5 \dots, \quad (\omega)_i = \omega_i,$$

More correctly, we should write  $\omega = (\omega_0, \omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \dots)$ , but we often use this abuse of notation.

Define a metric on  $\mathcal{A}^{\mathbb{N}}$  by

$$d(\omega, \omega') = 2^{-\min\{n \in \mathbb{N}: \omega_n \neq \omega'_n\}}, \quad \text{if } \omega \neq \omega', \quad \text{and} \quad d(\omega, \omega') = 0 \quad \text{if } \omega = \omega'.$$

We see that two points are close when their initial terms are equal. This metric is different to the metric used on  $\mathcal{A}^{\mathbb{N}}$  in Chapter 5, but has similar properties.

**Theorem 18.1.1** *The space  $\mathcal{A}^{\mathbb{N}}$  is a compact metric space. In particular, it is complete, totally bounded and sequentially compact.*

**Proof.** The fact that  $\mathcal{A}^{\mathbb{N}}$  is a metric space is similar to the proof given in Chapter 4, and is left as an exercise. We show that  $\mathcal{A}^{\mathbb{N}}$  is sequentially compact, and hence is compact.

Let  $\omega^n = (a_0^n, a_1^n, a_2^n, \dots)$  be an infinite sequence in  $\mathcal{A}^{\mathbb{N}}$ . We will show that  $(\omega^n)$  has a limit point  $\omega = (a_0, a_1, a_2, \dots)$ .

Since  $\mathcal{A}$  is a finite set, there exists  $a_0 \in \mathcal{A}$  such that

$$A_0 = \{n \in \mathbb{N} : a_0^n = a_0\}$$

is an infinite set. We construct the sequence  $(a_0, a_1, a_2, \dots)$  inductively as follows:

(i) We have found  $a_0$  above.

(ii) Suppose that for  $m > 0$  we have found  $a_k \in \mathcal{A}$ ,  $k = 0, 1, \dots, m-1$ , and infinite sets  $A_0, A_1, \dots, A_{m-1}$  contained in  $\mathbb{N}$  satisfying

$$A_0 \subseteq A_1 \subseteq A_2 \subseteq \dots \subseteq A_{m-1},$$

and such that  $a_k^n = a_k$  for all  $k \in A_k$ ,  $k = 0, \dots, m-1$ .

If we can find a set  $A_m$  and  $a_m \in \mathcal{A}$  such that

$$A_m = \{n \in A_{m-1} : a_m^n = a_m\}$$

is an infinite set, then by the principle of induction, we have constructed the sequence

$$\omega = (a_0, a_1, a_2, \dots),$$

which we show is a limit point of our given sequence.

Let  $\delta > 0$  and choose  $n \in \mathbb{N}$  so large that  $1/2^n < \delta$ . We need to show that  $\{\omega^n : n \in \mathbb{N}\} \cap B_{\delta}(\omega)$  contains points besides  $\omega$ . It suffices to show that a member of  $\{\omega^n : n \in \mathbb{N}\}$  coincides with  $\omega$  in its first  $n+1$  coordinates.

Let  $p \in \cap_{k=0}^n A_k$ . Then  $p \in A_0$ , and  $a_0^p = a_0$ ,  $p \in A_1$ ,  $a_1^p = a_1$ ,  $\dots$ ,  $p \in A_n$ , so  $a_n^p = a_n$ , i.e.,

$$\omega^p = (a_0^p, a_1^p, a_2^p, \dots, a_n^p, \dots),$$

coincides with  $\omega$  in the first  $n+1$  coordinates, so  $d(\omega^p, \omega) \leq 1/2^{n+1} < 1/2^n < \delta$  and the result follows.  $\square$

The following result says that  $\mathcal{A}^{\mathbb{N}}$  is a type of Cantor set. Recall that we defined a subset of  $\mathbb{R}$  to be totally disconnected if it contains no open subsets. We give a more general version of this notion for metric spaces. A metric space  $X$  is said to be *disconnected* if there exists a *clopen* subset of  $X$  (i.e., a set which is both open and closed), other than  $X$  or  $\emptyset$ . In other words, there exist  $A$  and  $B$  open, with  $A \neq \emptyset$ ,  $B \neq \emptyset$ ,  $A \cap B = \emptyset$  and  $A \cup B = X$ . If  $X$  is not disconnected, it is *connected*. A subset  $Z$  of  $X$  is disconnected if there are open sets  $A$  and  $B$  that have non-empty intersection with  $Z$ , and satisfy  $Z \subseteq A \cup B$  and  $A \cap B \cap Z = \emptyset$ . A subset of  $X$  is connected if it is not a disconnected subset.

For example,  $\mathbb{R}$  with its usual metric is connected, but the rationals, the Cantor set  $C$ , and the set of irrationals (each with the metric coming from  $\mathbb{R}$ ), are not connected. In fact, they are what we call *totally disconnected*.

A subset  $Z \subseteq X$  is a *component* of  $X$  if it is maximally connected, i.e.,  $Z$  is a connected set, and if  $Z \subseteq Y$  with  $Y$  connected, then  $Z = Y$ . It can be shown that the components of  $X$  partition  $X$  into clopen sets.

**Definition 18.1.2** The metric space  $(X, d)$  is *totally disconnected* if the only components of  $X$  are singletons (i.e., sets of the form  $\{x\}$  for  $x \in X$ ).

**Theorem 18.1.3** *The metric space  $\mathcal{A}^{\mathbb{N}}$  is totally disconnected, and every member is a limit point.*

**Proof.** Let  $\omega \in \mathcal{A}^{\mathbb{N}}$ . Then  $\omega$  is a limit point since we can construct a sequence  $(\omega^n)$  in  $\mathcal{A}^{\mathbb{N}}$  so that  $\omega^n$  is equal to  $\omega$  in the first  $n$ -coordinates, but with  $\omega^n \neq \omega$  and  $\omega^n \neq \omega^m$  for every  $m, n \in \mathbb{N}$ ,  $m \neq n$ . Then it is clear that  $d(\omega^n, \omega) \rightarrow 0$  as  $n \rightarrow \infty$ .

To show that  $\mathcal{A}^{\mathbb{N}}$  is totally disconnected, let  $Z \subseteq \mathcal{A}^{\mathbb{N}}$  be a component containing at least two points, say  $\omega, \omega' \in Z$  with  $\omega \neq \omega'$ . Suppose that  $\omega_k = \omega'_k$  for  $k = 0, 1, \dots, n-1$ , but  $\omega_n \neq \omega'_n$ . Set

$$U = \{x \in \mathcal{A}^{\mathbb{N}} : d(x, \omega) < 1/2^n\} = \{x \in \mathcal{A}^{\mathbb{N}} : d(x, \omega) \leq 1/2^{n+1}\}.$$

Then  $U$  is both open and closed,  $\omega \in U$ ,  $\omega' \notin U$  (see Exercises 18.1). This contradicts  $Z$  being a connected component.

□

Applying results from this section and Section 17.3, we are able to show the existence of fixed points of substitutions:

**Theorem 18.1.4** *Let  $\theta$  be a substitution on  $\mathcal{A} = \{0, 1, \dots, s - 1\}$ , ( $s > 1$ ), extended to be a function  $\theta : \mathcal{A}^{\mathbb{N}} \rightarrow \mathcal{A}^{\mathbb{N}}$ . Then*

- (a)  *$\theta$  is a continuous function.*
- (b) *Corresponding to each  $a \in \mathcal{A}$  with  $|\theta(a)| > 1$ , and for which  $\theta(a)$  starts with  $a$ ,  $\theta$  has a unique fixed point  $u_a \in \mathcal{A}^{\mathbb{N}}$ .*
- (c) *If both  $w \in \mathcal{A}^{\mathbb{N}}$  and  $\theta(a)$  start with  $a \in \mathcal{A}$ , ( $|\theta(a)| > 1$ ), then  $u_a = \lim_{n \rightarrow \infty} \theta^n(w)$ , is a fixed point of  $\theta$ .*

**Proof.** (a) We may assume that  $\theta(0)$  starts with 0. Set  $X_0 = \{\omega \in \mathcal{A}^{\mathbb{N}} : (\omega)_0 = 0\}$ .  $X_0$  is a closed compact subspace of  $\mathcal{A}^{\mathbb{N}}$ . For suppose  $\omega^n \rightarrow \omega$  as  $n \rightarrow \infty$ , with  $\omega^n \in X_0$  and  $(\omega^n)_0 \neq 0$ , then clearly  $d(\omega^n, \omega) = 1$ , and this is impossible.

Now  $d(\theta(\omega), \theta(\omega')) \leq d(\omega, \omega')$  for all  $\omega, \omega' \in \mathcal{A}^{\mathbb{N}}$ , and so  $\theta$  is a continuous function.

(b) Clearly  $X_0$  is invariant under  $\theta$ , and we can think of  $\theta$  as a continuous map on a compact space  $\theta : X_0 \rightarrow X_0$ . Let  $\omega, \omega' \in X_0$ , and suppose that  $d(\omega, \omega') = 1/2^k$  for some  $k > 0$ . Then clearly  $d(\theta(\omega), \theta(\omega')) < d(\omega, \omega')$  since  $(\theta(\omega))_i = (\theta(\omega'))_i$  for  $i = 0, 1, \dots, kr$ , where  $r \geq |\theta(0)| > 1$ .

Thus, we can apply Theorem 17.3.1 to deduce that  $\theta$  has a unique fixed point in  $X_0$ , and the result follows as we can do this for each  $a \in \mathcal{A}$  where  $\theta(a)$  starts with  $a$ .

(c) From (b), again assuming  $a = 0$ ,  $\theta$  has a unique fixed point  $u$  in  $X_0$  starting with 0. Then

$$d(\theta^n(w), u) = d(\theta^n(w), \theta^n(u)) \leq \frac{1}{2^{nr}} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and the result follows.

□

**Examples 18.1.5** 1. From statement (c) in Theorem 18.1.4, we see that if we iterate any  $w \in \mathcal{A}^{\mathbb{N}}$ , where both  $w$  and  $\theta(a)$  start with  $a$ , then in the limit we get a fixed

point of  $\theta$ . In Chapter 15, we saw that if the first letter of  $\theta(a)$  is  $a$  and  $\theta(a)$  is a word having length at least two, then  $u = \lim_{n \rightarrow \infty} \theta^n(a)$  is a fixed point of  $\theta$ . However,  $\theta^n(a)$  does not belong to  $\mathcal{A}^{\mathbb{N}}$ , so this procedure is not well defined. To make this rigorous we could instead iterate any sequence  $w \in \mathcal{A}^{\mathbb{N}}$ , which starts with  $a$ .

The Thue-Morse sequence arises from the substitution on  $\mathcal{A} = \{0, 1\}$ :

$$\theta(0) = 01, \quad \theta(1) = 10,$$

giving two fixed points in  $\{0, 1\}^{\mathbb{N}}$ . These are

$$u = 011010011001\dots, \quad \text{and} \quad R(u) = 100101100110\dots.$$

2. Let  $\mathcal{A} = \{0, 1\}$  and  $\theta(0) = 010, \theta(1) = 101$ . Then we have two fixed points which are both periodic sequences. This substitution is not interesting as we shall see that periodic sequences give rise to trivial dynamical behavior. For  $\theta(0) = 101, \theta(1) = 010$ , we have no fixed points (but we do have period 2-points), and again trivial dynamics.

3. The Fibonacci substitution is defined on  $\mathcal{A} = \{0, 1\}$  by  $\theta(0) = 01, \theta(1) = 0$ . It can be seen that  $\theta$  has a unique fixed point:

$$u = 0100101001001\dots$$

### Exercises 18.1

1. Suppose  $\mathcal{A} = \{0, 1, 2, \dots, s - 1\}$ ,  $s \geq 2$ , is an alphabet, and  $\mathcal{A}^{\mathbb{N}}$  is the set of all sequences with values in  $\mathcal{A}$  ( $\mathcal{A}^{\mathbb{N}} = \{\omega = (a_0, a_1, a_2, \dots) : a_i \in \mathcal{A}\}$ ). Prove that if we define a distance  $d$  on  $\mathcal{A}^{\mathbb{N}}$  by

$$d(\omega_1, \omega_2) = 2^{-\min\{k \geq 0 : a_k \neq b_k\}}, \quad \omega_1 \neq \omega_2, \quad d(\omega_1, \omega_1) = 0$$

where  $\omega_1 = (a_0, a_1, a_2, \dots)$  and  $\omega_2 = (b_0, b_1, b_2, \dots)$ , then

(i)  $d$  defines a metric on  $\mathcal{A}^{\mathbb{N}}$ ,

(ii) if  $d(\omega_1, \omega_2) < 1/2^n$ , then  $a_0 = b_0, a_1 = b_1, \dots, a_{n-1} = b_{n-1}$ , and if  $a_0 = b_0, a_1 = b_1, \dots, a_{n-1} = b_{n-1}$ , then  $d(\omega_1, \omega_2) \leq 1/2^n$ ,

(iii) if  $\alpha \in \mathcal{A}^{\mathbb{N}}$ , then

$$\{\omega \in \mathcal{A}^{\mathbb{N}} : d(\alpha, \omega) < 1/2^{n-1}\} = \{\omega \in \mathcal{A}^{\mathbb{N}} : d(\alpha, \omega) \leq 1/2^n\}.$$

2. (i) Given  $\alpha_0, \alpha_1, \dots, \alpha_{n-1} \in \mathcal{A}$ , the set

$$C = [\alpha_0, \alpha_1, \dots, \alpha_{n-1}] = \{\omega = (a_0, a_1, a_2, \dots) \in \mathcal{A}^{\mathbb{N}} : a_j = \alpha_j, 0 \leq j \leq n-1\},$$

is called a *cylinder set*. Prove that cylinder sets are both open and closed (called *clopen*), and in fact if  $\alpha \in C$

$$C = \{\omega \in \mathcal{A}^{\mathbb{N}} : d(\alpha, \omega) < 1/2^{n-1}\} = \{\omega \in \mathcal{A}^{\mathbb{N}} : d(\alpha, \omega) \leq 1/2^n\}.$$

(ii) Show that any clopen set is a finite union of cylinder sets.

3. Let  $\Sigma = \mathcal{A}^{\mathbb{N}}$ , where  $\mathcal{A} = \{0, 1\}$  with the metric given in this section. Define a map  $T : \Sigma \rightarrow \Sigma$  by

$$T(a_0, a_1, a_2, a_3, \dots) = 1 + (a_0, a_1, a_2, a_3, \dots).$$

Two examples of the addition are:

$$1 + (0, 0, 0, \dots) = (1, 0, 0, \dots), \quad 1 + (1, 1, 0, 0, 1, 0, 0, 0, \dots) = (0, 0, 1, 0, 1, 0, 0, 0, \dots),$$

(carry to the right).  $T$  is called the *adding machine* (see [79]), a type of *group rotation*.

(a) Show that  $T : \Sigma \rightarrow \Sigma$  is a homeomorphism.

(b) Show that the orbit of  $(0, 0, 0, \dots)$  is dense under  $T$ , so  $T$  is transitive (in fact  $T$  is minimal).

(c) If  $f : \Sigma \rightarrow \mathbb{C}$  is defined by  $f(a_0, a_1, a_2, \dots) = e^{\pi i a_0}$ , show that  $f$  is a continuous eigenfunction of  $T$ , and find the corresponding eigenvalue. Deduce that  $T$  is not weakly-mixing (see Exercises 17.4 and 17.5).

(d)\* Show that the  $2^n$ th roots of unity are eigenvalues of  $T$ .

4. Let  $\theta : \mathcal{A}^{\mathbb{N}} \rightarrow \mathcal{A}^{\mathbb{N}}$  be the map induced by the Morse substitution  $\theta(0) = 01$ ,  $\theta(1) = 10$ . Is  $\theta$  one-to-one or onto?

5. (a) Prove that the metric space  $(X, d)$  is connected if and only if the only sets that are both open and closed in  $X$  are  $X$  and the empty set.
- (b) Show that if  $(X, d)$  is a metric space having at least two points and the discrete metric, then  $X$  is a totally disconnected set.
6. Prove that a connected metric space  $(X, d)$ , with at least two points is uncountable. (Hint: Let  $a, b \in X$ ,  $a \neq b$  and  $\lambda \in (0, 1)$ . Suppose there is no  $x \in X$  with  $d(a, x) = \lambda d(a, b)$ . Consider  $A = \{y \in X : d(a, y) < \lambda d(a, b)\}$  and  $B = \{y \in X : d(a, y) > \lambda d(a, b)\}$ ).
7. Prove that  $A \subset \mathbb{R}$  is totally disconnected, if and only if  $A$  contains no non-empty open intervals. Deduce that the rationals and the Cantor set are totally disconnected subsets of  $\mathbb{R}$ .

## 18.2 Languages.

Some basic properties of sequences arising from substitutions are given. See the books by N. Pytheas Fogg [45], and M. Queffelec [104] for additional detail and more advanced material.

### 18.2.1 Languages and Words.

Let  $\mathcal{A} = \{0, 1, 2, \dots, s-1\}$ ,  $s \geq 2$ , be a finite set.  $\mathcal{A}$  is the *alphabet* of the sequence under consideration, and its members are the *letters* of the alphabet. A *word* or *finite block* of members of  $\mathcal{A}$  is a finite string of letters whose *length* is the number of letters in the string. We denote the set of all finite words using the alphabet  $\mathcal{A}$  by  $\mathcal{A}^*$ . This includes the *empty word* (the word having 0 length), which is denoted by  $\epsilon$ .

### 18.2.2 The Complexity Function of a Sequence.

**Definition 18.2.3** Suppose that  $u = u_0u_1u_2\dots = (u_n)_{n \in \mathbb{N}}$  is a one-sided infinite sequence of letters from  $\mathcal{A}$  (we think of  $u$  as being a member of  $\mathcal{A}^\mathbb{N}$ ). The *language* of the sequence  $u$  is the set  $\mathcal{L}(u)$  of all finite words that appear in the sequence  $u$ . We use  $\mathcal{L}_n(u)$  to denote the subset of  $\mathcal{L}(u)$  consisting of those words of length  $n$ . We call a finite word appearing in the sequence  $u$ , a *factor* of  $u$ .

We write  $p_u(n) = |\mathcal{L}_n(u)|$ , the number of different factors of length  $n$ .  $p_u(n)$  is called the *complexity function* of the sequence. It can be thought of as measuring the

“randomness” of the sequence. The greater the variety of different factors appearing, the more complex is the sequence. For example, a sequence which is periodic has little complexity (see [45]).

Clearly  $p_u(n) \leq p_u(n+1)$  for  $n \in \mathbb{N}$ , and  $1 \leq p_u(n) \leq s^n$  where  $s$  is the number of members of the alphabet.

**Definition 18.2.4** The sequence  $u = (u_n)_{n \in \mathbb{N}}$  is *periodic*, if there exists  $N \geq 1$  such that  $u_n = u_{n+N}$  for all  $n \in \mathbb{N}$ . The minimum such  $N$  is the *period* of the sequence. The sequence is *ultimately periodic* if it is periodic from some term onwards. If the sequence is neither periodic, nor ultimately periodic, we say it is *aperiodic*.

**Examples 18.2.5** We will show in Chapter 19, that the sequence  $u$  arising from the Fibonacci substitution,  $\theta(0) = 01$ ,  $\theta(1) = 0$  has complexity function  $p_u(n) = n + 1$  for all  $n \in \mathbb{N}$ . A sequence with this property is said to be a *Sturmian sequence*. The Morse sequence can be shown to have complexity function satisfying  $p_u(n) \leq C \cdot n$  for some constant  $C > 0$ , for all  $n \in \mathbb{N}$ . It is not a Sturmian sequence.

The next result tells us that sequences with a bounded complexity function have to be periodic. In particular, Sturmian sequences have the minimal complexity of all aperiodic sequences. Neither the Fibonacci nor the Morse sequence is periodic or eventually periodic. This proposition is an early result due to Coven and Hedlund [30].

**Proposition 18.2.6** *If  $u = (u_n)_{n \in \mathbb{N}}$  is a periodic or an ultimately periodic sequence, then the complexity function  $p_u(n)$  is a bounded sequence. If there exists  $n \in \mathbb{N}$  such that  $p_u(n) \leq n$ , then  $u$  is an ultimately periodic sequence.*

**Proof.** If  $u$  is a periodic sequence with period  $n$  and  $|\mathcal{L}_n(u)| = m$ , then  $|\mathcal{L}_{n+1}(u)| = m$ , because a letter can be added to a member of  $\mathcal{L}_n(u)$  in only one way. We see that  $|\mathcal{L}_k(u)| = m$  for all  $k \geq n$ . Similar considerations apply to ultimately periodic sequences.

On the other hand, assume that  $p_u(n) \leq n$  for some  $n \in \mathbb{N}$ . We may assume that  $p_u(1) \geq 2$ , for otherwise  $u$  is a constant sequence. Then we have

$$2 \leq p_u(1) \leq p_u(2) \leq \cdots \leq p_u(n) \leq n.$$

Here we have an increasing sequence of  $n$  integers, the smallest being at least 2 and the largest less than or equal to  $n$ . By the pigeon-hole principle we must have  $p_u(k+1) = p_u(k)$  for some  $k$ . If  $\omega = \omega_0 \omega_1 \dots \omega_{k-1} \in \mathcal{L}_k(u)$ , then there is at least one factor of  $u$  of the form  $\omega a$  occurring in  $\mathcal{L}_{k+1}(u)$  for some letter  $a \in \mathcal{A}$ . Since  $p_u(k+1) = p_u(k)$  there

can be at most one such factor. It follows that if  $u_i u_{i+1} \dots u_{i+k-1}$  and  $u_j u_{j+1} \dots u_{j+k-1}$  are two factors of length  $k$  in  $u$  with  $u_i u_{i+1} \dots u_{i+k-1} = u_j u_{j+1} \dots u_{j+k-1}$ , then  $u_{i+k} = u_{j+k}$  since the additional letter can be added in only one way. Since there are only finitely many factors of length  $k$ , we must have  $u_i u_{i+1} \dots u_{i+k-1} = u_j u_{j+1} \dots u_{j+k-1}$  for some  $j > i$ . If we repeatedly add a letter to the right in a unique way, then delete a letter from the left to go back to a factor of length  $k$ , we see that  $u_{i+p} = u_{j+p}$  for every  $p \geq 0$ , and hence we have a periodic sequence.

□

## Exercises 18.2

1. Give an example of a sequence on  $\mathcal{A} = \{0, 1\}$  having complexity function  $p_u(n) = 2^n$  for all  $n \geq 1$ .
2. (a) An infinite sequence  $u = u_0 u_1 \dots$  is *square free* if there is no word of the form  $w w$  appearing in  $u$ . Show that an infinite sequence on a two-letter alphabet cannot be square free.  
(b) It can be shown that the Thue-Morse sequence is *cube free*. In this direction, show that the words 000 and 111 never appear in the Thue-Morse sequence
3. Show that the complexity function  $p_u(n)$  of a sequence satisfies  $p_u(m+n) \leq p_u(m)p_u(n)$ , for  $m, n \geq 1$ .

### 18.3 Dynamical Systems Arising from Sequences.

Our aim is to define the dynamical system arising from a substitution  $\theta$  on a finite alphabet  $\mathcal{A} = \{0, 1, \dots, s-1\}$ ,  $s \geq 2$ . We first examine a more general situation of a dynamical system arising from some arbitrary sequence  $u \in \mathcal{A}^{\mathbb{N}}$ . Denote by  $\sigma : \mathcal{A}^{\mathbb{N}} \rightarrow \mathcal{A}^{\mathbb{N}}$  the shift map defined in the usual way:

$$\sigma(\omega_0, \omega_1, \omega_2, \dots) = (\omega_1, \omega_2, \dots) \quad \text{or} \quad \sigma((\omega_n)_{n \in \mathbb{N}}) = (\omega_{n+1})_{n \in \mathbb{N}}.$$

We have seen that the space  $\mathcal{A}^{\mathbb{N}}$  with the metric  $d$  given previously, is a compact metric space, and is therefore complete and totally disconnected. Clearly the shift

map is onto, but not one-to-one on  $\mathcal{A}^{\mathbb{N}}$ . Since  $d(\sigma(\omega_1), \sigma(\omega_2)) \leq 2d(\omega_1, \omega_2)$  for all  $\omega_1, \omega_2 \in \mathcal{A}^{\mathbb{N}}$ ,  $\sigma$  is continuous.

Let  $u = u_0u_1u_2u_3\dots \in \mathcal{A}^{\mathbb{N}}$  be a sequence. The orbit of  $u$  is the set  $O(u) = \{u, \sigma(u), \sigma^2(u), \dots, \sigma^n(u), \dots\}$ . The orbit closure  $\overline{O(u)}$  is a closed subset of  $\mathcal{A}^{\mathbb{N}}$  and hence it is compact as a subset of  $\mathcal{A}^{\mathbb{N}}$ , so is a compact metric space in its own right. Note that  $\overline{O(u)}$  is a finite set if and only if  $u$  is a shift periodic sequence (i.e.,  $\sigma^k(u) = u$  for some  $k > 0$ ).

**Definition 18.3.1** The dynamical system associated with a sequence  $u = u_0u_1u_2\dots \in \mathcal{A}^{\mathbb{N}}$  is  $(\overline{O(u)}, \sigma)$ , the shift map restricted to the orbit closure of  $u$ .

**Lemma 18.3.2**  $w \in \overline{O(u)}$ , if and only if there is a sequence of integers  $(k_n)_{n \geq 1}$  with  $\lim_{n \rightarrow \infty} \sigma^{k_n}(u) = w$ .

**Proof.** If  $w \in \overline{O(u)}$ , then  $w$  is a limit point of  $O(u)$ . For each  $n \in \mathbb{N}$ , the set  $B_{1/n}(w) \cap O(u)$  contains an element other than  $w$ . Choose one such point  $\sigma^{k_n}(u)$ , then  $d(w, \sigma^{k_n}u) < 1/n$ , so  $w = \lim_{n \rightarrow \infty} \sigma^{k_n}(u)$ .

Conversely, if  $\epsilon > 0$  and  $w = \lim_{n \rightarrow \infty} \sigma^{k_n}(u)$ , then there exists  $N \in \mathbb{N}$  with  $d(w, \sigma^{k_n}u) < \epsilon$  for all  $n > N$ . It follows that  $w$  is a limit point of  $O(u)$ , or  $w \in \overline{O(u)}$ .  $\square$

**Proposition 18.3.3** The shift map  $\sigma : \overline{O(u)} \rightarrow \overline{O(u)}$  is a continuous, well defined map.

**Proof.** We have seen that  $\sigma$  is continuous on  $\mathcal{A}^{\mathbb{N}}$ .  $\overline{O(u)}$  is an invariant subset, for if  $w \in \overline{O(u)}$ , then  $\omega = \lim_{n \rightarrow \infty} \sigma^{k_n}(u)$  for some sequence  $(k_n)$ . By the continuity of  $\sigma$ ,  $\sigma(\omega) = \lim_{n \rightarrow \infty} \sigma^{k_n+1}(u) \in \overline{O(u)}$ . Thus, the restriction of  $\sigma$  to  $\overline{O(u)}$  is well defined. Since  $\sigma$  is continuous on  $\mathcal{A}^{\mathbb{N}}$ , it will be continuous on  $\overline{O(u)}$ .  $\square$

The above dynamical system is an example of a *sub-shift*. The shift map  $\sigma$  on  $\mathcal{A}^{\mathbb{N}}$  is then the *full shift*. There are ways of constructing sub-shifts without taking the orbit closure of sequences. See for example [77].

We can give an alternative description of the set  $\overline{O(u)}$ , using the language  $\mathcal{L}(u)$  of  $u$  as follows:

**Proposition 18.3.4** Let  $u = u_0u_1\dots$ . For every sequence  $\omega \in \overline{O(u)}$ , the following are equivalent:

- (i)  $\omega \in \overline{O(u)}$ .
- (ii) there exists a sequence  $k_n$  such that  $\omega_0\omega_1\dots\omega_n = u_{k_n}u_{k_n+1}\dots u_{k_n+n}$ .
- (iii)  $\mathcal{L}_n(\omega) \subset \mathcal{L}_n(u)$  for every  $n \in \mathbb{N}$ .

**Proof.** Suppose that  $\omega \in \overline{O(u)}$ , then for each  $n \in \mathbb{N}$ , there exists  $k_n \in \mathbb{N}$  with  $d(\omega, \sigma^{k_n}(u)) < 1/2^n$ . This implies that  $\omega_0\omega_1\dots\omega_n = u_{k_n}u_{k_n+1}\dots u_{k_n+n}$ , so (ii) holds. Conversely, if (ii) holds, then (i) follows in a similar way.

If (ii) holds then every word of length  $n$  appearing in  $\omega$  appears in  $u$ , so (iii) holds. Conversely, if (iii) holds then (ii) must hold. □

**Corollary 18.3.5**  $\overline{O(u)} = \{\omega \in \mathcal{A}^{\mathbb{N}} : \mathcal{L}(\omega) \subset \mathcal{L}(u)\}$ .

**Proof.** If  $\omega \in \overline{O(u)}$ , then from Proposition 18.3.4, every word appearing in  $\omega$  appears in  $u$ , so that  $\mathcal{L}(\omega) \subset \mathcal{L}(u)$ . On the other hand, if  $\omega \in \mathcal{A}^{\mathbb{N}}$  with  $\mathcal{L}(\omega) \subset \mathcal{L}(u)$ , then clearly  $\mathcal{L}_n(\omega) \subset \mathcal{L}_n(u)$  for every  $n$ , and the result follows from the same proposition. □

Below, we define the fundamental notion of *almost periodicity* for a sequence  $u \in \mathcal{A}^{\mathbb{N}}$ .

**Definition 18.3.6** The sequence  $u$  is *recurrent* if every factor of  $u$  appears infinitely often. It is *uniformly recurrent* or *almost periodic*, if every factor appears infinitely often with bounded gaps. More precisely, if  $\omega = \omega_0\omega_1\dots\omega_\ell$  is a factor of  $u$ , there exists  $K > 0$  and a sequence of integers  $k_n$  with  $(n-1)K \leq k_n < nK$ ,  $n = 1, 2, \dots$ , such that  $u_{k_n}u_{k_n+1}\dots u_{k_n+\ell} = \omega$ .

It will be clear that a (shift) periodic sequence is almost periodic, but the converse is not true. Almost periodic sequences are often said to be *minimal*.

**Examples 18.3.7** We saw in Chapter 15 (exercises), that the Thue-Morse sequence is recurrent. The same argument shows that the Thue-Morse sequence is almost periodic. In the next section, we give a criterion for a sequence arising from a substitution to be almost periodic.

Here is an example of a sequence  $u \in \mathcal{A}^{\mathbb{N}}$ ,  $\mathcal{A} = \{0, 1\}$ , which is recurrent but not almost periodic. The idea is to make sure every word is repeated infinitely often, but with unbounded gaps between repetitions. Start with  $u_0 = 0$ , then  $u_1 = 010$ , and  $u_2 = 010111010$ , and then  $u_4 = 0101110101111111010111010$ .

In other words, to move from  $u_n$  to  $u_{n+1}$ , we define  $u_{n+1}$  to be the concatenation  $u_{n+1} = u_n w_n u_n$ , where  $w_n = 111\dots 11$ , a string of 1's having length equal to the number of 1's appearing in  $u_n$ . Thus

$$u = 010111010111111101011101011111111111\dots,$$

Each initial segment recurs, so  $u$  is recurrent, but the gap between certain words grows exponentially, thus  $u$  is not almost periodic. Such sequences cannot give rise to minimal dynamical systems.

We can now show that dynamical systems arising from almost periodic sequences are minimal, and have some of the properties of chaotic maps (transitive and sensitive dependence).

Recall that the dynamical system  $(X, T)$  is *transitive* if there is a point  $x_0 \in X$  having a dense orbit, and it is *minimal* if every point has a dense orbit. If the sequence  $u$  is shift periodic, the set  $O(u)$  is finite, so trivially, every point in  $\overline{O(u)} = O(u)$  has a dense orbit. Thus the following result is only of interest when  $u$  is an aperiodic sequence.

**Theorem 18.3.8** *The dynamical system  $(\overline{O(u)}, \sigma)$  is minimal if and only if  $u$  is an almost periodic sequence.*

**Proof.** Suppose that  $u$  is almost periodic, but the dynamical system is not minimal. There exists  $\alpha \in \overline{O(u)}$  for which the set  $\{\sigma^n(\alpha) : n \in \mathbb{N}\}$  is not dense in  $\overline{O(u)}$ . It follows that  $\alpha \notin \overline{O(\alpha)}$ , for otherwise  $\overline{O(u)} \subset \overline{O(\alpha)}$ , and the two sets would be equal.

By the Proposition 18.3.4, there exists  $j$  such that  $u_0 u_1 \dots u_j \notin \mathcal{L}(\alpha)$ . Now there is a sequence  $(k_n)$  for which  $\alpha = \lim_{n \rightarrow \infty} \sigma^{k_n}(u)$ , and using the fact that  $u$  is almost periodic

$$u_0 u_1 \dots u_j = u_{k_n+s} u_{k_n+s+1} \dots u_{k_n+s+j},$$

for some  $s$  and infinitely many  $k_n$ . By the continuity of  $\sigma$ ,  $\sigma^{k_n+s}(u)$  approaches  $\sigma^s(\alpha)$ , so

$$\alpha_s \alpha_{s+1} \dots \alpha_{s+j} = u_{k_n+s} u_{k_n+s+1} \dots u_{k_n+s+j},$$

for all  $n$  large enough.. It follows that  $u_0 u_1 \dots u_j = \alpha_s \alpha_{s+1} \dots \alpha_{s+j}$ , which is a contradiction.

Conversely, suppose  $(\overline{O(u)}, \sigma)$  is minimal, and let  $\alpha \in \overline{O(u)}$ , so  $\overline{O(\alpha)} = \overline{O(u)}$ . Let  $V = B_\epsilon(u)$  be some open ball containing  $u$ , where  $\epsilon > 0$ , then  $V \cap \overline{O(\alpha)} \neq \emptyset$ . In particular,  $\sigma^k(\alpha) \in V$  for some  $k \in \mathbb{N}$  and  $\overline{O(u)} = \bigcup_{k \geq 0} \sigma^{-k}V$ , which is an open cover

for  $\overline{O(u)}$ . By the compactness of  $\overline{O(u)}$  there is a finite subcover, say

$$\overline{O(u)} = \sigma^{-k_1}V \cup \sigma^{-k_2}V \cup \dots \cup \sigma^{-k_n}V.$$

If  $K = \max_{1 \leq i \leq n} k_i$ , then iterations of  $\alpha$  under  $\sigma$  must enter  $V$  after at most  $K$  steps.

If  $\alpha = \sigma^j u$ , then one of the points  $\sigma^j u, \dots, \sigma^{j+K} u$  must lie in  $V$ . If we start with

$$V = \{\omega \in \mathcal{A}^{\mathbb{N}} : \omega_0 = u_0, \omega_1 = u_1, \dots, \omega_\ell = u_\ell\},$$

(the *cylinder set*  $[u_0 u_1 \dots u_\ell]$  - see Exercise 18.1), then for every  $j > 0$ ,  $u_0 u_1 \dots u_\ell$  is one of the words

$$u_j \dots u_{j+\ell}, \quad u_{j+1} \dots u_{j+\ell+1}, \quad u_{j+K} \dots u_{j+K+\ell},$$

and  $u$  is almost periodic. □

**Remark 18.3.9** In Appendix B we indicate that a complete metric space  $X$  having no isolated points must be uncountable. This is a consequence of the Baire Category Theorem. Theorem 17.1.6(d) shows that a compact metric space is complete, and we know that any closed subset of a compact metric space is compact. It follows that sets of the form  $\overline{O(u)}$  in Theorem 18.3.8 are complete, and if the shift map  $\sigma$  is minimal, where  $u$  is not a periodic sequence, then there can be no isolated points. In particular the space  $\overline{O(u)}$  has to be uncountable. For example, if  $u$  is the Morse sequence,  $\overline{O(u)}$  is an uncountable set. This need not be true when  $u$  is merely a recurrent sequence, even if it is aperiodic.

### Exercises 18.3

1. (a) Let  $\sigma : \Sigma \rightarrow \Sigma$  be the shift map on  $\Sigma = \mathcal{A}^{\mathbb{N}}$ , where  $\mathcal{A} = \{0, 1\}$ . Show that  $\sigma$  is continuous.
- (b) Show that the set of all shift periodic sequences (respectively eventually periodic sequences), is a countably infinite dense subset of  $\Sigma$ . Deduce that the set of aperiodic sequences is uncountable.
- (c) Show that the set of all recurrent sequences in  $\Sigma$  is not closed. (Hint: Note that it is a dense subset of  $\Sigma$ , since it contains all the periodic points).

2. Write  $1^n = 11\dots1$ , ( $n$  times). A sequence  $w \in \Sigma$  is defined by

$$w = 0101^201^301^40\dots01^n0\dots = 01011011101110\dots$$

- (a) Note that  $w$  is not periodic, is  $w$  recurrent? Does  $w$  arise from a substitution?
- (b) Show that  $1^\infty = 11111111\dots \in \overline{O(w)}$ , and  $01^\infty \in \overline{O(w)}$ , but  $0^\infty \notin \overline{O(w)}$ . Are there any other limit points of the orbit of  $w$ ?
- (c) Deduce that there are sequences  $u$  for which  $\overline{O(u)}$  is a countably infinite set.

3. Show that if  $u \in \Sigma = \mathcal{A}^{\mathbb{N}}$  is a recurrent sequence on some finite alphabet  $\mathcal{A}$ , then  $u$  is a recurrent point in  $\Sigma$  under the shift map  $\sigma$  in the sense of Exercises 17.4 (i.e., there is a sequence  $n_k \rightarrow \infty$  such that  $\sigma^{n_k}(u) \rightarrow u$  as  $k \rightarrow \infty$ ).

- 4. (a) Show that if  $u$  is periodic under the shift map, then  $\overline{O(u)}$  is finite. Conversely, show that if  $\overline{O(u)}$  is finite, then  $u$  is periodic.
- (b) If  $O(u)$  has finitely many limit points, show that the limit points are periodic or eventually periodic points under the shift map.
- (c) Let  $u \in \Sigma = \mathcal{A}^{\mathbb{N}}$  for some finite alphabet  $\mathcal{A}$ . Show that  $O(u)$  is dense in  $\Sigma$  if and only if every finite block of symbols from  $\mathcal{A}$  appears somewhere in  $u$ .

5. (a) A sequence  $w \in \Sigma$  is defined by taking all finite words on  $\mathcal{A} = \{0, 1\}$  in lexicographic order, so that

$$w = 0100011011000001010011100101\dots$$

Show that  $w$  is recurrent under the shift map, but it is not almost periodic.

- (b) On the other hand, we have seen in Section 6.5 that the sequence  $w$  defined in (a) is a transitive point for the full shift map, i.e.,  $\overline{O(u)} = \Sigma$ . Give a direct argument that shows that the full shift map is not minimal.

6. The substitution  $\theta$  on  $\mathcal{A}$  has constant length  $p$  if  $|\theta(a)| = p$  for all  $a \in \mathcal{A}$ . For example,  $\theta(0) = 010$ ,  $\theta(1) = 121$ ,  $\theta(2) = 202$  on  $\{0, 1, 2\}$  has constant length 3.

Show that for a substitution of constant length  $p$

$$\sigma^p \circ \theta(\omega) = \theta \circ \sigma(\omega), \quad \text{for all } \omega \in \overline{\mathcal{O}(u)},$$

where  $\sigma$  is the shift map and  $u$  is a fixed point of  $\theta$ .

#### 18.4 Substitution Dynamics.

In order for a substitution  $\theta$ , defined on a finite alphabet  $\mathcal{A}$ , to give rise to a non-trivial dynamical system, certain minimum requirements are necessary. These requirements, which we shall usually assume, are as follows:

**Properties 18.4.1** 1. There exists  $a \in \mathcal{A}$  such that  $\theta(a)$  begins with  $a$  (this is necessary in order for  $\theta$  to have a fixed point  $u = \lim_{n \rightarrow \infty} \theta^n(a)$ ).

2.  $\lim_{n \rightarrow \infty} |\theta^n(b)| = \infty$  for all  $b \in \mathcal{A}$ .

3. All letters of  $\mathcal{A}$  actually occur in the fixed point  $u$ .

A related property that is useful for proving minimality is the requirement that the substitution be primitive. This notion was defined in a different way in Section 16.4 using the incidence matrix of the substitution. We shall see in the exercises that the two definitions are equivalent:

**Definition 18.4.2** A substitution  $\theta$  defined on  $\mathcal{A}$  is said to be *primitive* if there exists  $k \in \mathbb{N}$  such that for all  $a, b \in \mathcal{A}$ , the letter  $a$  occurs in the word  $\theta^k(b)$ .

We shall show that primitive substitutions give rise to minimal dynamical systems. It is clear that the Morse and Fibonacci substitutions are primitive, but the *Chacon substitution*:

$$\theta(0) = 0010, \quad \theta(1) = 1,$$

is clearly not primitive, and, does not satisfy (2) above (nor does  $\theta^n$  for any  $n$ ). However, the Chacon substitution can be shown to give rise to a minimal dynamical system which is also weakly-mixing, but not mixing (see Exercises 17.5 and [45]). Note that “non-trivial looking” substitutions such as  $\theta(0) = 010$ ,  $\theta(1) = 10101$  on  $\mathcal{A} = \{0, 1\}$ , may give rise to a shift periodic fixed point, even though they satisfy Properties 18.4.1.

The following proposition, together with the observation that if  $\theta$  is a primitive substitution, then some power  $\theta^n$  will satisfy Properties 18.4.1, can be used to show that certain substitutions are minimal, and so have non-trivial dynamics. Recall that  $u$  is a periodic point of the substitution  $\theta$  if there is some  $p \in \mathbb{N}$  with  $\theta^p(u) = u$ . It

follows from Exercises 18.3 # 4 that if  $u$  is a periodic point of  $\theta$ , then the dynamical system  $(\overline{O(u)}, \sigma)$  is finite (i.e.,  $\overline{O(u)}$  is a finite set), if and only if  $u$  is periodic under the shift map  $\sigma$ .

**Proposition 18.4.3** *If  $\theta$  is a primitive substitution, then any periodic point  $u$  of  $\theta$ , is an almost periodic sequence. It follows that the shift dynamical system  $(\overline{O(u)}, \sigma)$  is minimal.*

**Proof.** Suppose that  $u = \theta^p(u)$  is a periodic point of the substitution  $\theta$ ,  $u = u_0u_1u_2\dots$  say. Let  $k \in \mathbb{N}$  be chosen so that for any  $b \in \mathcal{A}$ ,  $a$  appears in  $(\theta^p)^k(b)$ . Now

$$u = (\theta^p)^k(u) = (\theta^p)^k(u_0)(\theta^p)^k(u_1)\dots,$$

where  $a$  occurs in each  $(\theta^p)^k(u_i)$  ( $a$  occurs infinitely often with bounded gaps). It follows that  $(\theta^p)^n(a)$  appears in  $u = (\theta^p)^n(u)$  infinitely often with bounded gaps, and hence so does every factor occurring in  $u$ . □

It is always true that minimal maps are onto (see Exercises 17.4), but the following shows that this is also true for the shift map when  $u$  is recurrent.

**Proposition 18.4.4** *If  $u$  is a recurrent sequence, then the shift map  $\sigma : \overline{O(u)} \rightarrow \overline{O(u)}$  is onto.*

**Proof.** Let  $\omega = \omega_0\omega_1\dots \in \overline{O(u)}$ . Then since  $\mathcal{L}(\omega) \subseteq \mathcal{L}(u)$  and  $u$  is recurrent,  $\omega_0\omega_1\dots\omega_n$  appears in  $u$  infinitely often. It follows that there exists  $a_n \in \mathcal{A}$  such that

$$a_n\omega_0\omega_1\dots\omega_n$$

appears in  $u$ . For each  $n \in \mathbb{N}$ , we can construct  $\omega^n \in \overline{O(u)}$  of the form  $\omega^n = a_n\omega_0\omega_1\dots\omega_n * * * \dots$ . Using the compactness of  $\overline{O(u)}$ , the infinite sequence  $(\omega^n)_{n \in \mathbb{N}}$  must have a limit point  $\omega' \in \overline{O(u)}$ . We must have  $\omega' = a\omega$  for some  $a \in \mathcal{A}$ , for otherwise  $d(\omega', \omega^n)$  would be bounded below by some fixed  $\epsilon > 0$ . It is now clear that  $\sigma(a\omega) = \omega$ , so that  $\sigma$  is onto. □

One might ask whether the shift map  $\sigma : \overline{O(u)} \rightarrow \overline{O(u)}$  is one-to-one, as this would imply that  $\sigma$  is a homeomorphism (see Exercises 17.2 # 1, which says that a continuous, one-to-one and onto map between compact metric spaces is necessarily a homeomorphism). However, this is not the case. Recall that we stated, without

proof, that the complexity function of the Morse sequence  $u$  satisfies  $p_u(n) \leq Cn$  for some constant  $C > 0$ . For such sequences we have:

**Theorem 18.4.5** *Let  $u$  be a recurrent sequence with  $p_u(n+1) - p_u(n) \leq C$  for all  $n$ , and some constant  $C > 0$ . Then there is a  $\sigma$ -invariant countable set  $D$  such that  $\sigma : \overline{O(u)} \setminus D \rightarrow \overline{O(u)} \setminus D$  is one-to-one.*

**Proof.** We have  $p_u(n+1) - p_u(n) \leq C$  where we may assume  $C$  is an integer. Since  $u$  is recurrent, from the argument used in the previous proposition, every word  $w$  in  $u$  has at least one left extension (i.e., a word  $aw$  occurring in  $u$  for some  $a \in \mathcal{A}$ ). It follows that there can be at most  $C$  words of length  $n$  which have two or more left extensions. Set

$$F = \{v \in \overline{O(u)} : \sigma^{-1}(v) \text{ contains at least two members}\}.$$

We show that  $F$  is a finite set.

If  $v = v_0v_1v_2\dots \in F$ , there exists  $a, b \in \mathcal{A}$ ,  $a \neq b$  such that  $av_0v_1v_2\dots$  and  $bv_0v_1v_2\dots$  belong to  $\overline{O(u)}$ . Clearly these have the same image under  $\sigma$ . It follows that if  $v^{(1)}, \dots, v^{(k)}$  are  $k$  sequences in  $F$ , for  $1 \leq i \leq k$ , each of the words  $v_0^{(i)}v_1^{(i)}\dots v_{n-1}^{(i)}$  in  $\mathcal{L}_n(u)$ , has at least two left extensions to  $\mathcal{L}_{n+1}(u)$ , for every  $n$ . Consequently, we must have  $k \leq C$ . Thus the set  $F$  contains at most  $C$  elements.

Now set  $D = \bigcap_{n=1}^{\infty} \sigma^n F$ , an at most countable set which is invariant under  $\sigma$ . We need to exclude  $F$  and all of its orbits under  $\sigma$ , in order to ensure that  $\sigma$  is well defined on  $\overline{O(u)} \setminus D$ , and is one-to-one.

□

In Remark 18.3.9, we mentioned that if  $u$  is recurrent and not eventually periodic, then the set  $\overline{O(u)}$  is uncountable. It follows that  $\overline{O(u)} \setminus D$  is non-empty.

## Exercises 18.4

1. (a) Let  $\theta$  be a substitution on  $\mathcal{A} = \{a_1, a_2, \dots, a_p\}$ . Why does there exist  $a \in \mathcal{A}$  and  $n \in \mathbb{N}$  for which  $\theta^n(a)$  starts with an  $a$ ?
- (b) Deduce that if  $|\theta^n(a)| \rightarrow \infty$  for all  $a \in \mathcal{A}$ , then  $\theta$  has a periodic point  $u$  ( $\theta^n(u) = u$  for some  $n \in \mathbb{N}$ ).
- (c) Deduce that if  $\theta$  is a primitive substitution, then some power of  $\theta$  satisfies Properties 18.4.1.

2. Let  $\theta(0) = 01$ ,  $\theta(1) = 0$  be the Fibonacci substitution on  $\mathcal{A} = \{0, 1\}$ . Show that if  $u = \theta^\infty(0)$ , then both  $w_1 = 0u$  and  $w_2 = 1u$  belong to  $\overline{\mathcal{O}}(u)$  (so in particular  $\sigma(w_1) = \sigma(w_2)$ , with  $w_1 \neq w_2$ ), and the shift map  $\sigma$  is not one-to-one.
3. Show that the Chacon substitution  $\theta(0) = 0010$ ,  $\theta(1) = 1$  on the alphabet  $\mathcal{A} = \{0, 1\}$ , gives rise to a minimal dynamical system.
4. Let  $M_\sigma$  be the *incidence matrix* of a substitution  $\sigma$  (see Section 16.4). Show
- $M_{\sigma^n} = (M_\sigma)^n$ .
  - $\sigma$  is primitive if and only if there exists  $k \in \mathbb{Z}^+$  such  $M_\sigma^k$  has only positive constant entries.
  - Deduce that the Fibonacci and tribonacci substitutions are primitive.
- 5\*. (a) Show that the Toeplitz dynamical system is a factor of the Morse dynamical system. (Hint: If  $(\overline{\mathcal{O}(u)}, \sigma)$  and  $(\overline{\mathcal{O}(v)}, \sigma)$  denote the respective dynamical systems, where  $\sigma$  is the shift map, define a map  $b : \{00, 01, 10, 11\} \rightarrow \{0, 1\}$  by  $b(ab) = a + b \pmod{2}$ . Extend  $b$  as a map  $b : \mathcal{A}^\mathbb{N} \rightarrow \mathcal{A}^\mathbb{N}$  in the obvious way and show that  $b(u) = v$ . Show that the restriction  $b : \overline{\mathcal{O}(u)} \rightarrow \overline{\mathcal{O}(v)}$  is well defined, continuous and onto, and in fact is a factor map).
- (b) Use the method of (a) to show that the Toeplitz dynamical system is a factor of the Rudin-Shapiro dynamical system.

6. **Block Codes.** Let  $\mathcal{A}$  be a finite alphabet,  $p \geq 1$  and  $F : \mathcal{A}^p \rightarrow \mathcal{A}$ . Denote by  $\Sigma = \mathcal{A}^\mathbb{N}$  the shift space with usual metric. Define a map  $f : \Sigma \rightarrow \Sigma$  by

$$[f(w)]_i = F(x_i, x_{i+1}, \dots, x_{i+p-1}), \text{ when } w = (x_0, x_1, x_2, \dots).$$

$f$  is called a *block code*. For example, the maps  $[f(w)]_i = x_{i+1}$  and  $b$  from the previous exercise, define block codes.

- Show that  $f$  is a continuous map that commutes with the shift map  $\sigma : \Sigma \rightarrow \Sigma$  (i.e.,  $f \circ \sigma = \sigma \circ f$ ).

(b)\* Prove the Curtis-Hedlund-Lyndon Theorem, which says that every code is a block code [20]. To be more specific, show that if  $\mathcal{A}$  and  $\mathcal{B}$  are two finite alphabets with corresponding shift spaces  $\Sigma$  and  $\Sigma'$ , and  $f : \Sigma \rightarrow \Sigma'$  is a continuous map commuting with the respective shifts, then there is a map  $F : \mathcal{A}^p \rightarrow \mathcal{B}$  for some  $p \geq 1$  with  $[f(w)]_i = F(x_i, x_{i+1}, x_{i+2}, \dots, x_{i+p-1})$ . (Hint: The inverse image under  $f$  of the set  $\{w : x_0 = a\}$  is both open, closed and compact. Deduce that it is the finite union of cylinder sets. This holds for each  $a \in \mathcal{B}$ , a finite set, so we may choose  $p$  sufficiently large that for every  $a$ , the inverse image of  $\{w : x_0 = a\}$  is a finite union of cylinders defined on coordinates  $0, 1, \dots, p-1$ . Now set  $F(x_0, x_1, \dots, x_{p-1}) = a$  if the cylinder set  $[x_0, x_1, \dots, x_{p-1}]$  (see Exercises 18.1 # 2), is contained in the inverse image of  $\{w : x_0 = a\}$  and use the fact that  $f$  commutes with the shift).

(c) The map  $b : \overline{O(u)} \rightarrow \overline{O(v)}$  in 4(a) is called a *2-block code*. Generally, if  $u$  and  $v$  are sequences taking values in finite alphabets  $\mathcal{A}$  and  $\mathcal{B}$ , and  $b : \overline{O(u)} \rightarrow \overline{O(v)}$  is a continuous map such that for every  $w \in \overline{O(u)}$ ,  $[b(w)]_i$  depends only on  $(x_i, x_{i+1}, \dots, x_{i+n-1})$ , then  $b$  is an *n-block code*. Show that if  $\phi : \overline{O(u)} \rightarrow \overline{O(v)}$  is a factor map, then  $\phi$  is an *n-block code*, for some  $n \in \mathbb{N}$ .

7. A substitution  $\theta$  on  $\mathcal{A}$  is *one-to-one* (or *injective*), if  $\theta(a) = \theta(b)$ , implies  $a = b$ , for any  $a, b \in \mathcal{A}$ . For example, the substitution  $\theta(0) = 01$ ,  $\theta(1) = 01$  is not one-to-one on  $\{0, 1\}$ , but  $\theta(0) = 01$ ,  $\theta(1) = 1$  is one-to-one. The purpose of this exercise is to show how a substitution that is not one-to-one can be reduced to one that is, by eliminating superfluous members of the alphabet  $\mathcal{A}$ .

Suppose  $\theta$  is not one-to-one on  $\mathcal{A}$ . Let  $\mathcal{B} \subseteq \mathcal{A}$  be defined so that for all  $a \in \mathcal{A}$ , there exists a unique  $b \in \mathcal{B}$  with  $\theta(a) = \theta(b)$ . Let  $\phi : \mathcal{A} \rightarrow \mathcal{B}$  be the map defined by  $\phi(a) = b$  if  $\theta(a) = \theta(b)$ .

Denote by  $\tau$  the unique substitution on  $\mathcal{B}$  satisfying

$$\tau \circ \phi = \phi \circ \theta,$$

(where  $\phi$  is extended in the obvious way).

(a) Find  $\tau$  for the substitutions (i)  $\theta(0) = 01$ ,  $\theta(1) = 01$ ,  $\theta(2) = 20$ , (ii)  $\theta(0) = 021$ ,  $\theta(1) = 12$ ,  $\theta(2) = 021$ . (Although the substitution  $\tau$  may not be one-to-one, by continuing this reduction, a one-to-one substitution is obtained in a finite number of steps).

(b) If  $\theta$  is primitive, show that  $\tau$  is also primitive.

- (c) If  $X_\theta = \overline{O(u)}$ , where  $u$  is a fixed point of  $\theta$ , show that  $\phi(u)$  is a fixed point of  $\tau$  and the dynamical system  $(X_\tau, \sigma)$  is a factor of the dynamical system  $(X_\theta, \sigma)$  (where  $\sigma$  is the shift map, and where  $\phi : X_\theta \rightarrow X_\tau$  is extended as before).
- (d) Show that  $X_\theta$  is a finite set if and only if  $X_\tau$  is a finite set.
- (e) If  $X_\theta$  is not finite, show that  $(X_\tau, \sigma)$  and  $(X_\theta, \sigma)$  are conjugate dynamical systems.

## CHAPTER 19

### Sturmian Sequences and Irrational Rotations.

Let  $u = u_0u_1u_2\dots$  be an infinite sequence. Recall that  $p_u(n) = |\mathcal{L}_n(u)|$  is the number of different factors of length  $n$  appearing in  $u$ , and is called the *complexity function* of  $u$ . In this chapter we investigate sequences whose complexity function satisfies  $p_u(n) = n + 1$  for all  $n \in \mathbb{N}$ . These are the aperiodic sequences of minimal complexity, called *Sturmian sequences*. (We saw in Section 18.2 that if there exists  $n \in \mathbb{N}$  with  $p_u(n) \leq n$ , then  $u$  is an ultimately periodic sequence). We start by showing that the sequence  $u$  arising as the fixed point of the Fibonacci substitution,  $\theta(0) = 01$ ,  $\theta(1) = 0$ , is a Sturmian sequence, and so too are sequences arising from a “coding” of irrational rotations. We end the chapter with the celebrated *Three Distance Theorem*, conjectured by Steinhaus, and proved by V. T. Vos. The proof given uses dynamical methods related to Sturmian sequences, and is due to V. Berthé [13]. Roughly speaking, the Three Distance Theorem says that iterations of an irrational rotation  $T_\alpha$ , partition  $[0, 1)$  into intervals of only three possible different lengths.

#### 19.1 Sturmian Sequences.

**Definition 19.1.1** A sequence  $u$  having the property that the complexity function  $p_u(n) = n + 1$  for all  $n \in \mathbb{N}$  is said to be a *Sturmian sequence*.

Since  $p_u(1) = 2$ , the sequence must use only two letters, so we write the alphabet as  $\mathcal{A} = \{0, 1\}$ . We will assume that  $\mathcal{A} = \{0, 1\}$  is our alphabet throughout this chapter. In addition,  $p_u(2) = 3$ , so one of the pairs  $00$ ,  $11$  does not appear in  $u$  ( $01$  and  $10$  have to appear for otherwise the sequence would be constant).

If  $u$  is a Sturmian sequence, then it has to be aperiodic (neither periodic, nor ultimately periodic), for otherwise  $p_u(n)$  would be bounded. In addition,  $u$  has to be recurrent; for suppose the factor  $w = w_1\dots w_n$  only occurs a finite number of times in  $u = u_1u_2\dots$ , and does not occur after  $u_N$ . Let

$$v = u_{N+1}u_{N+2}\dots,$$

a new sequence whose language  $\mathcal{L}(v)$  does not contain  $w$ . It follows that  $p_v(n) \leq n$ , so that  $v$  is eventually periodic. This implies  $u$  is eventually periodic, a contradiction.

Sturmian sequences have a long history with contributions from Jean Bernoulli III, Christoffel, A. A. Markov, M. Morse, G. Hedlund, E. Coven, M. Keane, and others. Our treatment is mostly based on that in Allouche and Shallit [2], Fogg [45], Rauzy [107], and Lothaire [85]. Our aim is to show that sequences arising as “codings” of irrational rotations are Sturmian, and some (but not all), Sturmian sequences can be represented as substitutions. We mention, without proof, that all Sturmian sequences arise as codings of irrational rotations. We start by showing that Sturmian sequences do exist, and in fact, the Fibonacci sequence is Sturmian.

**Definition 19.1.2** Let  $u \in \mathcal{A}^{\mathbb{N}}$ , where  $\mathcal{A} = \{0, 1\}$ . A *right special factor* of  $u$  is a factor  $w$  of  $u$ , that appears in  $u$  such that  $w0$  and  $w1$  also appear in  $u$ . *Left special factors* are defined in a similar way. A *right extension* of a factor  $w$  of  $u$ , is a word of the form  $wx$ ,  $x \in \mathcal{A}$ , where  $wx$  is also a factor of  $u$ . Left extensions are defined similarly. Thus right special factors are words that appear in  $u$  having two distinct right extensions.

**Proposition 19.1.3** Let  $\mathcal{A} = \{0, 1\}$  and  $u \in \mathcal{A}^{\mathbb{N}}$ . The sequence  $u$  is Sturmian if and only if it has exactly one right special factor of each length.

**Proof.** If  $u$  is Sturmian and  $w_1, w_2, \dots, w_{n+1}$  are the distinct factors of length  $n$ , then all but one factor can be extended in a unique way to form a factor of length  $n + 1$ , and exactly one factor must be extendable in two ways.

Conversely, suppose that for each  $n \in \mathbb{N}$ , there is exactly one right special factor. When  $n = 1$ ,  $p_u(1) = 2$ , since the set of words is just  $\{0, 1\}$ . Using induction, suppose that  $p_u(n) = n + 1$ , i.e., there are  $n + 1$  different factors of length  $n$  in the sequence  $u$ . Then there are  $n + 2$  factors of length  $n + 1$ , since all the factors of length  $n$  have one right extension except for one, which has two.

□

**Example 19.1.4** The Fibonacci substitution is defined on  $\mathcal{A} = \{0, 1\}$  by  $\theta(0) = 01$ ,  $\theta(1) = 0$ . We have seen that  $\theta$  has a unique fixed point:

$$u = 0100101001001 \dots$$

The sequence  $u$  is a Sturmian, i.e.,  $p_u(n) = n + 1$  for all  $n \geq 0$ .

**Proof.** We will show that for each  $n$  there is exactly one right special factor of length  $n$ .

Recall that  $\epsilon$  is the empty word. Now  $u = \theta(u)$  is a concatenation of the words 0 and 01, so that the word 11 does not appear as a factor of  $u$ , and  $p_u(2) = 3$ . This gives

$$\mathcal{L}_0(u) = \{\epsilon\}, \quad \mathcal{L}_1(u) = \{0, 1\}, \quad \mathcal{L}_2(u) = \{00, 01, 10\}.$$

We now show that

$$\mathcal{L}_3(u) = \{001, 010, 100, 101\}.$$

Since 11 cannot appear in  $u$ , the words 011, 110, and 111 do not appear. The factor 00 can only result from  $\theta(1)\theta(0) = \theta(10)$ , but then 000 must result from  $\theta(1)\theta(1)\theta(0) = \theta(110)$ . This is impossible as 11 does not appear.

Note that for all factors  $w$  of  $u$ , either  $0w0$  or  $1w1$  is not a factor of  $u$ . This is clear if  $w = \epsilon$ , the empty word, and also if  $w = 0$ , or  $w = 1$ . We use induction on the length to prove this generally.

Consequently, assume that both  $0w0$  and  $1w1$  are factors of  $u$ . It follows that  $w$  must start and end with a 0 since 11 cannot appear:  $w = 0y0$  for some factor  $y$ , so that  $00y00$  and  $10y01$  are factors of  $u$  (one can continue this analysis to see that  $y$  cannot start or end with a “0”). Since these are factors of  $\theta(u)$ , on examining the possibilities, we see that there exists a factor  $z$  of  $u$  with  $\theta(z) = 0y$ . In other words, the only way these words can occur, is as

$$\theta(1z10) = 00y001 \quad \text{and} \quad \theta(0z0) = 010y01,$$

so that  $1z1$  and  $0z0$  are both factors of  $u$ . Since the length of  $z$  is less than the length of  $w$ , we have a contradiction of our assumption that there can be at most one factor of this type having a lesser length.

We now see that  $u$  has at most one right special factor of each length, because if  $w$  and  $v$  are right special factors of the same length, let  $x$  be the longest sub-word of both  $w$  and  $v$  (possibly  $x = \epsilon$ ), with  $w = zx$ ,  $v = yx$  for some words  $y$  and  $z$ . Then the words  $0x0$ ,  $0x1$ ,  $1x0$  and  $1x1$  are all factors of  $u$  (because the last letters of  $y$  and  $z$  must be different), contradicting the last observation.

Finally, we show that there is at least one right special factor of each length. Set  $f_n = \theta^n(0)$ , so  $f_0 = 0$ . Let  $f_{-1} = 1$ . We use the identity

$$f_{n+2} = v_n \tilde{f}_n \tilde{f}_n t_n, \quad n \geq 2,$$

where  $v_2 = \epsilon$ , and for  $n \geq 3$

$$v_n = f_{n+3} \cdots f_1 f_0, \quad \text{and} \quad t_n = \begin{cases} 01; & n \text{ odd} \\ 10; & \text{otherwise} \end{cases},$$

(where  $\tilde{w}$  denotes the reflection of the word  $w$ , so if  $w = a_0 a_1 \dots a_n$ , then  $\tilde{w} = a_n a_{n-1} \dots a_1$ ). The identity can be proved by induction (see Exercises 19.1).

We see that the word  $\tilde{f}_n$  appears in two different places, and since the first letter of  $\tilde{f}_n$  is different to the first letter of  $t_n$ ,  $\tilde{f}_n$  is a right special factor (it is easy to see that the last letter of  $f_n$  alternates - 0 when  $n$  is even, 1 when  $n$  is odd). Now any word of the form  $w\tilde{f}_n$  is clearly, also a right special factor of  $u$ , so there are right special factors of any length, and the result follows.

□

### Exercises 19.1

1. Show that the Morse sequence is not Sturmian. (Hint: Look at the number of factors of length 2).
2. Let  $u$  be the fixed point of the substitution  $\theta(0) = 01$  and  $\theta(1) = 010$ , over  $\mathcal{A} = \{0, 1\}$ . Write down the sets  $\mathcal{L}_i(u)$ , for  $i = 0, 1, 2, 3$ , and show that  $p_n(i) = |\mathcal{L}_i(u)| = i + 1$  in each case.
3. Show that Sturmian sequences are almost periodic, and hence give rise to minimal dynamical systems.
4. Let  $f_n = \theta^n(0)$ ,  $n \geq 0$  and  $f_{-1} = 1$ , where  $u = u_0 u_1 \dots$  is the Fibonacci sequence generated by the substitution  $\theta(0) = 01$ ,  $\theta(1) = 0$ .
  - (a) Show that  $f_4 = \epsilon(010)(010)10$ , and  $f_5 = 0(10010)(10010)01$ , and use induction on the length of a word to show that for any word  $w$

$$\theta(\tilde{w})0 = \widetilde{\theta(w)}.$$

(where  $\tilde{w}$  denotes the reflection of the word  $w$ , so if  $w = a_0 a_1 \dots a_n$ , then  $\tilde{w} = a_n a_{n-1} \dots a_1$ ).

(b) Use (a) to prove the identity

$$f_{n+2} = v_n \tilde{f}_n \tilde{f}_n t_n, \quad n \geq 2,$$

where  $v_2 = \epsilon$ , and for  $n \geq 3$

$$v_n = f_{n+3} \cdots f_1 f_0, \quad \text{and} \quad t_n = \begin{cases} 01; & n \text{ odd} \\ 10; & \text{otherwise} \end{cases}.$$

(Hint: Show that  $\theta(\tilde{f}_n t_n) = 0 \tilde{f}_{n+1} t_{n+1}$ , then use  $\theta(v_n) 0 = v_{n+1}$ ).

5. The topological entropy of an infinite sequence  $u = u_0 u_1 \dots$ , is defined as:

$$H(u) = \lim_{n \rightarrow \infty} \frac{\log_d p_u(n)}{n},$$

where  $d = |\mathcal{A}|$  and  $p_u(n)$  is the complexity function of  $u$ .

(i) Find  $H(u)$  for  $u$  a Sturmian sequence.

(ii) Find  $H(u)$  for a sequence whose complexity function is  $p_u(n) = 2^n$ .

(iii)\* Show that  $\lim_{n \rightarrow \infty} \log_d p_u(n)/n$  exists. (Hint: If  $(a_n)$  is a non-negative sequence that satisfies  $a_{m+n} \leq a_n + a_m$ , ( $m, n \geq 1$ ), show that  $\lim_{n \rightarrow \infty} a_n/n$  exists. Now use the result of Exercises 18.2 # 3).

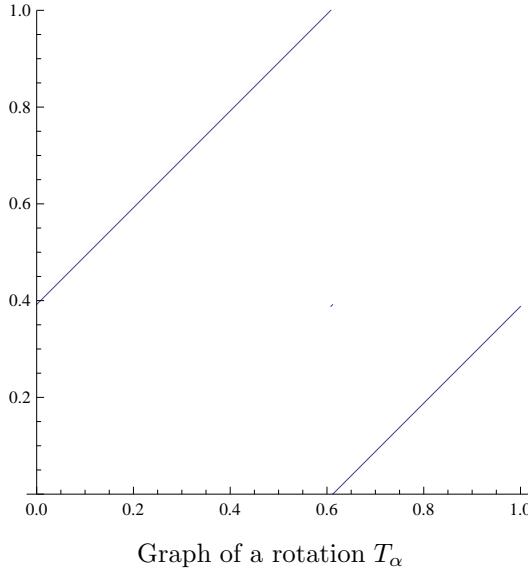
## 19.2 Sequences Arising From Irrational Rotations.

If  $x \in \mathbb{R}$ , we denote the *integer part* of  $x$  by  $\lfloor x \rfloor$  and its *fractional part* by  $\{x\}$ :

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\} \quad \text{and} \quad \{x\} = x - \lfloor x \rfloor.$$

$f(x) = \lfloor x \rfloor$  is called the *floor* function. With this notation, if  $0 < \alpha < 1$ , the rotation  $T_\alpha : [0, 1) \rightarrow [0, 1)$  is defined by  $T_\alpha x = \{x + \alpha\}$ . We can give an explicit description of  $T_\alpha$  as:

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



Define a sequence  $f_\alpha = f_\alpha(1)f_\alpha(2)\dots$  in  $\{0, 1\}^{\mathbb{Z}^+}$  by

$$f_\alpha(n) = \begin{cases} 1; & \{n\alpha\} \in [1 - \alpha, 1) \\ 0; & \text{otherwise,} \end{cases}$$

In this section, it is convenient to start our sequences at  $n = 1$ .

The sequence  $f_\alpha$  is obtained from following the orbit  $\{T_\alpha^n(0) : n \geq 1\}$  of 0 under the rotation  $T_\alpha$ , giving the value 1 if the orbit falls into the interval  $[1 - \alpha, 1)$ , and 0 otherwise. Our aim is to show that such sequences are Sturmian, and in some cases may be represented by a substitution (see Allouche and Shallit [2]). We look at a result due to Brown [23], giving a characterization of those irrationals  $\alpha$ , for which the sequence  $f_\alpha$  can be represented as the fixed point of a substitution.

From Exercises 19.2, we see that if  $0 < \alpha < 1$ , then

$$f_\alpha(n) = \lfloor (n + 1)\alpha \rfloor - \lfloor n\alpha \rfloor.$$

**Example 19.2.1** Let  $\alpha = (\sqrt{5} - 1)/2 = .61803\dots$ . We will show that

$$f_\alpha = 1011010110\dots,$$

which is just the Fibonacci sequence  $u = 0100101001\dots$ , with 0 and 1 interchanged. Thus  $f_\alpha$  is the fixed point of the substitution  $\theta(0) = 1$ ,  $\theta(1) = 10$ .

**Proposition 19.2.2** (a) For  $n \geq 1$  and  $0 < \alpha < 1$ ,

$$\sum_{i=1}^n f_\alpha(i) = \lfloor (n+1)\alpha \rfloor.$$

(b) If  $\alpha$  is irrational, then  $f_{1-\alpha} = R(f_\alpha)$  (where  $R(0) = 1, R(1) = 0$ ).

(c) Set  $g_\alpha = g_\alpha(1)g_\alpha(2)\dots$ , where

$$g_\alpha(n) = \begin{cases} 1; & \text{if } n = \lfloor k\alpha \rfloor \text{ for some integer } k \\ 0; & \text{otherwise,} \end{cases}$$

then if  $\alpha > 1$  is irrational,  $g_\alpha = f_{1/\alpha}$ .

**Proof.** (a) The series is telescoping since

$$\sum_{i=1}^n f_\alpha(i) = (\lfloor 2\alpha \rfloor - \lfloor \alpha \rfloor) + (\lfloor 3\alpha \rfloor - \lfloor 2\alpha \rfloor) + \dots + (\lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor) = \lfloor (n+1)\alpha \rfloor.$$

(b)  $f_{1-\alpha}(n) = \lfloor (n+1)(1-\alpha) \rfloor - \lfloor n(1-\alpha) \rfloor = \lfloor -(n+1)\alpha \rfloor - \lfloor -n\alpha \rfloor + 1$ , so that

$$f_\alpha(n) + f_{1-\alpha}(n) = \lfloor x \rfloor + \lfloor -x \rfloor - \lfloor y \rfloor - \lfloor -y \rfloor + 1$$

where  $x = (n+1)\alpha$  and  $y = n\alpha$ . It is easy to see that  $\lfloor x \rfloor + \lfloor -x \rfloor = -1$  if  $x$  is not an integer (otherwise it is 0), so  $f_\alpha(n) + f_{1-\alpha}(n) = 1$ , and the result follows.

(c)

$$\begin{aligned} g_\alpha(n) = 1 &\iff \exists k \text{ such that } n = \lfloor k\alpha \rfloor \\ &\iff \exists k \text{ such that } n \leq k\alpha < n+1 \\ &\iff \exists k \text{ such that } \frac{n}{\alpha} \leq k < \frac{n+1}{\alpha} \\ &\iff \exists k \text{ such that } \left\lfloor \frac{n}{\alpha} \right\rfloor = k-1 \text{ and } \left\lfloor \frac{n+1}{\alpha} \right\rfloor = k \\ &\iff \left\lfloor \frac{n+1}{\alpha} \right\rfloor - \left\lfloor \frac{n}{\alpha} \right\rfloor = 1 \iff f_{1/\alpha}(n) = 1. \end{aligned}$$

□

Given an alphabet  $\mathcal{A}$ , if  $a \in \mathcal{A}$  and  $n \in \mathbb{N}$ , we write  $a^n = a a \dots a$ , for the word of length  $n$  consisting of  $a$  repeated  $n$  times ( $a^0 = \epsilon$ , the empty word).

**Definition 19.2.3** Let  $k \geq 1$ , and denote by  $\theta_k$  the substitution defined on  $\mathcal{A} = \{0, 1\}$  by

$$\theta_k(0) = 0^{k-1} 1; \quad \theta_k(1) = 0^{k-1} 1 0.$$

When  $k = 1$ , the substitution is  $\theta_1(0) = 1$ ,  $\theta_1(1) = 1 0$ , essentially the Fibonacci substitution. When  $k = 2$ , we have  $\theta_2(0) = 0 1$ ,  $\theta_2(1) = 0 1 0$ .

**Definition 19.2.4 (Continued Fractions.)** It is usual to denote the *continued fraction*

$$a_0 + 1/(a_1 + 1/(a_2 + \dots)) = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots}},$$

by  $[a_0, a_1, a_2, \dots]$ , where  $a_0, a_1, \dots$  are positive integers. If  $\alpha = [0, \overline{a_1, a_2, \dots, a_n}]$ , we mean that the sequence  $a_1, a_2, \dots, a_n$  is repeated periodically. For example, if

$$\alpha = [0, 1, 1, \dots] = \frac{1}{1 + \frac{1}{1 + \dots}} = \frac{1}{1 + \alpha},$$

then  $\alpha$  is a root of the equation  $\alpha^2 + \alpha - 1 = 0$ , and it can be seen that  $\alpha = \frac{\sqrt{5}-1}{2}$ . Similarly, if  $\beta = [0, \overline{1, 2}] = 1/(1 + 1/(2 + \beta))$ , and  $\beta = \sqrt{3} - 1$ . These are examples of *quadratic irrationals*, since they are irrational roots of quadratic equations.

Given two substitutions of the form  $\theta_k, \theta_\ell$ , we compose them as functions to form a new substitution  $\theta_k \circ \theta_\ell$ . Using this we show that the sequence  $f_\alpha$  arises as the fixed points of a substitution, when  $\alpha$  is a certain type of quadratic irrational.

**Theorem 19.2.5** Let  $\alpha$  be irrational with  $\alpha = [0, \overline{a_1, a_2, \dots, a_n}]$ . Then  $f_\alpha$  is a fixed point of the substitution

$$\theta_{a_1} \circ \theta_{a_2} \circ \dots \circ \theta_{a_n}.$$

In particular, if  $\alpha = (\sqrt{5} - 1)/2$ , then  $f_\alpha$  is a fixed point of the substitution  $\theta(0) = 1$ ,  $\theta(1) = 1 0$ .

In order to prove this theorem, we shall use Lemma 19.2.6 below. A special case of this lemma tells us that if  $\alpha = (\sqrt{5} - 1)/2$ , then  $\theta_1(f_\alpha) = f_{1/(1+\alpha)} = f_\alpha$  (since  $\alpha = 1/(1 + \alpha)$ ), so that  $f_\alpha$  is a fixed point of the Fibonacci substitution:

**Lemma 19.2.6** Let  $k \geq 1$  and  $\alpha \in (0, 1)$  be irrational. Then

$$\theta_k(f_\alpha) = g_{k+\alpha} = f_{1/(k+\alpha)}.$$

**Proof.** We have  $f_\alpha = f_\alpha(1)f_\alpha(2)\dots f_\alpha(j)\dots$ , where  $f_\alpha(j) = \lfloor (j+1)\alpha \rfloor - \lfloor j\alpha \rfloor$ ,  $j \geq 1$ .

It follows that the sequence  $\theta_k(f_\alpha)$  is a sequence of 0's and 1's with

$$\theta_k(f_\alpha) = D_1 D_2 \dots D_q D_{q+1} \dots,$$

where  $D_j = \theta_k(f_\alpha(j))$ , for  $j \geq 1$ .

Because of the way the substitution  $\theta_k$  is defined, each block  $D_j$  contains exactly one “1”, in the  $k$ th position. The length of  $D_j$ ,  $|D_j|$  is either  $k$  or  $k+1$ .

Suppose that the  $(q+1)$ st “1” appears in position  $n$ , then 1 appears in the block  $D_{q+1}$ , so that

$$n = |D_1 D_2 \dots D_q| + k.$$

Since

$$|D_j| = k \quad \text{if } f_\alpha(j) = 0 \quad \text{and} \quad |D_j| = k+1 \quad \text{if } f_\alpha(j) = 1,$$

(by definition of  $\theta_k$ ), it follows that

$$|D_1 D_2 \dots D_q| = qk + f_\alpha(1) + f_\alpha(2) + \dots + f_\alpha(q) = qk + \lfloor (q+1)\alpha \rfloor,$$

(using Proposition 19.2.2 (a)).

If  $n$  is the position of the  $(q+1)$ st “1” in the sequence  $\theta_k(f_\alpha)$ , then

$$n = qk + \lfloor (q+1)\alpha \rfloor + k = \lfloor (q+1)(k+\alpha) \rfloor.$$

This shows that

$$\theta(f_\alpha)(n) = 1 \iff n = \lfloor (q+1)(k+\alpha) \rfloor, \quad \text{for some } q \geq 0 \iff g_{k+\alpha}(n) = 1.$$

We have therefore shown that  $\theta_k(f_\alpha) = g_{k+\alpha}$ , and Proposition 19.2.2 (c) gives  $\theta_k(f_\alpha) = g_{k+\alpha} = f_{1/(k+\alpha)}$ . □

**Proof of Theorem 19.2.5.** From Lemma 19.2.6, we have

$$\theta_{a_n}(f_\alpha) = f_{1/(a_n+\alpha)} = f_{[0,a_n+\alpha]},$$

$$\theta_{a_{n-1}} \circ \theta_{a_n}(f_\alpha) = \theta_{a_{n-1}}(f_{1/(a_n+\alpha)}) = f_{\frac{1}{a_{n-1}+1/(a_n+\alpha)}} = f_{1/(a_{n-1}+[0,a_n+\alpha])} = f_{[0,a_{n-1},a_n+\alpha]}.$$

Continuing in this way

$$\theta_{a_1} \circ \theta_{a_2} \circ \dots \circ \theta_{a_m}(f_\alpha) = f_\beta,$$

where  $\beta = [0, a_1, a_2, \dots, a_m + \alpha] = \alpha$ . □

## Exercises 19.2

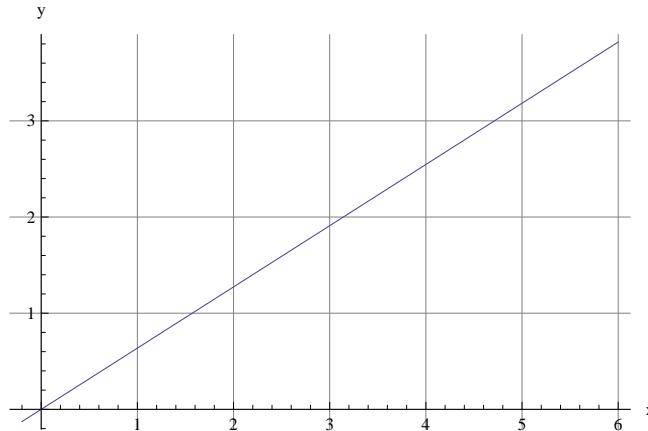
1. Prove that  $\sqrt{2} = 1 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \dots}}}$ .
2. Let  $T_\alpha$  be a rotation on  $[0, 1)$ ,  $0 < \alpha < 1$ . Show that  $T_\alpha^n(0) \in [1 - \alpha, 1)$  if and only if  $\lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor = 1$ , i.e.,  $\{n\alpha\} \in [1 - \alpha, 1)$  if and only if  $f_\alpha(n) = 1$ .
3. Find  $\alpha$  corresponding to the substitution  $\theta_k$  defined in 19.2.3. Do the same for the substitution  $\theta(0) = 010$ ,  $\theta(1) = 01001$ . (Hint: Use Theorem 19.2.5).

### 19.3 Cutting Sequences.

The sequence  $f_\alpha$  has a geometric interpretation. Let  $\beta > 0$  be irrational. Then the line  $\mathcal{L} : y = \beta x$  in the  $xy$ -plane passes through the origin and has slope  $\beta$ . Subdivide the first quadrant in the plane into squares using the *grid-lines*  $x = n$ ,  $y = m$  for  $n, m = 0, 1, 2, \dots$ . Since  $\beta$  is irrational, the line  $\mathcal{L}$  cannot intersect any point of the form  $(n, m)$  for  $n, m \in \mathbb{N}$ , for otherwise we would have  $\beta = m/n$ , a rational. We examine what happens when the line  $\mathcal{L}$  intersects the grid-lines. Consider a point on the line  $\mathcal{L}$  moving away from the origin and traveling upwards in the positive  $x$ -direction. Define an infinite sequence  $S_\beta = S_\beta(0)S_\beta(1)S_\beta(2)\dots$ , by

$$S_\beta(n) = \begin{cases} 0; & \mathcal{L} \text{ intersects a vertical grid-line,} \\ 1; & \mathcal{L} \text{ intersects a horizontal grid-line,} \end{cases}$$

where  $S_\beta(0) = 0$  (since the line starts at the origin).



The cutting sequence is determined by the point of intersection of the line  $y = \beta x$ , and the grid lines.

**Definition 19.3.1** The sequence  $S_\beta$  is called a *cutting sequence* for  $\beta$ .

We will now show that these sequences are of the form  $f_\alpha$  for some  $\alpha$ , and  $f_\alpha$  is actually a Sturmian sequence. In fact, all Sturmian sequences arise from an  $f_\alpha$  for some irrational  $0 < \alpha < 1$ . Since reflection in  $y = x$  interchanges the lines  $y = \beta x$  and  $y = (1/\beta)x$ , and reflection also interchanges the horizontal and vertical grid-lines, it can be seen that

$$R(S_\beta) = S_{1/\beta}, \quad \beta > 0,$$

where  $R$  is the reflection. Consequently, if  $\beta > 1$ , we can obtain the sequence  $S_\beta$  from that of  $S_{1/\beta}$ . So, without any loss of generality, we may assume  $0 < \beta < 1$ . In this case the sequence 11 cannot occur.

The connection between  $f_\alpha$  and  $S_\beta$  is given by Theorem 19.3.2, due to Crisp, Moran, Pollington and Shiue [31], who also completely characterized those  $\alpha$ 's for which  $f_\alpha$  is a substitution sequence:

We ask the question: given a sequence  $f_\alpha$ , what does the corresponding sequence  $S_\alpha$  look like? Suppose for example that  $f_\alpha = 01001010\dots$  (the Fibonacci sequence). This sequence has the property that when  $f_\alpha(n) = 0$ , we have  $\lfloor (n+1)\alpha \rfloor = \lfloor n\alpha \rfloor$ , so that consecutive 0's appear in the sequence  $S_\alpha$ .

If  $f_\alpha(n) = 1$ , then  $\lfloor (n+1)\alpha \rfloor > \lfloor n\alpha \rfloor$ , and we see that if

$$f_\alpha = 01001010\dots, \quad \text{then } S_\alpha = 00100010010\dots,$$

which can be written as

$$S_\alpha = h(f_\alpha), \quad \text{where } h(0) = 0, \quad h(1) = 01,$$

(the extra “0” at the start of  $S_\alpha$  comes from the fact that the sequence starts at the origin). This leads to:

**Theorem 19.3.2** *Let  $0 < \alpha < 1$  be irrational. Then*

$$S_\alpha = f_{\frac{\alpha}{1+\alpha}}.$$

**Proof.** Consider the segment of the line  $\mathcal{L}$  from  $P_n$  to  $P_{n+1}$ , where  $P_n = (n, n\alpha)$  (where we include the point  $P_{n+1}$ , but exclude  $P_n$ ). The block in  $S_\alpha$  corresponding to  $P_n P_{n+1}$  is  $1^i 0$  where

$$i = \lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor = f_\alpha(n).$$

This block is 0 if  $f_\alpha(n) = 0$  and 10 if  $f_\alpha(n) = 1$ . If  $h'(0) = 0, h'(1) = 10$ , then  $h'(f_\alpha) = S_\alpha$  gives the cutting sequence of  $y = \alpha x$  except for the block due to the line segment  $L_0$ . The missing block is 0. Applying instead the substitution  $h$ :  $h(0) = 0, h(1) = 01$  gives the entire cutting sequence, i.e.,  $h(f_\alpha) = S_\alpha$ .

Also, recall that  $\theta_1(0) = 1$  and  $\theta_1(1) = 10$ . Clearly  $h = R \circ \theta_1$ , where  $R(0) = 1$  and  $R(1) = 0$ , so that

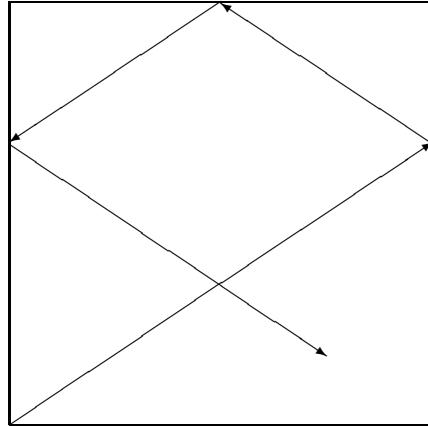
$$S_\alpha = h(f_\alpha) = R \circ \theta_1(f_\alpha) = R(f_{\frac{1}{1+\alpha}}),$$

(using Lemma 19.2.6.). This implies

$$S_{\frac{1}{\alpha}} = R(S_\alpha) = f_{\frac{1}{1+\alpha}}, \quad \text{or} \quad S_\alpha = f_{\frac{\alpha}{1+\alpha}}.$$

□

**19.3.3 Billiards in a Square.** Imagine shooting a ball from the origin in the  $xy$ -plane, at an irrational slope  $\alpha > 0$ , inside the unit square  $[0, 1] \times [0, 1]$ . We assume we have the usual laws of elastic reflection when the ball strikes the sides of the square (angle of incidence equals the angle of reflection). If we write down 0 when the ball strikes a vertical side and 1 when it strikes a horizontal side, we get a sequence which is Sturmian. This can be seen by extending the trajectory of the ball in a straight line from the origin, giving rise to a cutting sequence which is the same as the billiard sequence.



A billiard moving inside the square.

### Exercises 19.3

1. Show that if  $\beta$  and  $\gamma$  are irrational with cutting sequences  $S_\beta$  and  $S_\gamma$  respectively, then  $S_\beta = S_\gamma$  if and only if  $\beta = \gamma$ . It follows that there are uncountably many Sturmian sequences. Deduce from the fact that there are at most countably many substitutions on  $\mathcal{A} = \{0, 1\}$ , that not all Sturmian sequences can be represented as substitutions.
2. Show that if  $\beta > 0$  has cutting sequence  $S_\beta$ , then  $R(S_\beta) = S_{1/\beta}$  (where  $R(0) = 1$  and  $R(1) = 0$ ).

### 19.4 Sequences Arising from Irrational Rotations are Sturmian.

Let  $T_\alpha : [0, 1) \rightarrow [0, 1)$  be the rotation  $T_\alpha(x) = \{x + \alpha\}$  where  $0 < \alpha < 1$  is irrational. We know from Section 17.4 that in this case,  $T_\alpha$  is minimal, so that the sequence  $\{x + k\alpha\}_{k=0}^\infty$  is dense in  $[0, 1)$  for every  $x \in [0, 1)$ .

Define partitions  $\mathcal{P}$  and  $\mathcal{P}'$  of  $[0, 1)$  by

$$\mathcal{P} = \{[0, 1 - \alpha), [1 - \alpha, 1)\}, \quad \mathcal{P}' = \{[0, 1 - \alpha], (1 - \alpha, 1)\}.$$

A “coding” of the orbit  $O(x)$  of  $T_\alpha$ , with respect to the partition  $\mathcal{P}$  is obtained in the same way as in the last section. Define a sequence  $u = u_1 u_2 \dots$ , of 0’s and 1’s with

$$u_n = 1 \iff \{x + n\alpha\} \in [1 - \alpha, 1).$$

Since  $T_\alpha$  is minimal, the sequence can take the value  $1 - \alpha$  at most once, so we get essentially the same coding using  $\mathcal{P}$  or  $\mathcal{P}'$ . We will use the partition  $\mathcal{P}$  throughout.

If we let  $x = 0$ , we obtain the sequence  $f_\alpha$  defined previously. Set

$$I_0 = [0, 1 - \alpha), \quad I_1 = [1 - \alpha, 1).$$

The following proposition tells us that the coding obtained is independent of  $x \in [0, 1)$  in the sense that the set of factors  $\mathcal{L}(u)$  is independent of  $x$ .

**Proposition 19.4.1** *Let  $u = u_0 u_1 u_2 \dots$ , be the sequence with  $u_n = i \iff \{x + n\alpha\} \in I_i$ . The finite word  $w = w_1 w_2 \dots w_n$  is a factor of  $u$  if and only if there exists  $k \in \mathbb{N}$  such that*

$$\{x + k\alpha\} \in I(w_1, w_2, \dots, w_n) = \bigcap_{j=0}^{n-1} T_\alpha^{-j}(I_{w_{j+1}}).$$

**Proof.** The word  $w = w_1w_2 \dots w_n$  is a factor of  $u$  if and only if there exists  $k \in \mathbb{N}$  with

$$w_1 = u_k, \quad w_2 = u_{k+1}, \quad \dots, \quad w_n = u_{k+n-1},$$

or if and only if

$$\{x + k\alpha\} \in I_{w_1}, \quad \{x + (k+1)\alpha\} \in I_{w_2}, \quad \dots, \quad \{x + (k+n-1)\alpha\} \in I_{w_n},$$

if and only if

$$\{x + k\alpha\} \in I_{w_1}, \quad T_\alpha(\{x + k\alpha\}) \in I_{w_2}, \quad \dots, \quad T_\alpha^{n-1}(\{x + k\alpha\}) \in I_{w_n}.$$

Equivalently,

$$\{x + k\alpha\} \in \bigcap_{j=0}^{n-1} T_\alpha^{-j}(I_{w_{j+1}}).$$

□

Since the set  $I(w_1, \dots, w_n)$  is independent of  $x$ , the set of factors does not depend on the initial point  $x$ . Clearly there are  $2^n$  sets of the form  $I(w_1, \dots, w_n)$  where  $w_i = 0$  or  $w_i = 1$ , being pairwise disjoint with union equal to all of  $[0, 1]$ . We shall show that all but  $n + 1$  of these sets are empty, and the remaining  $n + 1$  are (connected), intervals. It follows that there are exactly  $n + 1$  factors of length  $n$ , so that for each  $x \in [0, 1]$ , the sequence is Sturmian. In particular, for  $0 < \alpha < 1$  irrational,  $f_\alpha$  is a Sturmian sequence.

Since  $T_\alpha$  is a one-to-one and onto map,  $T_\alpha^{-1}$  is defined, and given by

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

Notice that

$$\{T_\alpha^{-k}(0) : 0 \leq k \leq n - 1\} = \{0, \{1 - \alpha\}, \{1 - 2\alpha\}, \dots, \{1 - (n - 1)\alpha\}\},$$

and these points are spread around the unit interval (not necessarily in this order). Set

$$E_n = \{\{1 - k\alpha\}\} : k = 0, 1, 2, \dots, n - 1\} \cup \{1\},$$

and suppose that  $E_n = \{\beta_0^n, \beta_1^n, \dots, \beta_n^n, \beta_{n+1}^n\}$ , where the superscript denotes the dependence on  $n$  and

$$0 = \beta_0^n < \beta_1^n < \dots < \beta_n^n < \beta_{n+1}^n = 1.$$

$E_n$  is a partition of  $[0, 1]$  into  $n+1$  subintervals. Suppose that  $T_\alpha^{-n}(0) = \{1 - n\alpha\} = \gamma$ . Then  $E_{n+1}$  is the same finite set with the addition of  $\gamma$ ,  $\beta_i < \gamma < \beta_{i+1}$  for some  $i$ . We can write

$$E_{n+1} = \{\beta_0^{n+1}, \beta_1^{n+1}, \dots, \beta_{n+1}^{n+1}, \beta_{n+2}^{n+1}\} = E_n \cup \{\gamma\}$$

We will use induction to prove that for each  $n \geq 1$ , there are  $n+1$  non-empty intervals of the form  $I(w_1, \dots, w_n)$ . For example, when  $n=1$ , we have  $E_1 = \{0, 1 - \alpha, 1\} = \{\beta_0^1, \beta_1^1, \beta_2^1\}$ , and the two intervals are  $I(0) = I_0 = [0, 1 - \alpha]$  and  $I(1) = I_1 = [1 - \alpha, 1]$ .

There are two possibilities when  $n=2$ , depending on whether or not the discontinuity  $\alpha$  of  $T_\alpha^{-1}$  lies in the interval  $[0, \beta_1]$  or not. Suppose that  $\beta_1 < \alpha$ . We can check that if  $\gamma = \{1 - 2\alpha\} = 2 - 2\alpha$ , then  $E_2 = \{0, \beta_1^1, \gamma, 1\}$  and

$$I(0, 0) = \emptyset, \quad I(0, 1) = [0, \beta_1^1], \quad I(1, 0) = [\beta_1^1, \gamma], \quad I(1, 1) = [\gamma, 1].$$

If  $\beta_1^1 > \alpha$ , then  $E_2 = \{0, \gamma, \beta_1^1, 1\}$  and

$$I(0, 0) = [0, \gamma], \quad I(0, 1) = [\gamma, \beta_1^1], \quad I(1, 0) = [\beta_1^1, 1], \quad I(1, 1) = \emptyset,$$

where  $\gamma = 1 - 2\alpha = \{1 - 2\alpha\}$ . Notice that in each case we have a partition of  $[0, 1]$  into three intervals. We use these ideas to prove the general case, and deduce that the resulting sequences are Sturmian:

**Proposition 19.4.2** *Let  $x \in [0, 1)$ ,  $0 < \alpha < 1$  irrational, and  $u$  be the sequence of 0's and 1's with*

$$u_n = 1 \iff \{x + n\alpha\} \in [1 - \alpha, 1).$$

*If  $p_u(n)$  is the number of factors of length  $n$  of the sequence  $u$ , then*

$$p_u(n) = n + 1, \quad \text{for all } n \in \mathbb{N}.$$

**Proof.** We need the following lemma:

**Lemma 19.4.3** *Set  $E_n = \{\{1 - k\alpha\} : k = 0, 1, 2, \dots, n-1\} \cup \{1\}$  and suppose that  $E_n = \{\beta_0^n, \beta_1^n, \beta_2^n, \dots, \beta_{n+1}^n\}$ , where*

$$0 = \beta_0^n < \beta_1^n < \dots < \beta_n^n < \beta_{n+1}^n = 1.$$

*If  $I(w_1, \dots, w_n) \neq \emptyset$ , then  $I(w_1, \dots, w_n)$  is an interval of the form  $[\beta_k^n, \beta_{k+1}^n]$ , for some  $0 \leq k \leq n$ . Conversely, any interval of the form  $[\beta_k^n, \beta_{k+1}^n]$  is equal to  $I(w_1, \dots, w_n)$  for some factor  $w_1 w_2 \dots w_n$  of  $u$ .*

**Proof.** The proof is by induction. We have seen that the result is true for  $n = 1$  and  $n = 2$ . Suppose it is true for some fixed  $n$ . Then any non-empty set of the form  $I(w_1, \dots, w_n)$  is an interval equal to  $[\beta_k^n, \beta_{k+1}^n]$ , for some  $0 \leq k \leq n$ . Consider now

$$I(w_1, w_2, \dots, w_n, w_{n+1}) = \bigcap_{j=0}^n T_\alpha^{-j}(I_{w_{j+1}}) = I_{w_1} \cap T_\alpha^{-1}[I(w_2, w_3, \dots, w_{n+1})],$$

where we know by our induction hypothesis, that  $I(w_2, w_3, \dots, w_{n+1})$  is an interval of the form  $[\beta_k^n, \beta_{k+1}^n]$ . There are two cases to consider.

**Case 1.** The discontinuity  $\alpha$  of  $T_\alpha^{-1}$  does not lie in the interval  $[\beta_k^n, \beta_{k+1}^n]$ :

Then  $T_\alpha^{-1}[\beta_k^n, \beta_{k+1}^n] = [\beta_\ell^{n+1}, \beta_{\ell+1}^{n+1}]$  for some  $0 \leq \ell \leq n + 1$  (the integers  $\ell$  and  $\ell + 1$  have to be consecutive, for otherwise there would be some  $\gamma \in E_{n+1}$ , belonging to the interval  $T_\alpha^{-1}[\beta_k^n, \beta_{k+1}^n]$ , with corresponding  $\beta \in E_n$  in the interval  $(\beta_k^n, \beta_{k+1}^n)$ , giving a contradiction).

The intersection of this interval with the interval  $I_{w_1}$  (where  $I_{w_1}$  is either  $I_0$  or  $I_1$ ), is clearly an interval of the correct form.

**Case 2.** The discontinuity  $\alpha$  lies in the interval  $[\beta_k^n, \beta_{k+1}^n]$ :

Then

$$T_\alpha^{-1}[\beta_k^n, \beta_{k+1}^n] = [0, \beta_1^{n+1}) \cup [\beta_{n+1}^{n+1}, 1),$$

( $\beta_1^{n+1}$  and  $\beta_{n+1}^{n+1}$  appear here since other values from  $E_{n+1}$  would lead to a contradiction as in Case 1). Since  $\beta_1^{n+1} < 1 - \alpha$  and  $\beta_{n+1}^{n+1} > 1 - \alpha$  (because  $\beta_k^n \in [0, \alpha]$  and  $\beta_{k+1}^n \in (\alpha, 1)$ ), the intersection of this set with either  $I_0$  or  $I_1$  is again an interval of the required form.

By induction, it follows that for each  $n$ , the sets  $I(w_1, \dots, w_n)$  are intervals of the form  $[\beta_k^n, \beta_{k+1}^n]$ .

Since sets of the form  $I(w_1, \dots, w_n)$  give rise to a partition of  $[0, 1]$ , these sets must account for all sets  $[\beta_k^n, \beta_{k+1}^n]$  for  $0 \leq k \leq n$ , and the lemma follows directly.  $\square$

**Proof of Proposition 19.4.2** We saw from Proposition 19.4.1 that a word  $w_1 w_2 \dots w_n$  appears in the sequence  $u$  if and only if  $I(w_1, \dots, w_n) \neq \emptyset$ . From the lemma, there are exactly  $n + 1$  such sets (all being intervals), and the result follows.  $\square$

### 19.5 Semi-Topological Conjugacy Between $([0, 1], T_\alpha)$ and $(\overline{O(u)}, \sigma)$ .

Let  $u$  be the Sturmian sequence arising as the coding of some irrational rotation  $T_\alpha$ . In other words, if  $\mathcal{P} = \{[0, 1 - \alpha), [1 - \alpha, 1)\}$  is the partition of  $[0, 1]$  defined in Section 19.4, then  $u = u_0u_1\dots$ , where  $u_n = 0$  if  $T_\alpha^n(0) \in [0, 1 - \alpha)$  and  $u_n = 1$  if  $T_\alpha^n(0) \in [1 - \alpha, 1)$ . It is reasonable to ask whether there is a topological conjugacy between the dynamical systems  $([0, 1], T_\alpha)$  and  $(\overline{O(u)}, \sigma)$ . Clearly no such conjugation is possible as there would have to be a homeomorphism between the underlying metric spaces  $[0, 1]$  and  $\overline{O(u)}$ , the first space being connected (an interval), and the second being totally disconnected (a type of Cantor set). In a similar way, the dynamical systems  $([0, 1], T_\alpha)$  and  $(\mathbb{S}^1, R_a)$ , where  $R_a(z) = az$ ,  $a = e^{2\pi i\alpha}$ , are not conjugate as the underlying spaces are not homeomorphic. However, all these maps are what we call *semi-topologically conjugate*.

**Definition 19.5.1** Two dynamical systems  $(X, f)$  and  $(Y, g)$  are *semi-topologically conjugate* if there exist countable sets  $B \subset X$ ,  $C \subset Y$ , and a one-to-one onto function  $\phi : X_1 \rightarrow Y_1$ , where  $X_1 = X \setminus B$ ,  $Y_1 = Y \setminus C$ , which is continuous with continuous inverse, and satisfies  $\phi \circ f(x) = g \circ \phi(x)$ , for all  $x \in X_1$ .

In the case that  $(X, f)$  is a *symbolic system* (for example, a substitution dynamical system), we say that it is a *coding* of  $(Y, g)$ .

**Example 19.5.2** Let  $0 < \alpha < 1$  be irrational. The dynamical systems  $([0, 1], T_\alpha)$  and  $(\mathbb{S}^1, R_a)$  are semi-topologically conjugate when  $a = e^{2\pi i\alpha}$ .

**Proof.** Define  $\phi : [0, 1] \rightarrow \mathbb{S}^1$  by  $\phi(x) = e^{2\pi i x}$ . Then  $\phi(T_\alpha(x)) = R_a(\phi(x))$  for all  $x \in [0, 1]$ . Let  $B = O(0)$ , the orbit of 0 under  $T_\alpha$ , and  $C = \phi(B)$ . We can check that  $\phi$  restricted to  $[0, 1] \setminus B$  has the required properties. □

**Theorem 19.5.3** Let  $0 < \alpha < 1$  be irrational and  $u = u_0u_1\dots$  be the Sturmian sequence given by  $u_n = 0$  if  $T_\alpha^n(0) \in [0, 1 - \alpha)$ , and  $u_n = 1$  if  $T_\alpha^n(0) \in [1 - \alpha, 1)$ . Denote by  $(\overline{O(u)}, \sigma)$  the corresponding shift dynamical system. The dynamical systems  $([0, 1], T_\alpha)$  and  $(\overline{O(u)}, \sigma)$  are semi-topologically conjugate.

**Proof.** Write  $\mathcal{P} = \{[0, 1 - \alpha), [1 - \alpha, 1)\}$ . The partition of  $[0, 1]$  given above. Let  $x \in [0, 1]$ . Define a function  $\phi : [0, 1] \rightarrow \overline{O(u)}$  by

$$\phi(x) = \mathcal{P}\text{-name of } x,$$

where  $\phi(x)$  is the sequence of 0's and 1's whose  $n$ th term is 0 if  $T_\alpha^n(x) \in [0, 1 - \alpha]$ , and is 1 if  $T_\alpha^n(x) \in [1 - \alpha, 1]$ . In particular,  $\phi(0) = u$ , and  $\phi(T_\alpha^n(0)) = \phi(\{n\alpha\}) = \sigma^n(u)$  (where  $\{x\}$  denotes the fractional part of  $x$ ). Clearly  $\phi$  is well defined, and if  $x$  and  $y$  have the same  $\mathcal{P}$ -name, they are equal, since they are not separated by arbitrarily small intervals, so  $\phi$  is one-to-one.  $\phi$  is onto, for if  $\omega \in \overline{\mathcal{O}(u)}$ , then every factor of  $\omega$  is a factor of  $u$ , so if  $\omega = w_1 w_2 \dots w_n \dots$ , then we can find a sequence  $x_n \in I(w_1, w_2, \dots, w_n)$ ,  $n = 1, 2, \dots$ , with  $x_n \rightarrow x$  and  $\phi(x) = \omega$ .

Let  $D = \{\{n\alpha\} : n \in \mathbb{N}\}$ , a countable set dense in  $[0, 1]$ , and  $\phi(D) = \{\sigma^n(u) : n \in \mathbb{N}\}$ . Clearly  $\phi : [0, 1] \setminus D \rightarrow \overline{\mathcal{O}(u)} \setminus \phi(D)$  is also a bijection, and we show that it is a continuous map.

Let  $\epsilon > 0$  and  $x \in [0, 1] \setminus D$ , and choose  $n$  so large that  $1/2^n < \epsilon$ . Choose  $\delta > 0$  so that if  $y \in (x - \delta, x + \delta)$ , then  $(x - \delta, x + \delta) \subset I(w_1, w_2, \dots, w_n)$  for some factor  $w_1 w_2 \dots w_n$  of  $u$ . Now  $\phi(x)$  and  $\phi(y)$  have the same first  $n$  coordinates, so that  $d(\phi(x), \phi(y)) \leq 1/2^n < \epsilon$ , so  $\phi$  is continuous at  $x$ . This argument does not work if  $x \in D$ , since we cannot ensure that  $(x - \delta, x + \delta) \subset I(w_1, w_2, \dots, w_n)$  for any factor  $w_1 \dots w_n$ . A similar argument will show that  $\phi^{-1}$  is a continuous map.

Finally, note that the  $n$ th term of  $\phi(x + \alpha)$ :  $[\phi(x + \alpha)]_n = 0$ , if and only if  $T_\alpha^n(x + \alpha) \in [0, 1 - \alpha]$ . This is equivalent to  $T_\alpha^{n+1}(x) \in [0, 1 - \alpha]$ , and hence to  $[\sigma(\phi(x))]_n = 0$ . It follows that  $\phi \circ T_\alpha(x) = \sigma \circ \phi(x)$  for  $x \in [0, 1]$ . □

**Remark 19.5.4** We showed earlier that if  $u$  is a recurrent sequence, then the dynamical system  $(\overline{\mathcal{O}(u)}, \sigma)$  has the property of being onto. Since a Sturmian sequence is recurrent, the shift map  $\sigma : \overline{\mathcal{O}(u)} \rightarrow \overline{\mathcal{O}(u)}$ , is necessarily onto. The next result tells us that it is essentially one-to-one.

**Proposition 19.5.5** *The shift map  $\sigma : \overline{\mathcal{O}(u)} \rightarrow \overline{\mathcal{O}(u)}$ , for  $u$  a Sturmian sequence, is one-to-one except at one point.*

**Proof.** Since  $u$  is recurrent, every factor in  $\mathcal{L}_n(u)$  (the words in  $u$  of length  $n$ ), appears infinitely often in  $u$ , so can be extended on the left. Since  $|\mathcal{L}_n(u)| = n + 1$ , there is exactly one factor of length  $n$  that can be extended on the left in two different ways (called a *left special factor*).

Let  $L_n$  be the unique word in  $\mathcal{L}_n(u)$  that can be extended in 2 different ways. In particular,

$$0 L_n, 1 L_n \in \mathcal{L}_{n+1}(u).$$

Consider the word  $L_{n+1}$ , which can be extended in two ways to give  $0 L_{n+1}$  and  $1 L_{n+1}$ . It is clear that

$$L_{n+1} = L_n a, \quad \text{for some } a \in \{0, 1\},$$

since an extension of  $L_{n+1}$  on the left gives rise to an extension of  $L_n$  on the left, and  $L_n$  is unique.

Suppose now that  $w \in \overline{O(u)}$  is a sequence with two preimages:  $v_1, v_2 \in \overline{O(u)}$  with

$$\sigma(v_1) = w = \sigma(v_2).$$

Then every initial factor of  $w$  of length  $n$  has two extensions: i.e.,  $w = L_n v$  for some sequence  $v$ , for each  $n \in \mathbb{N}$ . Such a  $w$  is clearly unique. □

### Exercises 19.5

1. Let  $F = \{v_1, v_2\}$  where  $v_1$  and  $v_2$  are the two sequences in  $\overline{O(u)}$  from Proposition 19.5.5 with  $\sigma(v_1) = \sigma(v_2)$  and  $v_1 \neq v_2$ . Show that if  $D = \bigcup_{n \in \mathbb{Z}} \sigma^n(F)$ , then  $\sigma : \overline{O(u)} \setminus D \rightarrow \overline{O(u)} \setminus D$ , is a homeomorphism.
2. (a) Let  $\theta(0) = 01$ ,  $\theta(1) = 0$  be the Fibonacci substitution. Set  $u = \theta^\infty(0)$ , a Sturmian sequence. Find the left special factors  $L_1, L_2, L_3, \dots$ , mentioned in Proposition 19.5.5. Give a conjecture on what the sequence  $w = \lim_{n \rightarrow \infty} L_n$  equals, and prove your conjecture. (Hint: You may find Exercise 18.4 # 2 useful).
3. Prove that every prefix of a Sturmian sequence  $u$  appears at least twice in  $\overline{O(u)}$ . Deduce that the shift map  $\sigma : \overline{O(u)} \rightarrow \overline{O(u)}$  is not one-to-one.
4. Show that the angle doubling map  $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ,  $f(z) = z^2$  and the full shift map  $\sigma : \Sigma \rightarrow \Sigma$ , where  $\Sigma = \{0, 1\}^{\mathbb{N}}$ , are semi-topologically conjugate. (Hint: Define  $h$  on the subset of  $\Sigma$  consisting of those sequences consisting of infinitely many 0's and 1's).
5. Prove that semi-topological conjugacy is an equivalence relation.

## 19.6 The Three Distance Theorem.

Starting with  $n = 1$  we defined the partition  $E_1 = \{0, 1 - \alpha, 1\} = \{\beta_0^1, \beta_1^1, \beta_2^1\}$ , giving rise to subintervals of  $[0, 1)$  having lengths  $1 - \alpha$  and  $\alpha$  respectively. The sets  $E_2, E_3, \dots$ , can be defined inductively by adding a single member of the orbit of 0 under  $T_\alpha^{-1}$  at each stage, giving rise to new subintervals of  $[0, 1)$ , of decreasing lengths. It is surprising that for each fixed  $n \geq 1$ , the lengths of the intervals  $[\beta_k^n, \beta_{k+1}^n)$  created by the partition  $E_n$  take at most three values. This is the celebrated “*Three Distance Theorem*”, which was conjectured by Steinhaus and proved by V. T. Vos. We digress somewhat from our studies of dynamical systems to give a proof which uses the properties of Sturmian sequences, and also uses results from the theory of equidistribution (see Appendix D on Weyl’s Equidistribution Theorem). The proof in 19.6.1 below, is due to V. Berthé, [13], [14]. The Three Distance Theorem can be stated in terms of the orbit of 0 under  $T_\alpha$  as:

**Theorem 19.6.1** (The Three Distance Theorem.)

*Let  $0 < \alpha < 1$  be irrational and  $n \in \mathbb{Z}^+$ . The points  $\{k\alpha\}$ , for  $0 \leq k \leq n$ , partition the interval  $[0, 1)$  into  $n + 1$  intervals, the lengths of which take at most three values, one being the sum of the other two.*

There are many proofs of the Three Distance Theorem, usually of a combinatorial nature. Our dynamical proof due to V. Berthé, uses the theory developed so far concerning Sturmian sequences. In order to prove this theorem we need to define the *Rauzy graph* of a sequence.

**Definition 19.6.2** Let  $u \in \mathcal{A}^\mathbb{N}$  be a sequence over a finite alphabet  $\mathcal{A}$ . The *Rauzy graph*  $\Gamma_n$  of  $u$  is an oriented graph whose vertices and edges are defined as follows:

- (i) The vertices  $U, V, \dots$  of  $\Gamma_n$  are the factors (words) of length  $n$  appearing in  $u$  (i.e., the members of  $\mathcal{L}_n(u)$ ).
- (ii) There is an edge from the vertex  $U$  to the vertex  $V$  if  $V$  follows  $U$  in the sequence  $u$ .

More precisely:  $V$  follows  $U$  if there is a factor  $W$  of  $u$  and  $x, y \in \mathcal{A}$  such that

$$U = xW, \quad V = Wy \quad \text{and} \quad xWy \quad \text{is a factor of } u.$$

Such an edge is labeled  $xWy$ .

Recall that the complexity function  $p_u(n) = |\mathcal{L}_n(u)|$ , is the number of distinct factors of length  $n$ . We need some straightforward properties of the Rauzy graph:

**Proposition 19.6.3**  $p_u(n) = \text{the number of vertices of } \Gamma_n$ ,  $p_u(n+1) = \text{the number of edges of } \Gamma_n$ .

**Proof.** The first statement is clear. If  $w = u_1u_2\dots u_{n+1}$  is a factor of length  $n+1$ , set  $W = u_2\dots u_n$ . Then if  $U = u_1W$  and  $V = Wu_{n+1}$ ,  $U$  and  $V$  are factors of length  $n$ , so are vertices of the graph.  $w = u_1Wu_{n+1}$  is a factor of length  $n+1$ , which is an edge joining them. In other words, every factor of length  $n+1$  appears as an edge, and it cannot appear in more than one place.

On the other hand, by definition, every edge corresponds to a factor of length  $n+1$ . □

Recall that a *right extension* of a factor  $w = w_1\dots w_n$  of  $u$ , is a factor of the form  $w_1w_2\dots w_nx$ . *Left extensions* are defined similarly. A factor having more than one right extension is called a *right special factor* (respectively *left special factor*). Following, we use the symbol  $\#$  to denote the number of members of a set.

**Definition 19.6.4** Let  $U$  be a vertex of  $\Gamma_n$ . Set

$$U^+ = \# \text{ of edges of } \Gamma_n \text{ originating at } U, \quad U^- = \# \text{ of edges of } \Gamma_n \text{ ending at } U.$$

**Proposition 19.6.5**

$$p_u(n+1) - p_u(n) = \sum_{|U|=n} (U^+ - 1) = \sum_{|U|=n} (U^- - 1).$$

**Proof.** Here we are summing over all the vertices (factors of length  $n$ ), so that  $\sum_{|U|=n} 1 = p_u(n)$ . Since  $U^+$  gives the total number of edges (factors of length  $n+1$ ), originating at the vertex  $U$ ,  $\sum_{|U|=n} U^+$  will give the total number of factors of length  $n+1$ . □

**Definition 19.6.6** A *branch* of the Rauzy graph  $\Gamma_n$  is a sequence of adjacent edges  $(U_1, U_2, \dots, U_m)$  of maximal length (possibly empty), with the property:

$$U_i^+ = 1 \quad \text{for } i < m, \quad U_i^- = 1, \quad \text{for } i > 1.$$

In order to prove the Three Distance Theorem, we need to relate the frequency of a factor of some length  $n$  appearing in  $u$ , with the lengths of the intervals  $I(w_1, w_2, \dots, w_n)$ , which are intervals of the form  $[\beta_k^n, \beta_{k+1}^n]$  (defined in the last section).

**Definition 19.6.7** Let  $w$  be a factor of the sequence  $u = u_1 u_2 \dots$ . The *frequency*  $f(w)$  of  $w$  in  $u$  is the limit (if it exists):

$$f(w) = \lim_{k \rightarrow \infty} \frac{\#\{\text{occurrences of } w \text{ in } u_1 u_2 \dots u_k\}}{k}.$$

From Proposition 19.6.8 below, we can deduce that frequencies always exist for factors of Sturmian sequences. In addition, for each fixed  $n$ , the frequencies of factors of length  $n$  can take at most three different values. Certain words necessarily have the same frequencies.

**Proposition 19.6.8** *If  $U$  and  $V$  are vertices of  $\Gamma_n$  linked by an edge  $xWy$  with  $U^+ = 1$ ,  $V^- = 1$ , then  $f(U) = f(V)$ . In addition, the vertices of a branch have the same frequencies.*

**Proof.**  $U = xW$  and  $V = Wy$ . Since  $U^+ = 1$ ,  $U$  has the unique right extension  $Uy$ , and since  $V^- = 1$ ,  $V$  has the unique left extension  $xV$ . It follows for example that the frequency of the appearance of  $U$  in  $u$  and  $Uy$  in  $u$  are identical, so that

$$f(U) = f(Uy) = f(xWy) = f(xV) = f(V).$$

Continuing this argument, we see that the vertices of any branch must have the same frequencies. □

The following is fundamental in proving the Three Distance Theorem (see [13]).

**Proposition 19.6.9** *Let  $u$  be a recurrent sequence having complexity function  $p_u(n)$ . The number of different values the frequencies of factors of length  $n$  can have, is at most*

$$3[p_u(n+1) - p_u(n)].$$

**Proof.** Let  $V_1$  be the set of all factors of length  $n$  of the sequence  $u$ , having more than one extension. In other words,  $V_1$  is the set of those vertices  $U$  in  $\Gamma_n$  for which

$U^+ \geq 2$ . The cardinality of  $V_1$  must satisfy:

$$|V_1| = \sum_{|U|=n, U^+ \geq 2} 1 \leq \sum_{|U|=n} (U^+ - 1) = p_u(n+1) - p_u(n).$$

Let  $V_2$  be the subset of the set of vertices of  $\Gamma_n$  defined in the following way:

$U \in V_2$  if and only if  $U^+ = 1$ , and if  $V$  is the unique vertex such that there is an edge from  $U$  to  $V$  in  $\Gamma_n$ , then  $V^- \geq 2$ .

In other words,  $U \in V_2$  if and only if  $U = xW$  where  $x \in \mathcal{A}$ , and where the factor  $xW$  of the sequence  $u$  has a unique right extension  $xWy$ , and  $Wy$  has at least two left extensions. The cardinality of  $V_2$  satisfies:

$$|V_2| \leq \sum_{V^- \geq 2} V^- = \sum_{V^- \geq 2} (V^- - 1) + \sum_{V^- \geq 2} 1 \leq 2(p_u(n+1) - p_u(n)).$$

It follows that there are at most  $3(p_u(n+1) - p_u(n))$  factors in  $V_1 \cup V_2$ . Now let  $U$  be a factor of length  $n$  that does not belong to either of  $V_1$  or  $V_2$ . Then  $U^+ = 1$ , and the unique word  $V$  for which there is an edge from  $U$  to  $V$  in  $\Gamma_n$  satisfies  $V^- = 1$ . From the previous proposition,  $U$  and  $V$  have the same frequencies:  $f(U) = f(V)$ . Now consider the path in  $\Gamma_n$ , starting at  $U$  and consisting of vertices that do not belong to  $V_1$  or  $V_2$ . The last vertex of this path belongs to either  $V_1$  or  $V_2$  and has the same frequency as  $U$ .

□

**Proof of the Three Distance Theorem.** We are assuming that the sequence  $u$  arises as the coding of the sequence  $\{\{x + k\alpha\} : k \in \mathbb{N}\}$ , so that it is Sturmian and  $p_u(n) = n + 1$  and  $p_u(n+1) - p_u(n) = 1$  for each  $n$ . Proposition 19.6.8 tells us that there are at most three values that the frequencies  $f(w)$  can take for any factor of length  $n$ . Suppose  $w = w_1 \dots w_n$ . Then we showed in Section 19.4 that  $w$  appears in  $u$  if and only if  $I(w_1, \dots, w_n) \neq \emptyset$ , and in this case,  $I(w_1, \dots, w_n)$  is an interval formed by the partitioning of  $[0, 1]$  using the points  $\{T_\alpha^{-k}(0) : 0 \leq k \leq n\}$ . Now using Weyl's Equidistribution Theorem (see Appendix D), the frequency of the sequence  $\{x + k\alpha\}$  in a subinterval of  $[0, 1]$  is equal to the length of that interval. It now follows that each of these subintervals can have at most three different lengths.

□

## Exercises 19.6

1. Construct a sequence  $u$  for which the frequencies of letters do not exist.

2. Show that the Rauzy graph  $\Gamma_n$  is always connected, i.e., given any two vertices  $U$  and  $V$ , there are edges that connect  $U$  and  $V$ .

## CHAPTER 20

### The Multiple Recurrence Theorem of Furstenberg and Weiss.

In 1978, H. Furstenberg and B. Weiss [50], gave a new proof of van der Waerden’s Theorem on arithmetic progressions in the set of integers. If we partition the integers as  $\mathbb{Z} = \cup_{i=1}^N C_i$  into a finite partition, then it is clear that at least one of the sets must contain infinitely many integers (an *infinite pigeonhole principle*). It is not so obvious that one of the sets must contain arithmetic progressions of arbitrary finite length. This is an important result due to van der Waerden [125], which according to Khintchine [75], is one of the “Three pearls of number theory”.

#### 20.1 van der Waerden’s Theorem.

**Theorem 20.1.1** (van der Waerden, 1927). *Let  $\mathbb{Z} = \cup_{j=1}^N C_j$  be a finite partition of the integers. There exists  $1 \leq i \leq N$  such that  $C_i$  contains arithmetic progressions of arbitrary length, (i.e., for each  $m \geq 1$ , there exists  $c \in \mathbb{Z}$  and  $d \geq 1$ , such that  $c + jd \in C_i$ , for  $0 \leq j \leq m - 1$ ).*

Our proof of van der Waerden’s Theorem uses the Multiple Recurrence Theorem of Furstenberg and Weiss concerning a topological dynamical system  $(X, T)$ . Theorem 20.1.2 is the topological version of a much deeper result about measure preserving transformations due to Furstenberg ([51]). Furstenberg’s measure theoretic multiple recurrence theorem revolutionized the field of combinatorial number theory, enabling the proof of many results seemingly out of reach using other methods. It also led to the proof of the Green-Tao Theorem [64], which says there are arbitrarily long arithmetic progressions within the prime numbers. It is a surprising fact that the fields of topological dynamics, and measurable dynamics (ergodic theory), have analogous theories. It was therefore natural to prove a topological version of the recurrence theorem after the discovery of the measure theoretic version.

Following is the topological version of the multiple recurrence theorem that we shall prove. There are numerous generalizations of this result, but to avoid minor complications, we shall assume that  $T : X \rightarrow X$  is a homeomorphism. Just as we

saw in Section 17.4,  $T$  will have a minimal set  $Y$  (a  $T$ -invariant set on which  $T$  is minimal). It follows as in Theorem 17.4.4, that for any open set  $V \subset Y$ , there exists  $M > 0$  with  $Y = \bigcup_{|n| \leq M} T^{-n}V$ . The Birkhoff Transitivity Theorem tells us that for any open set  $V \subset Y$ , there exists  $n \in \mathbb{Z}^+$  with  $V \cap T^nV \neq \emptyset$ . The following is a significant generalization of this result. Our approach is based on [103].

**Theorem 20.1.2** (Multiple Recurrence Theorem: Furstenberg and Weiss, 1978). *Let  $T : X \rightarrow X$  be a homeomorphism of a compact metric space  $X$ . Let  $Y$  be a minimal subset of  $X$ , and  $V$  any open subset of  $Y$ . For all  $N \in \mathbb{N}$  there exists  $k \geq N$  and  $d \geq 1$  such that*

$$V \cap T^{-d}V \cap T^{-2d}V \cap \cdots \cap T^{-(k-1)d}V \neq \emptyset.$$

For  $k \geq 1$ , let  $P(k)$  be the statement: “there exists  $d \geq 1$  such that  $V \cap T^{-d}V \cap T^{-2d}V \cap \cdots \cap T^{-(k-1)d}V \neq \emptyset$ .” We use induction to prove Theorem 20.1.2. When  $k = 1$ , the result is clear:  $P(1)$  is true. Assume that  $P(k - 1)$  holds. We show that  $P(k)$  holds and this will be sufficient to prove Theorem 20.1.2. From the transitivity on  $Y$ , we know that there is an  $M > 0$  with  $Y = \bigcup_{n=-M}^M T^{-n}V$ . First we give a preliminary lemma.

**Lemma 20.1.3** *For each  $\ell \geq 0$ , we can choose points  $x_j$  and open sets  $T^{-n_j}V$ ,  $-M \leq n_j \leq M$ , with  $x_j \in T^{-n_j}V$ ,  $j = 0, 1, \dots, \ell$ , and natural numbers  $N_0 < N_1 < \cdots < N_\ell$  such that for all  $r, s \in \mathbb{N}$ :  $0 \leq r \leq s \leq \ell$ ,*

$$T^{j(N_s - N_r)}x_r \in T^{-n_s}V, \quad \text{for } j = 0, 1, \dots, \ell.$$

**Proof.** Call the above statement  $Q(\ell)$ . We use an inductive argument on  $\ell$ . It is trivial that  $Q(0)$  holds. Assume we know  $Q(\ell - 1)$  holds (within the induction on  $k$ , where we are assuming  $P(k - 1)$  holds).

Choose a small open set  $V_0$  containing  $x_{\ell-1}$ . Since  $P(k - 1)$  holds, select  $d$  such that there exists

$$y \in V_0 \cap T^{-d}V_0 \cap T^{-2d}V_0 \cap \cdots \cap T^{-(k-2)d}V_0.$$

Set  $x_\ell = T^{-d}y$ , and  $N_\ell = N_{\ell-1} + d$ . We can choose some  $n_\ell$  with  $x_\ell \in T^{-n_\ell}V_0$  and  $|n_\ell| \leq M$  (using our remark about  $Y$  being a minimal set). In particular, for each  $0 \leq r < \ell$ :

$$T^{j(N_\ell - N_r)}x_\ell = T^{j(N_{\ell-1} - N_r)}(T^{(j-1)d}y) \in T^{j(N_{\ell-1} - N_r)}(V_0),$$

for  $j = 1, 2, \dots, k - 1$ . Then provided  $V_0$  is sufficiently small,  $Q(\ell)$  follows from  $Q(\ell - 1)$ . The additional results for  $x_\ell$  come from those for  $x_{\ell-1}$ , and the continuity of  $T^{j(N_\ell - N_r)}$ .

(To understand what is happening here, it is instructive to consider going from  $\ell = 1$  to  $\ell = 2$ . In this case we have  $x_0 \in T^{-n_0}V$ ,  $x_1 \in T^{-n_1}V$  and  $T^{(N_1 - N_0)}x_1 \in T^{-n_1}V$ . Taking  $V_0$  small containing  $x_1$ , set  $x_2 = T^{-d}y$ , ( $y$  chosen as above), with  $n_2$  such that  $x_2 \in T^{-n_2}V_0$  and  $N_2 = N_1 + d$ . It needs to be shown that  $T^{j(N_2 - N_1)}x_0$  and  $T^{j(N_2 - N_0)}x_1$  both belong to  $T^{-n_2}V$  for  $j = 0, 1, 2$ . Take a look at what the above argument shows (with a suitable picture) and use the continuity for  $V_0$  small enough).  $\square$

**Proof of Theorem 20.1.2.** It suffices to show that  $P(k)$  follows from  $P(k - 1)$ . We apply Lemma 20.1.3, with  $\ell = 2M + 1$ . By the pigeonhole principle, we can select  $r, s$ ,  $0 \leq r < s \leq 2M + 1$  such that  $n_r = n_s \in \{-M, \dots, M\}$ . Setting  $x = x_r$ , and  $d = N_s - N_r$ , gives a point in the intersection for  $P(k)$ . This completes the inductive step and the proof of the Theorem 20.1.2.  $\square$

To prove van der Waerden's Theorem, we apply the Multiple Recurrence Theorem to the (two-sided) shift map  $\sigma$  defined on a suitable sequence space. Recall that  $\mathbb{Z} = \bigcup_{j=1}^N C_j$ . Set

$$\Sigma = \{1, 2, \dots, N\}^{\mathbb{Z}} = \{(\dots, x_{-1}, x_0, x_1, x_2, \dots) : x_i \in \{1, 2, \dots, N\}\},$$

the space of all two-sided sequences with entries from  $\{1, 2, \dots, N\}$ . Give  $\Sigma$  the metric defined, for example, by

$$d(\omega_1, \omega_2) = \sum_{n \in \mathbb{Z}} \frac{|x_i - y_i|}{N^{|i|}},$$

where  $\omega_1 = (\dots, x_{-1}, x_0, x_1, x_2, \dots)$  and  $\omega_2 = (\dots, y_{-1}, y_0, y_1, y_2, \dots)$ , again a compact metric space and  $\sigma : \Sigma \rightarrow \Sigma$  is a homeomorphism. Define a sequence  $u = (u_n) \in \Sigma$  associated with the partition  $\{C_j : 1 \leq j \leq N\}$  by

$$u_n = i, \text{ if and only if } n \in C_i.$$

Set  $Y = \overline{O(u)}$ , the closure of the two-sided orbit of  $u$  under  $\sigma$ . The restriction  $\sigma : Y \rightarrow Y$  is a transitive homeomorphism. We may assume that  $Y$  is a minimal set, for if not we can replace it by a minimal set that it contains. Van der Waerden's Theorem now follows directly from multiple recurrence theorem and the following lemma:

**Lemma 20.1.4** Let  $\sigma : Y \rightarrow Y$  be the two-sided shift defined above. Set  $[i] = \{w = (x_n) \in \Sigma : x_0 = i\}$ , where  $1 \leq i \leq N$ . Assume that

$$Y \cap [i] \cap \sigma^{-d}[i] \cap \sigma^{-2d}[i] \cap \cdots \cap \sigma^{-(k-1)d}[i] \neq \emptyset,$$

for some  $d \geq 1$ , and  $k \geq 1$ . Then  $C_i$  contains an arithmetic progression of length  $k$ .

**Proof.** Since the intersection above is an open set in  $Y$ , it contains  $\sigma^n(u)$  for some  $n \in \mathbb{Z}$ . It follows that  $u_{n+jd} = i$  for  $0 \leq j \leq k - 1$ . This says that  $n, n + d, \dots, n + (k - 1)d \in C_i$ , proving the lemma.  $\square$

**20.1.5 A Birkhoff Multiple Recurrence Theorem.** If  $T : X \rightarrow X$  is a homeomorphism of a compact metric space  $X$ , Birkhoff's Recurrence Theorem may be stated in the form: *there exists  $x \in X$  and a sequence  $r_i \rightarrow \infty$  such that  $d(T^{r_i}x, x) \rightarrow 0$  as  $i \rightarrow \infty$ .* Using the multiple recurrence theorem, we can generalize this result as follows:

**Proposition 20.1.6** Let  $T : X \rightarrow X$  be a homeomorphism of a compact metric space  $X$ . For each  $k \geq 1$ , there exists  $x \in X$  and a sequence  $r_i \rightarrow \infty$ , such that  $\max_{0 \leq n \leq k-1} \{d(T^{nr_i}x, x)\} \rightarrow 0$  as  $i \rightarrow \infty$ .

**Proof.** See Exercises 20.1.

### Exercises 20.1

1. If  $\mathbb{Z} = C_0 \cup C_1$ , where  $C_0 = \text{odd integers}$  and  $C_1 = \text{even integers}$ , find the sequence  $u$  described in the proof of Theorem 20.1.1, and deduce  $\overline{\mathcal{O}(u)}$ .
2. Show that van der Waerden's Theorem does not hold for infinite length progressions: there is a finite partition of  $\mathbb{Z}$  such that no partition member contains an infinite length arithmetic progression.
3. First use Theorem 20.1.2 to prove Birkhoff's Recurrence Theorem, and then generalize your proof to give a proof of the Multiple Recurrence Theorem (Proposition 20.1.6).

4. Show that if a set  $A \subseteq \mathbb{Z}$  contains arbitrarily long arithmetic progressions, and  $A = A_1 \cup \dots \cup A_n$ , then some  $A_i$ ,  $1 \leq i \leq n$ , also contains arbitrarily long arithmetic progressions.



## APPENDIX A

### Theorems from Calculus.

We list some important theorems from analysis which we assume without proof, to be used at various times in the text. We refer the reader to [117] for more detail concerning the reals and general metric spaces. In many cases we have provided hints to enable the reader to prove these results as exercises in earlier chapters.

The real numbers  $\mathbb{R}$ , with the metric  $d(x, y) = |x - y|$ , is a complete metric space, i.e., every Cauchy sequence in  $\mathbb{R}$  is convergent (see Section 10.4). This completeness is a consequence of the *Completeness Axiom* for  $\mathbb{R}$ : *Let  $S \subset \mathbb{R}$  be a non-empty set which is bounded above, then there exists  $U \in \mathbb{R}$  with the properties:*

- (i)  $U \geq x$  for all  $x \in S$ .
- (ii) If  $M \geq x$  for all  $x \in S$ , then  $U \leq M$ .

We call this number  $M$  the *supremum* or *least upper bound* of  $S$ , and we write  $M = \sup(S)$ . The *infimum* or *greatest lower bound* of a set which is non-empty and bounded below is defined similarly. It is denoted by  $\inf(S)$ .

**A1. The Completeness of  $\mathbb{R}$ .** *Every Cauchy sequence  $(x_n)$  in  $\mathbb{R}$  is convergent.*

**A2. The Intermediate Value Theorem.** *Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function. Suppose that  $w$  lies between  $f(a)$  and  $f(b)$ . Then there exists  $c \in (a, b)$  with  $f(c) = w$ .*

**A3. The Monotone Sequence Theorem.** (i) *Let  $(x_n)$  be a sequence of real numbers that is increasing and bounded above. Then  $\lim_{n \rightarrow \infty} x_n$  exists and is equal to  $\sup\{x_n : n \in \mathbb{Z}^+\}$ .*

(ii) *Similarly, if the sequence  $(x_n)$  is decreasing and bounded below, then  $\lim_{n \rightarrow \infty} x_n$  exists and is equal to  $\inf\{x_n : n \in \mathbb{Z}^+\}$ .*

**A4. The Closed Bounded Interval Theorem.** *If  $f : [a, b] \rightarrow \mathbb{R}$  is continuous, then  $f([a, b]) = [c, d]$ , a closed bounded interval, for some  $c, d \in \mathbb{R}$ .*

**A5. Rolles' Theorem and the Mean Value Theorem.** *Let  $f : [a, b] \rightarrow \mathbb{R}$  be a function continuous on  $[a, b]$  and differentiable on  $(a, b)$ .*

- (a) *There exists  $c \in (a, b)$  with  $f'(c) = \frac{f(b) - f(a)}{b - a}$ .*
- (b) *If  $f(a) = f(b)$ , then  $c$  can be chosen so that  $f'(c) = 0$ .*

**A5. The Bolzano-Weierstrass Theorem.** *If  $(x_n)$  is an infinite sequence in the closed interval  $[a, b]$ , then it has a limit point in  $[a, b]$ .*

A result about continuous functions that is frequently used is the following:

**A6.** *Let  $f : [a, b] \rightarrow \mathbb{R}$  be continuous at  $x = c$ , where  $c \in (a, b)$ . If  $f(c) > 0$ , there exists  $\delta > 0$  and  $I = (c - \delta, c + \delta)$  such that if  $x \in [a, b]$  and  $x \in I$ , then  $f(x) > 0$ .*

**Proof.** Set  $\epsilon = f(c)/2 > 0$ . Then since  $f$  is continuous at  $x = c$ , there exists  $\delta > 0$  such that if  $x \in [a, b]$  and  $|x - c| < \delta$ , then  $|f(x) - f(c)| < \epsilon$ . This implies that if  $x \in I$ , then  $f(x) > f(c) - \epsilon = f(c) - f(c)/2 = f(c)/2 > 0$ .

□

## APPENDIX B

### The Baire Category Theorem.

Metric spaces are defined in Chapter 4; with complete metric spaces being defined in Chapter 10. Let  $(X, d)$  be a metric space. A *Baire space* is a metric space with the following property: for each collection of sets  $\{U_n : n \in \mathbb{N}\}$  which are open and dense in  $X$ , their intersection  $\bigcap_{n=1}^{\infty} U_n$  is dense in  $X$ .

**B1. The Baire Category Theorem.** *Every complete metric space  $X$  is a Baire space.*

**Proof.** Let  $(U_n)$  be a sequence of dense open sets in  $X$ , and let  $V$  be non-empty and open in  $X$ . It suffices to show that  $V \cap \bigcap_{n=1}^{\infty} U_n \neq \emptyset$ .

Since  $U_1$  is dense, there exists  $x_1 \in U_1 \cap V$ .  $U_1 \cap V$  is open, so there exists  $B_1 = B_{r_1}(x_1) \subset U_1 \cap V$  for some  $r_1 > 0$ .

Since  $U_2$  is dense, there exists  $x_2 \in U_2 \cap B_1$ . Again we can find  $r_2 > 0$  so that  $B_2 = B_{r_2}(x_2)$  is contained in  $U_2$ , and by taking  $r_2$  small enough, we may assume that  $\overline{B}_2 \subset B_1$ .

Continuing inductively, we find open balls  $B_n = B_{r_n}(x_n)$  with  $\overline{B}_n \subset B_{n-1}$ ,  $B_n \subset U_n$ , and  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ .

Fix  $N \in \mathbb{N}$ . Then for all  $m, n > N$ ,  $x_m, x_n \in B_N$ . This says that  $d(x_n, x_m) \leq 2r_N$  for all  $m, n > N$ . Since  $r_n \rightarrow 0$ , it follows that  $(x_n)$  is a Cauchy sequence.  $X$  is a complete space, so there exists  $x \in X$  with  $x_n \rightarrow x$  as  $n \rightarrow \infty$ .

For all  $n > N$ ,  $x_n \in B_{N+1}$  and so  $x \in \overline{B}_{N+1} \subset B_N \subset U_N$ . Clearly  $x \in V$ , since all of the  $B_n$ 's are contained in  $V$ , so  $x \in V \cap \bigcap_{n=1}^{\infty} U_n$ , and the result follows.  $\square$

A subset of the metric space  $X$  is *nowhere dense* if it does not contain any open sets. A point  $x \in X$  is *isolated* if there exists  $r > 0$  with  $B_r(x) = \{x\}$ . The following are immediate consequences of the Baire Category Theorem.

**B2. Corollary** *A non-empty, complete metric space  $X$  is not the countable union of nowhere dense closed sets.*

**B3. Corollary** *Every complete metric space with no isolated points is uncountable.*

**Proof.** Suppose that  $X$  is a countable complete metric space having no isolated points. If  $x \in X$ , then the singleton set  $\{x\}$  is nowhere dense in  $X$ , so  $X$  is the countable union of nowhere dense sets, contradicting the Baire Category Theorem (Corollary B2). □

## APPENDIX C

### The Complex Numbers.

Various properties of the complex numbers are reviewed as needed in Section 6.1 and in Chapter 14. Here we give more detail on certain results that are used, but not proved in the main text. The complex numbers  $\mathbb{C} = \{a + ib : a, b \in \mathbb{R}, i^2 = -1\}$  forms a field with respect to the usual laws of addition and multiplication.  $\mathbb{C}$  is a complete metric space if a distance  $d$  is defined on  $\mathbb{C}$  by

$$d(z, w) = |z - w|,$$

where  $|z| = \sqrt{a^2 + b^2} = \sqrt{z\bar{z}}$ , when  $z = a + ib$ ,  $a, b \in \mathbb{R}$ , and  $\bar{z} = a - ib$ .

**C1. Definition.** Let  $V \subseteq \mathbb{C}$  be open. A function  $f : V \rightarrow \mathbb{C}$  is *analytic* if the derivative  $f'(z)$  is defined and continuous throughout  $V$ .

This is equivalent to  $f$  having a power series expansion about any point  $z_0$  in  $V$  which converges to  $f$  in some ball surrounding  $z_0$ .

**C2. Examples.** The exponential function  $e^z$  and the trigonometric functions  $\sin(z)$  and  $\cos(z)$  are analytic functions (analytic on the whole complex plane - these are called *entire functions*). Their power series representations, valid at any point in  $\mathbb{C}$ , are given by:

$$\begin{aligned} e^z &= 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots + \frac{z^n}{n!} + \cdots, \\ \cos(z) &= \frac{1}{2}(e^{iz} + e^{-iz}) = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots, \\ \sin(z) &= \frac{1}{2i}(e^{iz} - e^{-iz}) = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots, \end{aligned}$$

We remind the reader of the following standard theorems from a first course in complex analysis:

**C3. The Maximum Modulus Principle.** *A non-constant analytic function cannot attain its maximum absolute value at any interior point of its region of definition.*

**C4. Cauchy's Estimate.** *If an analytic function  $f$  maps  $B_r(z_0) = \{z \in \mathbb{C} : |z - z_0| < r\}$  inside some ball of radius  $s$ , then*

$$|f'(z_0)| \leq s/r.$$

**C5. Liouville's Theorem.** *A bounded entire function is constant.*

**Proof.** If  $M = \max\{|f(z)| : z \in \mathbb{C}\}$ , then by Cauchy's Estimate,  $|f'(z)| \leq M/r$ , where  $r$  can be made arbitrarily large. It follows that  $f'(z) = 0$  for all  $z \in \mathbb{C}$ , so  $f$  is constant.  $\square$

Write  $\mathbb{D} = \{z \in \mathbb{C} : |z| \leq 1\}$ , the closed unit ball in the complex plane.

**C6. The Schwarz Lemma.** *Let  $f : \mathbb{D} \rightarrow \mathbb{D}$  be an analytic function with  $f(0) = 0$ . Then*

- (i)  $|f'(0)| \leq 1$ .
- (ii) *If  $|f'(0)| = 1$ , then  $f(z) = cz$  for some  $c \in \mathbb{S}^1$ , the unit circle.*
- (iii) *If  $|f'(0)| < 1$ , then  $|f(z)| < |z|$  for all  $z \neq 0$ .*

**Proof.** (i) Since  $f : \mathbb{D} \rightarrow \mathbb{D}$ ,  $|f(z)| \leq 1$  always, so the maximum value  $|f(z)|$  can take is 1. It follows from the Maximum Modulus Principle, that if  $|z| < 1$ , then  $|f(z)| < 1$ .

Note that since  $f(0) = 0$ ,  $g(z) = f(z)/z$  is defined and analytic throughout  $\mathbb{D}$  (we can see this from the power series representation of  $g(z)$ ). If  $|z| = r < 1$ , then  $|g(z)| < 1/r$ . Since this is true for all  $0 < r < 1$ , it follows that  $|g(z)| \leq 1$  for all  $z \in \mathbb{D}$ . Now

$$|f'(0)| = \lim_{z \rightarrow 0} \left| \frac{f(z) - f(0)}{z - 0} \right| = \lim_{z \rightarrow 0} |g(z)| \leq 1,$$

so that (i) follows.

(ii) If  $|f'(0)| = 1$ , then  $\lim_{z \rightarrow 0} |g(z)| = 1$ , so by the continuity of  $g$ ,  $g(0) = 1$ . But again, the Maximum Modulus Principle tells us that it is not possible for  $g$  to take its maximum value at an interior point, so we must have  $g(z) = c \in \mathbb{S}^1$ , a constant on  $\mathbb{D}$ . (ii) follows.

(iii) If  $|f'(0)| < 1$ , then  $|g(0)| < 1$ , and we must have  $|g(z)| < 1$  for all  $z \in \mathbb{D}$ . If not, then  $|g(z)| = 1$  for some interior point  $z$ , so  $f(z) = cz$ , contradicting  $|f'(0)| < 1$ .  $\square$

## APPENDIX D

### Weyl's Equidistribution Theorem.

We include this theorem as it is needed to complete the proof of the Three Distance Theorem in Chapter 19. Let  $x \in \mathbb{R}$ , then the integer part of  $x$  is  $\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}$  (called the *floor function*), and the *fractional part* of  $x$  is  $\{x\} = x - \lfloor x \rfloor$ .

**D1. Definition.** Let  $(a_n)$  be a sequence in  $[0, 1]$ .  $(a_n)$  is *uniformly distributed* in  $[0, 1]$ , if for every  $a, b \in [0, 1]$  with  $a < b$ , we have

$$\lim_{N \rightarrow \infty} \frac{|\{n \leq N : a_n \in (a, b)\}|}{N} = b - a.$$

We have seen in Chapter 17, that the sequence  $(\{n\alpha\})$ ,  $n = 1, 2, \dots$ , is dense in  $[0, 1]$ . Our aim here is to show the much stronger fact:  $(\{n\alpha\})$  is equidistributed in  $[0, 1]$ . To say that the sequence is equidistributed means that the terms of the sequence enter any interval in proportion to the length of that interval. In order to prove this, we use Weyl's criterion for equidistribution, stated without proof (see [80]):

**D2. Weyl's Equidistribution Theorem.** Let  $(a_n)$  be a sequence in  $[0, 1]$ .  $(a_n)$  is equidistributed in  $[0, 1]$ , if and only if for each  $k \in \mathbb{N}$ ,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N e^{2\pi i k a_n} = 0.$$

Denote by  $||\alpha||$  the distance of  $\alpha$  to the nearest integer. We now prove:

**D3. Theorem.** Let  $\alpha$  be irrational. The sequence  $(\{n\alpha\})$  is equidistributed in  $[0, 1]$ .

We first prove a lemma:

**D4 Lemma.** Let  $\alpha \in \mathbb{R}$ . For  $N \in \mathbb{N}$ ,

$$\left| \sum_{n=1}^N e^{2\pi i n \alpha} \right| \leq \min\{N, \frac{1}{2||\alpha||}\}.$$

**Proof.** If  $\alpha = 0$ , the sum is  $N$ . If  $\alpha \neq 0$ , we have a finite geometric series whose sum is

$$\frac{e^{2\pi i\alpha}(1 - e^{2\pi i n\alpha})}{1 - e^{2\pi i\alpha}} = \frac{e^{2\pi i\alpha}(1 - e^{2\pi i n\alpha})}{-2ie^{\pi i\alpha} \sin(\pi\alpha)}.$$

Taking the absolute value, we see the sum is bounded by  $|\sin(\pi\alpha)|^{-1}$ . The result follows from  $|\sin(\pi\alpha)| \geq 2||\alpha||$  (for example, consider the graphs of  $y = \sin(\pi x)$  and  $y = 2x$ , for  $0 \leq x \leq 1/2$ ). □

**Proof of D3.** If we can show that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N e^{2\pi i k n \alpha} = 0,$$

for every  $k \in \mathbb{N}$ , then the result will follow from Weyl's Theorem. From Lemma D4, since  $k\alpha$  is irrational, the sum is at most  $1/2||k\alpha||$ . Therefore

$$\left| \frac{1}{N} \sum_{n=1}^N e^{2\pi i k n \alpha} \right| < \frac{C}{N},$$

where  $C$  is the constant  $C = 1/2||k\alpha||$ . The result follows. □

## Bibliography

- [1] D. S. Alexander, *A History of Complex Dynamics: From Schröder to Fatou and Julia*, Vieweg, 1994.
- [2] J-P. Allouche and J. Shallit, The ubiquitous Prouet-Thue-Morse sequence, *Sequences and their applications* (Singapore, 1998), 116, Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999.
- [3] J-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications and Generalizations*, Cambridge University Press, 2003.
- [4] L. Alseda, J. Libre and M. Misiurewicz, *Combinatorial Dynamics and Entropy in Dimension One*, 2nd ed., Advanced Series in Nonlinear Dynamics, **5**, World Scientific, River Edge, NJ, 2000.
- [5] A. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, N.Y. Benjamin, 1967.
- [6] P. Arnoux and E. Harriss, What is ... a Rauzy fractal, *Notices of the A.M.S.*, **61** (2014), 768-770.
- [7] J. Auslander and J. Yorke, Interval maps, factor maps and chaos, *Tohoku Math. J.* **32** (1980), 177-188.
- [8] J. Banks, J. Brooks, G. Cairns, G. Davis and P. Stacey, On Devaney's definition of chaos, *American Math. Monthly*, **99**, 1992, 332-334.
- [9] B. Barna, Über die Divergenzpunkte des Newtonschen Verfahrens zur Bestimmung von Wurzeln algebraischer Gleichungen, I, *Publ. Math. Debrecen*, **3** (1953), 109-118.
- [10] J. Barrow-Green, *Poincaré and the Three Body Problem*, London Math. Soc and Amer. Math. Soc., 1997.
- [11] S. Bassein, The dynamics of one-dimensional maps, *Amer. Math. Monthly*, **105** (1998), 118-130.
- [12] J. Bechhoefer, The birth of period 3, revisited. *Mathematics Magazine*, **69** (1996), 115-118.
- [13] V. Berthé, Fréquences des facteurs des suites sturmian, *Theor. Comp. Sci.*, **165** (1996), 295-309.
- [14] V. Berthé, Sequences of low complexity: automatic and Sturmian sequences, in Topics in Symbolic Dynamics and Applications. Ed: F. Blanchard, A. Maass and A. Nogueira. LMS Lecture Notes Ser. **279**, CUP, 1-34.
- [15] V. Berthé, A. Siegel and J. Thuswaldner, Substitutions, Rauzy fractals and tilings, *Combinatorics, Automata and Number Theory*, ed. V. Berthé and M. Rigo. CUP, (2010), 248-266.
- [16] F. Blanchard, Topological chaos: what may this mean? *J. of Difference Equations and Applications*, **15** (2009), 23-46.
- [17] P. Blanchard, The dynamics of Newton's method, *Proceedings of Symposia in Applied Mathematics* **49**, Complex Dynamical Systems: The Mathematics Behind the Mandelbrot and Julia Sets, AMS, (1994), 139-154.
- [18] P. Blanchard, Complex analytic dynamics on the Riemann sphere, *Bull. Amer. Math. Soc.*, **11** (1984), 85-141.
- [19] L. S. Block and W. A. Coppel, *Dynamics in One Dimension*, Lecture Notes in Math., **1513**, Springer-Verlag, Berlin, 1992.

- [20] M. Boyle, *Algebraic aspects of symbolic dynamics*, Topics in Symbolic Dynamics and Applications, Eds. F. Blanchard, A. Maass, A. Nogueira, LMS Lecture Note Series, **279** (2000), 57-88.
- [21] A. Boyarsky and P. Gora, Invariant measures for Chebyshev maps, *J. Applied Math. and Stochastic Analysis*, **14** (2001), 257-264.
- [22] P. Bracken, Dynamics of the mapping  $f(x) = (x + 1)^{-1}$ , *Fibonacci Quarterly*, **33** (1995), 357-358.
- [23] T. C. Brown, A characterization of the quadratic rationals, *Canad. Math. Bull.*, **34** (1991), 36-41.
- [24] K. Brucks and H. Bruin, *Topics from One-Dimensional Dynamics*, Cambridge University Press, 2004.
- [25] K. Burns and B. Hasselblatt, The Sharkovsky Theorem: A natural direct proof, *Amer. Math. Monthly*, **118** (2011), 229-244.
- [26] M. P. de Carvalho, Chaotic Newton's sequences, *Mathematical Intelligencer*, **24** (2002), 31-35.
- [27] A. Cayley, Applications of the Newton-Fourier method to an imaginary root of an equation, *Quarterly J. Pure and Applied Math.*, **16** (1879), 179-185.
- [28] K. Conrad, Contraction mapping theorem. [www.math.uconn.edu/~kconrad/blurbs/analysis/contraction.pdf](http://www.math.uconn.edu/~kconrad/blurbs/analysis/contraction.pdf)
- [29] W. A. Coppel, The solution of equations by iteration, *Proc. Cambridge Philos. Soc.* **51** (1955) 41-43.
- [30] E. M. Coven and G. A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory*, **7** (1973), 138-153.
- [31] D. Crisp, W. Moran, A. Pollington and P. Shiue, Substitution invariant cutting sequences, *Journal de Théorie des Nombres de Bordeaux*, **5** (1993), 123-137.
- [32] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Benjamin Cummings, 1986.
- [33] F. M. Dekking, M. Mendés France and A. van der Poorten, Folds! *The Mathematical Intelligencer*, **4** (1982), 130-137.
- [34] F. M. Dekking, M. Mendés France and A. van der Poorten, Folds! II. Symmetry disturbed, *Mathematical Intelligencer*, **4** (1982), 173-181.
- [35] F. M. Dekking, M. Mendés France and A. van der Poorten, Folds! III. More morphisms, *Mathematical Intelligencer*, **4** (1982), 190-195.
- [36] F. M. Dekking, On the distribution of digits in arithmetic sequences, *Sém. Théorie des Nombres*, Bordeaux, Exposé 32, 1-12.
- [37] A. Douady and J. Hubbard, Itération des polynomes quadratiques complexes, *C. R. Acad. Sci. Paris*, **29**, Ser. I-1982, 123-126.
- [38] B.-S. Du, A collection of simple proofs of Sharkovsky's Theorem, arXiv:math/0703592v3 [math.DS] 9 Sep 2007.
- [39] B.-S. Du, A simple proof of Sharkovsky's theorem, *Amer. Math. Monthly*, **111** (2004), 595-599.
- [40] B.-S. Du, A simple proof of Sharkovsky's theorem revisited, *Amer. Math. Monthly*, **114** (2007), 152-155.
- [41] S. Elaydi, *Discrete Chaos*, Chapman and Hall/CRC, 2000.
- [42] S. Elaydi, A converse to Sharkovsky's Theorem, *Amer. Math. Monthly*, **99**, 1992, 332-334.
- [43] P. Fatou, Sur les équations fonctionnelles, *Bull. Soc. Math. France*, **47** (1918), 161-271, and **48** 33-94, 208-314.
- [44] B. Y. Feng, A trick formula to illustrate the period three bifurcation diagram of the logistic map. *Journal of Mathematical Research and Exposition*, **30**, 2010, 286-290.
- [45] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*. Lecture Notes in Mathematics, **1794**, Springer, 2002.

- [46] N. Priebe Frank, A primer of substitution tilings of the Euclidean plane, *Expo. Math.* **26** (2008), 295-326.
- [47] N. Priebe Frank, Multidimensional constant-length substitution sequences, *Topology and its Applications*, **152** (2005), 44-69.
- [48] N. A. Friedman, Replication and stacking in ergodic theory, *American Mathematical Monthly*, **99** (1992), 31-41.
- [49] N. A. Friedman, *Introduction to Ergodic Theory*, Van Nostrand Reinhold, New York, 1970.
- [50] H. Furstenberg and B. Weiss, Topological dynamics and combinatorial number theory, *J. Analyse Math.*, **34** (1978), 61-85.
- [51] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.*, **31** (1977), 204-256.
- [52] H. Furstenberg, *Ergodic Theory and Fractal Geometry*, CBMS Regional Conference Series in Mathematics, **120**, American Mathematical Society, 2014.
- [53] H. Furstenberg, Disjointness in ergodic theory, minimal sets and a problem in Diophantine approximation, *Math. Systems Theory*, **1** (1967), 1-49.
- [54] M. Gardner, Extraordinary nonperiodic tiling that enriches the theory of tiles, *Scientific American*, **236** (1977), 110-119.
- [55] W. J. Gilbert, The complex dynamics of Newton's method for a double root, *Computers Math. Applic.*, **22** (1991), 115-119.
- [56] E. Glasner, *Ergodic Theory via Joinings*, American Mathematical Society, Providence R.I., 2003.
- [57] E. Glasner and B. Weiss, On the interplay between measurable and topological dynamics, arXiv:math/0408328v1 [math.DS] 24 Aug 2004.
- [58] S. Golomb, Tilings with polyominoes, *Journal of Combinatorial Theory*, **1** (1966), 280-296.
- [59] G. R. Goodson, M. Lemańczyk, On the rank of a class of substitution dynamical systems, *Studia Math.*, **96** (1990), 219-230.
- [60] G. R. Goodson, On the spectral multiplicity of a class of finite rank transformations, *Proc. Amer. Math. Soc.*, **93** (1985), 303-306.
- [61] G. R. Goodson, Groups having elements conjugate to their squares and applications to dynamical systems, *Applied Mathematics*, **1** (2010), 416-424.
- [62] G. R. Goodson, Conjugacies between ergodic transformations and their inverses, *Colloquium Mathematicum*, **84** (2000), 185-193.
- [63] W. B. Gordon, Period three trajectories of the logistic map. *Mathematics Magazine*, **69** (1996), 118-120.
- [64] B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions, *Annals of Math.* **167** (2008), 481-547.
- [65] D. Gulick, *Encounters with Chaos*, McGraw Hill, 1992.
- [66] A. Hardy and W-H. Steeb, *Mathematical Tools in Computer Graphics with C# Implementations*, World Scientific Pub. Co., 2008.
- [67] J. Heidel, The existence of periodic orbits of the tent map, *Physics Letters A*, **143** (1990), 195-201.
- [68] M. Hénon, A two-dimensional mapping with a strange attractor, *Comm. Math. Phys.*, **50** (1976), 69-77.
- [69] R. A. Holmgren, *A First Course in Discrete Dynamical Systems*, Springer-Verlag, 1994.
- [70] D. Huang and D. Scully, Periodic points of the open tent function, *Math. Mag.*, **76** (2003), 204-213.
- [71] J. E. Hutchinson, Fractals and self-similarity, *Indiana Univ. Math. J.* **30** (1981), 713-747.

- [72] G. Julia, Memoire sur l'iteration des fonctions rationnelles. *Journeel des mathematiques pures et appliquees*, **4** (1918), 47-245.
- [73] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Encyclopedia of Mathematics and its Applications **54**, Cambridge University Press, 1995.
- [74] H. B. Keynes and J. B. Robertson, Eigenvalue theorems in topological transformation groups, *Transactions of the American Mathematical Society*, **139**, (1969), 359-369.
- [75] A. Khintchine, *Three Pearls of Number Theory*, Dover Publications Inc., Mineola, NY., 1998.
- [76] J. Kigami, *Analysis on Fractals*. Cambridge University Press, 2001.
- [77] B. P. Kitchens, *Symbolic Dynamics*, Springer, 1998.
- [78] G. Koenigs, Recherches sur le substitutions uniformes, *Bull. des Sciences math. et astron.*, Series 2, **7** (1883), 340-357.
- [79] S. Kolyada and L. Snoha, Some aspects of topological transitivity - a survey, *Grazer Math. Berichte*, **334** (1997), 3-35.
- [80] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Dover, 2006.
- [81] J. S. Lee, Toral automorphisms and chaotic maps on the Riemann sphere, *Trends in Math, Info. Center for Math. Sci.*, **2** (2004), 127-133.
- [82] M. H. Lee, Analytical study of the superstable 3-cycle in the logistic map. *Journal of Mathematical Physics*, **50** (2009), 122702, 1-6.
- [83] M. H. Lee, Three-cycle problem in the logistic map and Sharkovskii's Theorem. *Acta Physica Polonica B*, **42** (2011), 1071-1080.
- [84] T-Y. Li and J. A. Yorke, Period three implies chaos, *Amer. Math. Monthly* **82** (1975), 985-992.
- [85] M. Lothaire, *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.
- [86] J. Ma and J. Holdener, When Thue-Morse meet Koch, *Fractals*, **13** (2005), 191-206.
- [87] R. M. May, Simple mathematical models with very complicated dynamics. *Nature*, **261** (1976), 459-467.
- [88] B. Mandelbrot, *The Fractal Geometry of Nature*. W. H. Freeman, New York, 1982.
- [89] A. McKane and J. Pearson, *Non-Linear Dynamics*, University of Manchester Lecture Notes, (2007), 1-49.
- [90] M. Mendé France, Paper folding, space filling curves and Rudin-Shapiro sequences, *Contemporary Mathematics*, **9** (1982), 85-95.
- [91] J. Milnor. *Dynamics in One Complex Variable: Introductory Lectures*, Vieweg, 2000.
- [92] M. Misiurewicz, Remarks on Sharkovsky's Theorem. *American Mathematical Monthly*. **104** (1997), 846-847.
- [93] M. Misiurewicz, On the iterates of  $e^z$ . *Ergodic Theory and Dynamical Systems*. **1** (1981), 103-106.
- [94] R. Nillsen, Chaos and one-to-oneness. *Mathematics Magazine*, **72** (1999), 14-21.
- [95] V. Ovsienko and S. Tabachnikov, What is ... the Schwarzian derivative. *Notices Amer. Math. Soc.*, **56**, (2009), 34-36.
- [96] J. C. Oxtoby, *Measure and Category*. Springer-Verlag, 1971.
- [97] R. Palais, A simple proof of the Banach contraction principle. *J. Fixed Point Theory and Applications*, **2** (2007), 221-223.
- [98] J. Palmore, Newton's method and Schwarzian derivatives. *J. Dynamics and Differential Equations*, **6** (1994), 507-511.
- [99] W. Parry, *Topics in Ergodic Theory*. Cambridge University Press, Cambridge, 1981.
- [100] R. Penrose, Tilings and quasi-crystals: a non-local growth problem? Introduction to the Mathematics of Quasicrystals (Ed. Marco Jarić). Academic Press, 1989, 53-80.
- [101] P. Petek, A nonconverging Newton sequence. *Mathematical Magazine*, **56** (1983), 43-45.
- [102] K. Petersen, *Ergodic Theory*, Cambridge University Press, 1989.

- [103] M. Pollicott, Van der Waerden's theorem on arithmetic progressions, preprint.
- [104] M. Queffelec, *Substitution Dynamical Systems - Spectral Analysis*. Lecture Notes in Mathematics **1294**, 2nd Edition, Springer-Verlag, 2010.
- [105] H. Rademacher, *Lectures on Elementary Number Theory*, A Blaisdell book in pure and applied sciences, 1964.
- [106] G. Rauzy, Nombres algébrique et substitutions, Bull. Soc. Math. France, **110** (1982), 147-178.
- [107] G. Rauzy, Low complexity and geometry. Dynamics of complex interacting systems (Santiago, 1994) 147-177, Nonlinear Phenom, Complex Systems **2**, Kluwer Acad. Publ. Dordrecht, 1996.
- [108] A. E. Robinson, The dynamical theory of tilings and quasicrystallography. Ergodic theory of  $Z^d$ -actions (Warwick, 1993-1994), 451-473, London Math. Soc. Lecture Note Ser., **228**, Cambridge Univ. Press, Cambridge, 1996.
- [109] A. E. Robinson, On the table and the chair, *Indag. Mathem., N.S.* **10** (1999), 581-599.
- [110] D. Ruelle, What is ... a strange attractor? *Notices of the AMS*, **53** (2006), 764-765.
- [111] D. G. Saari and J. B. Urenko, Newton's method, circle maps and chaotic motion. *Amer. Math. Monthly*, **91** (1984), 3-17.
- [112] P. Saha and S. H. Strogatz. The birth of period three. *Mathematics Magazine*, **68** (1995), 42-47.
- [113] E. Schröder, Ueber iterite funktionen, *Mathematische Annalen*, **3** (1871), 296-322.
- [114] H. Sedaghat, The impossibility of unstable, globally attracting fixed points for continuous mappings of the line. *American Mathematical Monthly*, **104** (1997), 356-358.
- [115] D. Singer, Stable orbits and bifurcation of maps of the interval. *SIAM J. Appl. Math.* **35** (1978), 260-267.
- [116] H. J. S. Smith, On the integration of discontinuous functions, *Proc. Lond. Math. Soc.*, **6** (1874), 140-153.
- [117] H. Sohrab, *Basic Real Analysis*, Second Edition, Birkhauser (Springer Science+Business Media New York 2003, 2014).
- [118] B. Solomyak, Dynamics of self-similar tilings, *Ergodic Theory, Dynamical Systems*, **17** (1997), 695-738.
- [119] D. Sprows, Digitally determined periodic points. *Mathematics Magazine*, **71** (1998), 304-305.
- [120] R. Stankewitz and J. Rolf, Chaos, fractals, the Mandelbrot set, and more. *Explorations in Complex Analysis*, 1-83, Classr, Res. Mater. MAA, Washington DC, 2012.
- [121] P. Štefan, A theorem of Sharkovsky on the existence of periodic orbits of continuous endomorphisms of the real line. *Commun. Math. Phys.* **54** (1977), 237-248.
- [122] S. Sternberg, *Dynamical Systems*, Dover, 2009.
- [123] P. D. Straffin, Jr., Periodic points of continuous functions, *Math. Mag.* **51** (1978), 99-105.
- [124] M. Vellekoop and R. Berglund, On intervals: transitivity → chaos. *American Math. Monthly*, **101** (1994), 353-355.
- [125] B. L. van der Waerden, Beweis eiener Baudetschen Vermutung, *Nieuw. Arch. Wisc.* **15** (1928), 212-216.
- [126] J. A. Walsh, The dynamics of Newton's method for cubic polynomials. *The College Mathematics Journal*, **26** (1995), 22-28.
- [127] P. Walters, *An Introduction to Ergodic Theory*. Springer Verlag, 1981.
- [128] E. W. Weisstein, "Logistic Map". From MathWorld – A Wolfram Web Resource: <http://mathworld.wolfram.com/LogisticMap.html>.



## Index

- $2^n$ -cycle, 220
- $2^n$ -cycles of the tent family, 174
- $\omega$ -limit set, 351
- 2-cycle, 45, 52, 57, 204
- 3-cycle, 206
- 3-cycles of the logistic map, 59, 65
- 3-cycles of the tent map, 50
- adding machine, 310, 366
- affine maps, 5
- affine transformation, 19
- algebraic number, 324
- almost everywhere, 108
- almost periodicity, 371
- alphabet, 367
- analytic function, 246, 415
- angle doubling map, 128, 148, 150, 354, 399
- angle tripling map, 151
- aperiodic sequence, 368
- asymptotically stable fixed point, 21, 22, 48, 96, 226, 247
- attracting fixed point, 18, 19, 22, 42, 96, 247
- attractor, 119
- B. Weiss, 405
- Baire category theorem, 340, 373, 413
- Baire space, 413
- Baker's transformation, 238
- Banach fixed-point theorem, 190
- basin of attraction, 41, 250
- basin of attraction of  $\infty$ , 255
- Bau-Sen Du, 209
- Benoit Mandelbrot, 179, 266
- Bernoulli shift, 135, 353
- bifurcation, 43
- bifurcation diagram, 59, 162
- bifurcation theory, 41
- bijective substitution, 322
- billiard sequence, 392
- binary expansion of a real number, 109, 197
- Binet's formula, 332
- Birkhoff multiple recurrence theorem, 408
- Birkhoff transitivity theorem, 130, 348, 406
- Birkhoff's recurrence theorem, 352
- bisection method, 106
- block code, 379
- Bolzano-Weierstrass theorem, 412
- box counting dimension, 182
- Brouwer's fixed point theorem, 196
- Cantor set, 112, 116, 179, 182, 195, 206, 314, 363
- Cantor's middle thirds set, 112
- Cantor's ternary function, 118
- cardinality, 110
- cardinality of  $\mathbb{A}$  set, 114
- Cauchy sequence, 188, 189, 413
- Cauchy's estimate, 416
- Cauchy's theorem, 262
- Cayley, 36, 179
- Chacon substitution, 324, 357, 375
- chaotic map, 131, 262, 354
- characteristic equation of a matrix, 224
- Charles Pisot, 325
- Chebyshev polynomials, 146, 151
- class  $C^1$  function, 23, 101
- clopen set, 366
- closed ball, 88
- closed bounded interval theorem, 341, 412
- closed form solution, 4, 6, 228
- closed set, 88
- closure of a set, 89
- commutative substitution, 322

- compact metric space, 336, 362
- compact set, 114
- compactness, 335, 340, 361
- complement of a set, 88
- complete metric space, 188
- completeness of  $\mathbb{R}$ , 411
- complex arctangent function, 283
- complex dynamics, 243
- complex exponential function, 281
- complex logarithm, 282
- complex numbers, 243, 415
- complex sine and cosine, 282
- complexity function, 367, 381, 401
- computer graphics for the Mandelbrot set, 272
- conjugacy, 143, 167, 346
- conjugacy - properties, 146
- conjugacy between  $T_2$  and  $L_4$ , 176
- conjugacy is an equivalence relation, 145
- conjugate maps, 143
- connected set, 100, 363
- continued fractions, 388
- continuity on  $\mathbb{C}$ , 245
- continuous function, 3, 341
- continuous functions - properties, 100
- continuous map on metric spaces, 94
- contraction mapping, 189
- contraction mapping theorem, 190, 342
- Coppel's theorem, 209, 212
- correlation function of a sequence, 304
- countable set, 107, 108, 377
- cube free sequence, 369
- cube roots of unity, 276
- cutting sequence, 390
- cycle, 48
- cylinder set, 366
- Daniel Alexander, 243
- David Singer, 162
- DeMoivre's theorem, 244
- dense set, 89
- denumerable set, 108
- diagonal argument, 109
- diagonalization of a matrix, 225
- diffeomorphism, 101, 233
- difference equation, 3
- direct product substitution, 321
- disconnected set, 363
- discrete dynamical system, 2
- discrete metric, 86
- disjoint dynamical systems, 359
- doubling map, 127, 150
- dragon curve, 314
- dyadic rationals, 200
- dynamical system, 3, 9, 361, 369
- dynamical systems arising from substitutions, 369
- eigenfunction of a continuous map, 349
- eigenspace, 224
- eigenvalue of a continuous map, 349
- eigenvalue of a linear map, 224
- eigenvectors, 225
- empty word, 367
- escape criterion, 257
- even function, 54
- eventual fixed point, 13
- eventually periodic point, 48
- exactness, 353
- extended complex plane, 250
- extended real line, 99
- factor, 356
- factor map, 144, 346
- factor of a sequence, 367
- Fatou, 179
- Fatou set, 255
- Fatou's theorem, 268
- Feigenbaum's number, 59
- Fibonacci sequence, 6, 10, 18, 291
- Fibonacci substitution, 295, 365, 375, 382, 399
- filled-in Julia set, 255, 258
- fixed point, 12, 14, 15, 17, 23, 37, 224
- fixed point at infinity, 250
- fixed point of a substitution, 297, 364
- fixed point theorem, 14, 195
- fixed points of linear fractional transformations, 253
- floor function, 386
- flow, 2
- fractal, 113, 179, 233, 266
- fractal dimension, 181, 255
- fractal dust, 262
- fractal geometry, 179
- full shift, 370
- fundamental dichotomy, 267
- fundamental domain, 167

- Gérard Rauzy, 323  
 Gaston Julia, 162, 249  
 George Cantor, 112  
 globally attracting fixed point, 19, 26, 31, 44  
 graph joinings, 359  
 graphical iteration, 18  
 Green-Tao theorem, 405  
 group rotation, 366
- H. Furstenberg, 359, 405  
 Hénon attractor, 232  
 Hénon map, 231  
 Hénon map - properties, 234  
 Halley's method, 284  
 Hausdorff metric, 192  
 Heighway dragon, 315  
 Heine-Borel theorem, 110, 335  
 Henry Smith, 112  
 Hermann Schwartz, 36  
 homeomorphism, 97, 141, 143, 167, 366, 408  
 Hutchinson's Theorem, 191  
 hyperbolic  $n$ -cycle, 51  
 hyperbolic fixed point, 22, 226  
 hyperbolic toral automorphisms, 239
- identity map, 135  
 immediate basin of attraction, 41, 97  
 incidence matrix, 332  
 incidence matrix of a substitution, 323, 375, 378  
 infimum, 411  
 intermediate value theorem, 14, 106, 341, 411  
 invariant subset, 42, 47, 97  
 involution, 146  
 irrational rotation, 134, 345, 347, 381, 385, 394  
 Isaac Newton, 10  
 isolated fixed point, 39  
 isolated point, 413  
 isometry, 99, 135, 349  
 iterated function system, 194
- Jacobian matrix, 231  
 John von Neumann, 7, 308  
 joinings, 359  
 Jordan form of a matrix, 230  
 Joseph Lagrange, 36  
 Joseph Raphson, 10  
 Julia, 179  
 Julia set, 249, 254, 260, 266
- Julia set - properties, 264  
 Julia set of a polynomial, 255
- K. Petersen, 343  
 Karl Weierstrass, 179  
 Koch curve, 311  
 Koch snowflake, 180  
 Kronecker product, 186
- languages, 367  
 languages and words, 367  
 left special factor, 401  
 Li-Yorke Theorem, 75, 125  
 limit point, 88  
 linear conjugacy, 152  
 linear conjugacy - properties, 154  
 linear fractional transformation, 159, 252  
 linear maps, 4  
 linear model, 4  
 linear transformations, 223  
 linearly conjugate maps, 152  
 Liouville's theorem, 416  
 logistic map, 6–8, 17, 24, 37, 43, 46, 47, 55, 57, 63, 66, 148, 161, 162, 168, 213
- M. Pollicott, 406  
 Möbius transformation, 252  
 Mandelbrot set, 179, 266, 267  
 Mandelbrot set - properties, 272  
 Maurice Hénon, 231  
 maximum modulus principle, 415  
 mean value theorem, 14, 23, 29, 412  
 Menger sponge, 184, 321  
 metric, 85  
 metric space, 85  
 minimal map, 134, 349, 372, 376  
 minimal maps - properties, 345  
 minimal set, 344, 349, 352  
 minimality, 344  
 mixing, 353  
 monic polynomial, 202  
 monotone sequence theorem, 411  
 Morse substitution, 375  
 multiple recurrence theorem, 405
- N. Pytheas Fogg, 375  
 negative Schwarzian derivative, 161  
 nested sequence, 337  
 neutral fixed point, 247

- Newton function, 11, 29, 196, 201, 276
- Newton's method, 10, 26, 197, 277, 278
- Newton's method - relaxed, 279
- Newton's method for complex cubics, 275
- Newton's method for complex quadratics, 273
- Newton's method for real cubics, 201
- Newton's method for real quadratics, 197, 199, 273
- Newton's method in the complex plane, 273
- non-hyperbolic  $n$ -cycle, 51
- non-hyperbolic fixed point, 22, 31, 38
- non-negative matrix, 323
- non-wandering point, 350
- nowhere dense set, 413
- odd function, 54
- one-dimensional map, 3
- one-to-one substitution, 379
- open ball, 87
- open cover, 108, 335
- open set, 87
- orbit, 3, 168
- order preserving diffeomorphism, 103
- P. Fatou, 243
- P. Walters, 237, 343
- palindrome, 314
- paperfolding sequence - properties, 308
- paperfolding sequences, 291, 304
- paperfolding substitution, 307
- Penrose tiling, 319
- perfect set, 116
- period, 48
- period doubling, 57
- period doubling bifurcation, 56, 221
- period doubling route to chaos, 60
- period doubling substitution, 299
- periodic point, 48, 247
- periodic points of polynomials - properties, 252
- periodic points of the doubling map, 138
- periodic sequence, 368
- Perron-Frobenius theorem, 324
- Pisot irreducible substitution, 326
- Pisot number, 324, 325, 333
- population dynamics, 4
- positive definite sequence, 304
- primary cardioid of the Mandelbrot set, 270
- primitive matrix, 323
- primitive substitution, 323, 375
- prototile, 319
- quadratic irrational, 388
- R. Devaney, 177, 249
- R. Palais, 190
- Rauzy fractal, 323, 327, 328
- Rauzy graph, 401
- real linear transformation, 223
- recurrence relations, 6
- recurrent point, 351
- recurrent sequence, 371, 376
- recursively, 6
- renormalization operator, 177
- rep-tiles, 318
- repelling 2-cycle, 204
- repelling fixed point, 18, 21, 226, 247
- Riemann sphere, 251
- Riemann sphere metric, 254
- right extension of a factor, 401
- right special factor, 401
- Rolles' theorem, 14, 412
- Rudin-Shapiro sequence, 291, 301
- Rudin-Shapiro sequence - properties, 303
- Rudin-Shapiro substitution, 302
- S. Kakutani, 308
- S. Sternberg, 4, 168
- saddle point, 226
- Schröder, 243
- Schröder's theorem, 247
- Schröder-Cayley theorem, 274, 285
- Schwarz lemma, 416
- Schwarzian derivative, 35, 38, 157
- self-similar sets, 182
- self-similarity, 113
- semi-stable fixed point, 31
- semi-topological conjugacy, 330, 397, 399
- sensitive dependence on initial conditions, 130, 139, 355
- sequence space, 92
- sequential compactness, 336
- set of measure zero, 107
- sets of measure zero - properties, 113
- Sharkovsky ordering, 75, 213
- Sharkovsky's theorem, 76, 82, 206, 209, 218
- Sharkovsky's theorem - converse, 80, 219
- shift dynamical system, 330

- shift map, 136, 353, 370, 399, 407
- shift map is mixing, 356
- Sierpinski carpet, 185, 321
- Sierpinski gadget, 186
- Sierpinski triangle, 194, 195
- Singer's Theorem, 159
- smooth function, 101
- spectrum of a sequence, 304
- square free sequence, 369
- stable fixed point, 22, 30, 96, 152
- stable orbit, 21
- stable periodic point, 48
- stacking transformation, 308
- Stanislaw Ulam, 7
- Stefan Banach, 190
- strange attractor, 231
- Sturmian sequence, 368, 381, 382, 391, 394, 398, 399
- sub-shift, 370
- subspace, 89
- substitution, 291, 293, 296, 361, 364
- substitution dynamical system, 361, 375
- substitution of constant length, 295, 316, 321
- super-attracting 2-cycle, 58
- super-attracting 3-cycle of the logistic map, 65
- super-attracting fixed point, 27, 28, 53, 57
- supremum, 189, 411
- symbolic dynamical system, 361
- table and chair tilings, 318, 320
- tent family, 69, 73, 172
- tent map, 13, 50, 69, 70, 133, 150, 151, 162, 352, 354
- tent map  $T_3$ , 142
- ternary expansion of a real number, 115
- Thomas Simpson, 10
- three distance theorem, 381, 400
- Thue-Morse sequence, 291, 312, 365, 371
- Thue-Morse sequence - properties, 298
- tiling of the plane, 323
- Toeplitz sequence, 291, 306
- Toeplitz substitution, 299
- Toeplitz substitution is a factor of the Morse substitution, 378
- topological dimension, 181
- topological dynamical system, 343, 361, 405
- topological dynamics, 335, 343
- topological entropy, 385
- toral maps, 235
- total boundedness, 336
- totally disconnected set, 116, 363
- transitive, 353
- transitive map, 128, 355, 366
- transitive maps - properties, 347
- transitive point, 128
- tribonacci number, 330
- tribonacci sequence, 10, 332
- tribonacci substitution, 323, 332
- truncated tent map, 82, 84, 219
- two-dimensional substitution, 316
- Tychonoff's theorem, 339, 361
- uncountable set, 109, 373, 377
- unimodal map, 59, 354
- unstable fixed point, 22
- unstable orbit, 21
- van der Waerden's theorem, 405
- wandering point, 350
- weakly mixing map, 357, 366, 375
- web diagram, 18, 29
- Weyl's equidistribution theorem, 403, 417
- words, 367